Lecture Notes in Mathematics 1665
Editors:
A. Dold, Heidelberg
F. Takens, Groningen

Springer
Berlin
Heidelberg
New York
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

E. Giné G.R. Grimmett L. Saloff-Coste

Lectures on Probability Theory and Statistics

Ecole d'Eté de Probabilités
de Saint-Flour XXVI – 1996

Editor: P. Bernard

Springer

Authors

Evarist Giné
Department of Mathematics
University of Connecticut
196 Auditorium Road (U-9, R. 111)
Storrs, CT 06269-3009, USA

Geoffrey R. Grimmett
Statistical Laboratory
University of Cambridge
16 Mill Lane
Cambridge, CB2 1SB, UK

Laurent Saloff-Coste
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
118, Route de Narbonne
F-31062 Toulouse, France

Editor

Pierre Bernard
Laboratoire de Mathématiques Appliquées
UMR CNRS 6620
Université Blaise Pascal
Clermont-Ferrand
F-63177 Aubière, France

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Lectures on probability theory and statistics / Ecole d'Eté de Probabilités de Saint-Flour XXVI – 1996. E. Giné ... Ed.: P. Bernard. – Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997
(Lecture notes in mathematics; Vol. 1665)
ISBN 3-540-63190-9

Mathematics Subject Classification (1991): 46N30, 47D07, 60-01, 60-06, 60B12, 60F15, 60G50, 60G60, 60J25, 60J27, 60K35, 62-01, 62-02, 62-06, 62G05

ISSN 0075-8434
ISBN 3-540-63190-9 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera-ready TEX output by the authors
SPIN: 10553291 46/3142-543210 - Printed on acid-free paper

INTRODUCTION

This volume contains lectures given at the Saint-Flour Summer School of Probability Theory during the period 19th August - 4th September, 1996.

We thank the authors for all the hard work they accomplished. Their lectures are a work of reference in their domain.

The school brought together 74 participants, 38 of whom gave a lecture concerning their research work.

At the end of this volume you will find the list of participants and their papers.

Finally, to facilitate research concerning previous schools we give here the number of the volume of "Lecture Notes" where they can be found :

Lecture Notes in Mathematics

1971 : n°307 -	1973 : n°390 -	1974 : n°480 -	1975 : n°539 -	1976 : n°598 -
1977 : n°678 -	1978 : n°774 -	1979 : n°876 -	1980 : n°929 -	1981 : n°976 -
1982 : n°1097 -	1983 : n°1117 -	1984 : n°1180 -	1985 - 1986 et	1987 : n°1362 -
1988 : n°1427 -	1989 : n°1464 -	1990 : n°1527 -	1991 : n°1541 -	1992 : n°1581
1993 : n°1608 -	1994 : n°1648			

Lecture Notes in Statistics

1986 : n°50

TABLE OF CONTENTS

Evarist GINE : "DECOUPLING AND LIMIT THEOREMS FOR U-STATISTICS AND U-PROCESSES" 1

1. Introduction 2
2. Decoupling inequalities 3
3. Limit theorems for U-statistics 12
4. Limit theorems for U-processes 24

References 32

Evarist GINE : "LECTURES ON SOME ASPECTS THEORY OF THE BOOTSTRAP" 37

Preface 39

1. On the bootstrap in $\mathbb{R}$ 41

   1.1 Efron's bootstrap of the mean in $\mathbb{R}$ 41
   1.2 The general exchangeable bootstrap of the mean 54
   1.3 The bootstrap of the mean for stationary sequences 59
   1.4 The bootstrap of U-statistics 64
   1.5 A general m out of n bootstrap 76

2. On the bootstrap for empirical processes 83

   2.1 Background from empirical process theory 83
   2.2 Poissonization inequalities and Efron's bootstrap in probability 103
   2.3 The almost sure Efron's bootstrap 113
   2.4 The exchangeable bootstrap 119
   2.5 Uniformly pregaussian classes of functions and the bootstrap 128
   2.6 Some remarks on applications 139

References 147

Geoffrey GRIMMETT : "PERCOLATION AND DISORDERED SYSTEMS" 153

Preface 154

1. Introductory Remarks 158

1.1 Percolation 158
1.2 Some possible questions 158
1.3 History 159

2. Notation and Definitions 161

2.1 Graph terminology 161
2.2 Probability 162
2.3 Geometry 162
2.4 A partial order 164
2.5 Site percolation 165

3. Phase Transition 166

3.1 Percolation probability 166
3.2 Existence of phase transition 166
3.3 A question 168

4. Inequalities for Critical Probabilities 169

4.1 Russo's formula 169
4.2 Strict inequalities for critical probabilities 171
4.3 The square and triangular lattices 172
4.4 Enhancements 176

5. Correlation Inequalities 178

5.1 FKG inequality 178
5.2 Disjoint occurrence 181
5.3 Site and bond percolation 182

6. Subcritical Percolation 185

6.1 Using subadditivity 185
6.2 Exponential decay 188
6.3 Ornstein-Zernike decay 198

7. Supercritical Percolation 199

7.1 Uniqueness of the infinite cluster 199
7.2 Percolation in slabs 201
7.3 Limit of slab critical points 202
7.4 Percolation in half-spaces 215
7.5 Percolation probability 216
7.6 Cluster-size distribution 218

8. Critical Percolation 219

8.1 Percolation probability 219
8.2 Critical exponents 219
8.3 Scaling theory 219
8.4 Rigorous results 221
8.5 Mean-field theory 221

9. Percolation in Two Dimensions 229

9.1 The critical probability is 1/2 229
9.2 RSW technology 230
9.3 Conformal invariance 233

10. Random Walks in Random Labyrinths 236

10.1 Random walk on the infinite percolation cluster 236
10.2 Random walks in two-dimensional labyrinths 240
10.3 General labyrinths 248

11. Fractal Percolation 254

11.1 Random fractals 254
11.2 Percolation 255
11.3 A morphology 258
11.4 Relationship to Brownian Motion 260

12. Ising and Potts Models 261

12.1 Ising model for ferromagnets 261
12.2 Potts models 262
12.3 Random-cluster models 263

13. Random-Cluster Models 265

13.1 Basic properties 265
13.2 Weak limits and phase transitions 266
13.3 First and second order transitions 268
13.4 Exponential decay in the subcritical phase 269
13.5 The case of two dimensions 273

References 281

Laurent SALOFF-COSTE : "LECTURES ON FINITE MARKOV CHAINS" 301

1. Introduction and background material 304

   1.1 Introduction 304
   1.2 The Perron-Frobenius Theorem 309
   1.3 Notation for finite Markov chains 319

2. Analytic tools 326

   2.1 Nothing but the spectral gap 326
   2.2 Hypercontractivity 332
   2.3 Nash inequalities 344
   2.4 Distances 358

3. Geometric tools 367

   3.1 Adapted edge sets 368
   3.2 Poincaré inequality 368
   3.3 Isoperimetry 379
   3.4 Moderate growth 394

4. Comparison techniques 396

   4.1 Using comparison inequalities 396
   4.2 Comparison of Dirichlet forms using paths 400

Bibliography 408

DECOUPLING AND LIMIT THEOREMS FOR U-STATISTICS AND U-PROCESSES

Evarist GINE (*)

(*) Work partially supported by NSF Grants Nos. DMS-9300725 and DMS-9625457 and by the CNRS through the Laboratoire de Mathématiques Appliquées de l'Université Blaise Pascal, Clermont-Ferrand, France.

1. Introduction. Recently discovered decoupling inequalities for U–processes (mainly, de la Peña, 1992, and de la Peña and Montgomery–Smith, 1995) have had important consequences for the asymptotic theory of U–statistics and U–processes (Giné and Zinn, 1994, and Arcones and Giné, 1993, 1995, among others). It is the object of these lectures to describe these developments.

U–statistics, first considered by Halmos (1946) in connection with unbiased estimators and formally introduced by Hoeffding (1948), are defined as follows: given an i.i.d. sequence of random variables $\{X_i\}_{i=1}^{\infty}$ with values in a measurable space $(S, \mathcal{S})$, and a measurable function $h : S^m \to \mathbb{R}$, the U–statistics of order m and kernel h based on the sequence $\{X_i\}$ are

$$U_n(h) = \frac{(n-m)!}{n!} \sum_{(i_1,\ldots,i_m)\in I_n^m} h(X_{i_1}, \ldots, X_{i_m}), \quad n \geq m, \tag{1.1}$$

where

$$I_n^m = \{((i_1, \ldots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}.$$

These objects appear often in Statistics either as unbiased estimators of parameters of interest or, perhaps more often, as components of higher order terms in expansions of smooth statistics (von Mises expansion, delta–method). Particularly in connection with von Mises expansions, it is sometimes convenient to also consider U–processes indexed by families $\mathcal{H}$ of kernels, that is, collections of U–statistics $\{U_n(h) : h \in \mathcal{H}\}$.

By a *decoupling* result for U–statistics we mean a (usually two–sided) inequality between the quantities

$$\mathbb{E}\Phi\big(\big|\sum_{I_n^m} h(X_{i_1}, \ldots, X_{i_m})\big|\big) \tag{1.2}$$

and

$$\mathbb{E}\Phi\big(\big|\sum_{I_n^m} h(X_{i_1}^1, \ldots, X_{i_m}^m)\big|\big), \tag{1.3}$$

possibly multiplied by constants that depend only on m, where the sequences $\{X_i^k\}$, $k = 1, \ldots, m$, are independent copies of the original sequence $\{X_i\}$, and Φ is a non–negative function. Two quite different types of functions Φ have been considered: Φ convex (thus including $\Phi(|x|) = |x|^p$, $p \geq 1$) and $\Phi(|x|) = I_{|x|\geq t}$.

The variables at the different coordinates of the domain of h in the decoupled statistic come from different independent sequences and therefore a decoupled U–statistic can be treated, conditionally on all but one of these sequences, as a sum of independent random variables. Clearly then, decoupling inequalities will allow for conditional use of Lévy type maximal inequalities and for randomization by Rademacher variables, which then turn U–statistics into Rademacher chaos processes conditionally on the X samples. In this way, the analysis of U–processes can proceed more or less by analogy with that of empirical processes.

In Section 2 we describe the pertinent decoupling results and the randomization lemma. Section 3 is devoted to the central limit theorem and to the law of the iterated logarithm for U–statistics, and Section 4 to U–processes.

Contrary to the case of the bootstrap lectures in this volume, which are almost self-contained, here we present no technical details and refer the reader, instead, to the book 'An Introduction to Decoupling inequalities and Applications' by de la Peña and myself, in preparation, or to the original articles.

I thank the organizers of, and the participants in, the Saint–Flour École d'Été de Calcul de Probabilités for the opportunity to present these lectures. I would like to mention here that both, these lectures and the bootstrap lectures in this volume have their origin in a short course on these same topics that I gave at the Université de Paris-Sud (Orsay) in 1993. It is therefore a pleasure for me to also extend my gratitude to the Orsay Statistics group.

2. Decoupling inequalities. *a) Decoupling.* Let $(S, \mathcal{S}, \mathrm{P})$ be a probability space. Consider collections $\mathcal{H}_{i_1\ldots i_m}$ of measurable functions $h : S^m \to \mathbb{R}$ for $(i_1, \ldots, i_m) \in I_n^m$, with $n > m$ (these functions can also be Banach space valued, but this would not actually change the level of generality of the results to be stated below). It is convenient to have the following definition: an *envelope* (or a measurable envelope) of a class of functions $\mathcal{H}_{i_1\ldots i_m}$ is any measurable function $H_{i_1\ldots i_m}$ such that $\sup_{h\in\mathcal{H}_{i_1\ldots i_m}} |h(x_1, \ldots, x_m)| \le H_{i_1\ldots i_m}(x_1, \ldots, x_m)$ for all $x_1, \ldots, x_m \in S$. All the classes of functions considered here will have everywhere finite envelope.

Some standard notation: We set

$$\|h(X_1, \ldots, X_m)\|_{\mathcal{H}} := \sup_{h\in\mathcal{H}} |h(X_1, \ldots, X_m)|$$

for any collection of kernels $\mathcal{H}$, and we even write $\|h\|$ for $\|h\|_{\mathcal{H}}$ if no confusion is possible. Since these are often uncountable suprema of random variables, they may not be measurable; in this case we write $\mathbb{E}^*$ and Pr^* for outer expectation and probability (see the lecture notes on the bootstrap in this volume, chapter 2).

The main result about decoupling that we use in this article is the following theorem of de la Peña (1992):

2.1. THEOREM. *For natural numbers $n \ge m$, let X_i, X_i^k, $i = 1, \ldots, n$, $k = 1, \ldots, m$, be the coordinate functions of the product probability space $(S^{n(m+1)}, \mathcal{S}^{n(m+1)}, (P_1 \times \cdots \times P_n)^{m+1})$, in particular the variables $\{X_i\}_{i=1}^n$ are independent S–valued random variables, X_i with probability law P_i, $i \le n$, and the sequences $\{X_i^k\}_{i=1}^n$, $k \le m$, are i.i.d. copies of the sequence $\{X_i\}_{i=1}^n$. For each $(i_1, \ldots, i_m) \in I_n^m$, let $\mathcal{H}_{i_1\ldots i_m}$ be a collection of measurable functions $h_{i_1\ldots i_m} : S^m \to \mathbb{R}$ admitting an everywhere finite measurable envelope $H_{i_1\ldots i_m}$ such that $\mathbb{E}H_{i_1\ldots i_m}(X_1, \ldots, X_m) < \infty$. Let $\Phi : [0, \infty) \to [0, \infty)$ be a convex non–decreasing function such that $\mathbb{E}\Phi(H_{i_1\ldots i_m}(X_1, \ldots, X_m)) < \infty$ for all $(i_1, \ldots, i_m) \in I_n^m$. Then,*

$$\mathbb{E}^*\Phi\Big(\sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}} \Big|\sum_{I_n^m} h_{i_1\ldots i_m}(X_{i_1}, \ldots, X_{i_m})\Big|\Big)$$
$$\le \mathbb{E}^*\Phi\Big(C_m \sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}} \Big|\sum_{I_n^m} h_{i_1\ldots i_m}(X_{i_1}^1, \ldots, X_{i_m}^m)\Big|\Big) \qquad (2.1)$$

where $C_m = 2^m(m^m - 1)((m-1)^{(m-1)} - 1) \times \cdots \times 3$. *If, moreover, the classes* $\mathcal{H}_{i_1\ldots i_m}$ *satisfy that for all* $h_{i_1\ldots i_m} \in \mathcal{H}_{i_1\ldots i_m}$, $x_1, \ldots, x_m \in S$ *and permutations* s *of* $\{1, \ldots, m\}$,

$$h_{i_1\ldots i_m}(x_1, \ldots, x_m) = h_{i_{s_1}\ldots i_{s_m}}(x_{s_1}, \ldots, x_{s_m}), \tag{2.2}$$

then

$$\mathbb{E}^*\Phi\Big(\frac{1}{2^{(2m-2)}(m-1)!} \sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}} \Big|\sum_{I_n^m} h_{i_1\ldots i_m}(X_{i_1}^1, \ldots, X_{i_m}^m)\Big|\Big)$$
$$\leq \mathbb{E}^*\Phi\Big(\sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}} \Big|\sum_{I_n^m} h_{i_1\ldots i_m}(X_{i_1}, \ldots, X_{i_m})\Big|\Big). \tag{2.3}$$

For the proof of this theorem we refer to the above mentioned article of de la Peña or to our forthcoming book. However, we indicate here the proof of Theorem 1.2 for $m = 2$ and $\mathcal{H}_{i_1 i_2} = \mathcal{H}$ for all $\{i_1, i_2\}$, countable. In this proof, $\|\cdot\|$ will denote the sup over $h \in \mathcal{H}$.

Proof of Theorem 2.1 under the stated restrictions. We replace X_i^1, X_i^2 respectively by X_i, X_i'. Let $\{\varepsilon_i\}_{i=1}^n$ be independent random variables uniformly distributed on a set with two elements, say $\{-1, 1\}$, independent of $\{X_i, X_i'\}_{i=1}^n$, and let Z_i and Z_i', $i = 1, \ldots n$, be defined as follows:

$$Z_i = \begin{cases} X_i & \text{if } \varepsilon_i = 1 \\ X_i' & \text{if } \varepsilon_i = -1 \end{cases}, \quad Z_i' = \begin{cases} X_i' & \text{if } \varepsilon_i = 1 \\ X_i & \text{if } \varepsilon_i = -1 \end{cases}. \tag{2.4}$$

If, for each $1 \leq i \leq n$, P_i is the law of X_i then the law of the vector $(Z_1, \ldots, Z_n, Z_1', \ldots, Z_n')$ is $(P_1 \times \cdots \times P_n)^2$ since for each fixed $\varepsilon_1, \ldots, \varepsilon_n$ the coordinates of this vector are just $2n$ independent variables such that P_i is the law of the i-th and the $(n+i)$-th, $i = 1, \ldots, n$. That is,

$$\mathcal{L}(Z_1, \ldots, Z_n, Z_1', \ldots, Z_n') = \mathcal{L}(X_1, \ldots, X_n, X_1', \ldots, X_n').$$

Likewise, $\mathcal{L}(Z_1, \ldots, Z_n) = \mathcal{L}(X_1, \ldots, X_n)$. Therefore, for any $(P_1 \times \cdots \times P_n)^2$–integrable functions f and $(P_1 \times \cdots \times P_n)$–integrable functions g we have

$$\mathbb{E}f(Z_1, \ldots, Z_n, Z_1', \ldots, Z_n') = \mathbb{E}f(X_1, \ldots, X_n, X_1', \ldots, X_n'),$$
$$\mathbb{E}g(Z_1, \ldots, Z_n) = \mathbb{E}g(X_1, \ldots, X_n). \tag{2.5}$$

Note also that, if $\mathcal{Z}$ is the σ–algebra generated by the X variables,

$$\mathcal{Z} = \sigma(X_i, X_i' : i = 1, \ldots, n),$$

then conditional integration with respect to $\mathcal{Z}$ of any function of the Z variables is simply integration with respect to the ε_i variables only. In particular, for all $i \neq j$,

$$\mathbb{E}(h(Z_i, Z_j)|\mathcal{Z}) = \mathbb{E}(h(Z_i, Z_j')|\mathcal{Z}) = \mathbb{E}(h(Z_i', Z_j)|\mathcal{Z}) = \mathbb{E}(h(Z_i', Z_j')|\mathcal{Z})$$
$$= \frac{1}{4}\Big(h(X_i, X_j) + h(X_i, X_j') + h(X_i', X_j) + h(X_i', X_j')\Big). \tag{2.6}$$

These observations i.e., equations (2.5) and (2.6), together with the convexity and monotonicity of Φ, the integrability of the functions involved, and Jensen's inequality, justify the following two strings of inequalities which, together, prove the theorem.

1) For h symmetric in its entries,

$$\begin{aligned}
\mathbb{E}\Phi(\|\sum_{I_n^2} h(X_i,X_j')\|) &= \mathbb{E}\Phi(\frac{1}{2}\|\sum_{I_n^2} h(X_i,X_j')+\sum_{I_n^2} h(X_i',X_j)\|)\\
&\le \frac{1}{2}\mathbb{E}\Phi\Big(\|\sum_{I_n^2}[h(X_i,X_j')+h(X_i',X_j)+h(X_i,X_j)+h(X_i',X_j')]\,\|\Big)\\
&\qquad + \frac{1}{4}\mathbb{E}\Phi(2\|\sum_{I_n^2} h(X_i,X_j)\|)+\frac{1}{4}\mathbb{E}\Phi(2\|\sum_{I_n^2} h(X_i',X_j')\|)\\
&= \frac{1}{2}\mathbb{E}\Phi(4\|\sum_{I_n^2}\mathbb{E}(h(Z_i,Z_j)|\mathcal{Z})\|)+\frac{1}{2}\mathbb{E}\Phi(2\|\sum_{I_n^2} h(X_i,X_j)\|)\\
&\le \frac{1}{2}\mathbb{E}\Phi(4\|\sum_{I_n^2} h(Z_i,Z_j)\|)+\frac{1}{2}\mathbb{E}\Phi(2\|\sum_{I_n^2} h(X_i,X_j)\|)\\
&= \frac{1}{2}\mathbb{E}\Phi(4\|\sum_{I_n^2} h(X_i,X_j)\|)+\frac{1}{2}\mathbb{E}\Phi(2\|\sum_{I_n^2} h(X_i,X_j)\|)\\
&\le \mathbb{E}\Phi(4\|\sum_{I_n^2} h(X_i,X_j)\|), \qquad (2.7)
\end{aligned}$$

proving (2.3). Note that symmetry is essential for the first identity.

2) For h not necessarily symmetric, letting $\mathcal{X}=\sigma(X_i : i=1,\dots,n)$, we have

$$\begin{aligned}
&\mathbb{E}\Phi(\|\sum_{I_n^2} h(X_i,X_j)\|)\\
&\le \frac{1}{2}\mathbb{E}\Phi\Big(2\|\sum_{I_n^2}\mathbb{E}\big[h(X_i,X_j)+h(X_i',X_j)+h(X_i,X_j')+h(X_i',X_j')|\mathcal{X}\big]\,\|\Big)\\
&\qquad + \frac{1}{2}\mathbb{E}\Phi\Big(2\|\sum_{I_n^2}\mathbb{E}\big[h(X_i,X_j')+h(X_i',X_j)+h(X_i',X_j')|\mathcal{X}\big]\,\|\Big)\\
&\le \frac{1}{2}\mathbb{E}\Phi\Big(2\|\sum_{I_n^2}\big[h(X_i,X_j)+h(X_i',X_j)+h(X_i,X_j')+h(X_i',X_j')\big]\,\|\Big)\\
&\qquad + \frac{1}{6}\mathbb{E}\Phi(6\|\sum_{I_n^2}\mathbb{E}(h(X_i',X_j)|\mathcal{X})\|)+\frac{1}{6}\mathbb{E}\Phi(6\|\sum_{I_n^2}\mathbb{E}(h(X_i,X_j')|\mathcal{X})\|)\\
&\qquad\qquad + \frac{1}{6}\mathbb{E}\Phi(6\|\sum_{I_n^2} E(h(X_i',X_j')|\mathcal{X})\|)\\
&\le \frac{1}{2}\mathbb{E}\Phi(8\|\sum_{I_n^2}\mathbb{E}(h(Z_i,Z_j')|\mathcal{Z})\|)+\frac{1}{3}\mathbb{E}\Phi(6\|\sum_{I_n^2} h(X_i,X_j')\|)
\end{aligned}$$

$$+\frac{1}{6}\Phi(6\|\sum_{I_n^2}\mathbb{E}h(X_i',X_j')\|)$$

$$\le \frac{1}{2}\mathbb{E}\Phi(8\|\sum_{I_n^2}h(Z_i,Z_j')\|)+\frac{1}{3}\mathbb{E}\Phi(6\|\sum_{I_n^2}h(X_i,X_j')\|)+\frac{1}{6}\Phi(6\|\sum_{I_n^2}\mathbb{E}h(X_i',X_j')\|)$$

$$=\frac{1}{2}\mathbb{E}\Phi(8\|\sum_{I_n^2}h(X_i,X_j')\|)+\frac{1}{3}\mathbb{E}\Phi(6\|\sum_{I_n^2}h(X_i,X_j')\|)+\frac{1}{6}\Phi(6\|\sum_{I_n^2}\mathbb{E}h(X_i,X_j')\|)$$

$$\le \mathbb{E}\Phi(8\|\sum_{I_n^2}h(X_i,X_j')\|), \tag{2.8}$$

proving (2.1), even with a better constant.

□

If $h_{i_1\ldots i_m}$ are functions with values in a separable Banach space then, taking

$$\mathcal{H}_{i_1\ldots i_m}=\{f\circ h_{i_1\ldots i_m}: f\in B_1'\},$$

the sup over the $\mathcal{H}$'s in Theorem 2.1 can be replaced by the norm of the Banach space. The same comment applies to the next theorem.

It is remarkable that not only expected values of convex functions of U-statistics can be decoupled, but also tail probabilities. This is due to de la Peña and Montgomery–Smith (1995). Their result contains Theorem 2.1 modulo constants, and is as follows:

2.2. THEOREM. *With the notation of Theorem 2.1 (but without any integrability assumptions on the envelopes $H_{i_1\ldots i_m}$), there are constants $C_m\in(0,\infty)$, depending on m only, such that for all $t>0$ and $n\ge m$,*

$$\Pr{}^*\Big\{\sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}}\big|\sum_{I_n^m}h_{i_1\ldots i_m}(X_{i_1},\ldots,X_{i_m})\big|>t\Big\}$$
$$\le C_m\Pr{}^*\Big\{C_m\sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}}\big|\sum_{I_n^m}h_{i_1\ldots i_m}(X_{i_1}^1,\ldots,X_{i_m}^m)\big|>t\Big\} \tag{2.9}$$

If moreover the classes $\mathcal{H}_{i_1\ldots i_m}$ satisfy the symmetry conditions (2.2), then there are constants $D_m\in(0,\infty)$, depending on m only, such that for all $t>0$ and $n\ge m$,

$$\Pr{}^*\Big\{\sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}}\big|\sum_{I_n^m}h_{i_1\ldots i_m}(X_{i_1}^1,\ldots,X_{i_m}^m)\big|>t\Big\}$$
$$\le D_m\Pr{}^*\Big\{D_m\sup_{h_{i_1\ldots i_m}\in\mathcal{H}_{i_1\ldots i_m}}\big|\sum_{I_n^m}h_{i_1\ldots i_m}(X_{i_1},\ldots,X_{i_m})\big|>t\Big\} \tag{2.10}$$

The proof of this theorem is much more involved than that of Theorem 2.1: it requires hypercontractivity of the Rademacher polynomials in conjunction with a Paley Zygmund type argument to obtain a sort of conditional Jensen inequality for tail probabilities, hypercontractivity of linear combinations of the coordinates of a multinomial $(1;1/n,\ldots,1/n)$ random vector, and (a simpler form of) the Lévy type maximal inequality of Montgomery–Smith (1994) for sums of i.i.d. random vectors. See de la Peña and Montgomery–Smith (1995) or our forthcoming book for the proof.

Decoupling theory started with decoupling of multilinear forms in i.i.d. random variables with distributional constraints (e.g., Gaussian, stable). Theorem 2.2 provides the most general decoupling inequality for multilinear forms, up to constants, as follows. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a vector of n independent real random variables X_i, let $\mathbf{X}_j = (X_1^j, \ldots, X_n^j)$, $j = 1, \ldots, m$, be m independent copies of $\mathbf{X}$ and let

$$Q_m := Q_m(\mathbf{X}, \ldots, \mathbf{X}) = \sum_{\mathbf{i} \in I_n^m} a_{i_1 \ldots i_m} X_{i_1} \cdots X_{i_m}, \tag{2.11}$$

where the coefficients $a_{i_1 \ldots i_m}$ are elements of some Banach space. Without loss of generality we can assume the coefficients $a_{i_1 \ldots i_m}$ symmetric in their entries (otherwise, we replace them by $\sum a_{i_{s(1)} \ldots i_{s(m)}}/m!$, the sum extended over all permutations s of $\{1, \ldots, m\}$). Q_m is a *tetrahedral* m–linear form in the variables $X_1, \ldots, X_n$ (its monomials are of degree at most one in each of these variables). The *decoupled version* of Q_m is defined to be

$$Q_m^{\mathrm{dec}} := Q_n(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \sum_{\mathbf{i} \in I_n^m} a_{i_1 \ldots i_m} X_{i_1}^1 \cdots X_{i_m}^m, \tag{2.12}$$

assuming the coefficients $a_{i_1 \ldots i_m}$ are invariant under permutations of its subindices. Application of Theorem 2.2 to the functions

$$h_{i_1 \ldots i_m}(x_1, \ldots, x_m) = a_{i_1 \ldots i_m} x_{i_1} \cdots x_{i_m}$$

(more concretely to the collections $\{f(h_{i_1 \ldots i_m}) : f \in B_1'\}$, where B_1' is the unit ball of the dual of B), immediately gives that the tail probabilities of the norms of Q_m and Q_m^{dec} are comparable. Actually, with a little extra care, this extends to not necessarily homogenous polynomials (Giné, 1997) as follows:

2.3. COROLLARY. *There exist constants $C_m \in (0, \infty)$ depending only on m such that if $Q_{(m)}$ is a tetrahedral polynomial of degree m in any set of n independent random variables $\{X_i\}_{i=1}^n$, $n > m$, with coefficients in any Banach space,*

$$Q_{(m)} = \sum_{k=0}^{m} \sum_{\mathbf{i} \in I_n^k} a_{i_1 \ldots i_k} X_{i_1} \cdots X_{i_k}$$

(with $I_n^0 = \{0\}$), and if $Q_{(m)}^{dec}$ is its decoupled version, defined as

$$Q_{(m)}^{dec} = \sum_{k=0}^{m} \frac{(m-k)!}{m!} \sum_{\mathbf{i} \in I_n^k} \sum_{\mathbf{r} \in I_m^k} a_{i_1 \ldots i_k} X_{i_1}^{r_1} \cdots X_{i_k}^{r_k},$$

where $\{X_i^j : i = 1, \ldots, n\}$, $j = 1, \ldots, m$, are m independent copies of $\{X_i\}_{i=1}^n$, then

$$\frac{1}{C_m} \Pr\{\|Q_{(m)}^{dec}\| > C_m t\} \le \Pr\{\|Q_{(m)}\| > t\} \le C_m \Pr\{C_m \|Q_{(m)}^{dec}\| > t\}.$$

This result should not be considered new since it is a trivial consequence of Theorem 2.2, but it is formally new in the sense that previously published versions of it require the variables X_i to be symmetric and the polynomials to be homogeneous (Kwapień and Woyczynski, 1992; de la Peña, Montgomery–Smith and Szulga, 1994), or the variables to be symmetric and to have expected values of convex functions instead of tail probabilities in the inequalities (Kwapień, 1987).

Neither Theorem 2.2 nor Corollary 2.3 will be used in the sequel.

b) Randomization of convex functions. What interests us about decoupling is the possibility of randomizing a degenerate U–process (or a degenerate U–statistic). In order to be more concrete, we will have to define the degree of degeneracy of a U–statistic and also recall Hoeffding's decomposition.

As usual, we let $(S, \mathcal{S})$ be a measurable space and P a probability measure on it, and let X_i, $X_i^{(j)}$ ε_i, $\varepsilon_i^{(j)}$ be the coordinate functions on the product of countably many copies of $(S, \mathcal{S}, \mathrm{P})$ and countably many copies of $(\{-1,1\}, (\delta_1 + \delta_{-1})/2)$. In particular these variables are all independent, the X's have law P, and the ε's are Rademacher variables.

2.4. DEFINITION. *A P^m–integrable symmetric function of m variables, $h : S^m \to \mathbb{R}$, is* P *-degenerate of order $r-1$, $1 < r \le m$, if*

$$\int h(x_1, \ldots, x_m) d\mathrm{P}^{m-r+1}(x_r, \ldots, x_m) = \int h d\mathrm{P}^m \quad \text{for all } x_1, \ldots, x_{r-1} \in S$$

whereas

$$\int h(x_1, \ldots, x_m) d\mathrm{P}^{m-r}(x_{r+1}, \ldots, x_m)$$

is not a constant function. If h is P^m–centered and is P*–degenerate of order $m-1$, that is, if*

$$\int h(x_1, \ldots, x_m) d\mathrm{P}(x_1) = 0 \quad \text{for all } x_2, \ldots, x_m \in S,$$

then h is said to be canonical or completely degenerate with respect to P. *If h is not degenerate of any positive order we say it is nondegenerate or degenerate of order zero.*

In this definition the identities are usually taken in the almost everywhere sense, however, when dealing with uncountable families of functions (and only then), we need them to hold pointwise.

With the notation $\mathrm{P}_1 \times \cdots \times \mathrm{P}_m h = \int h d(\mathrm{P}_1 \times \cdots \times \mathrm{P}_m)$, the *Hoeffding projections* of $h : S^m \to \mathbb{R}$ symmetric are defined as

$$\pi_k h(x_1, \ldots, x_k) := \pi_{k,m}^{\mathrm{P}} h(x_1, \ldots, x_k) := (\delta_{x_1} - \mathrm{P}) \times \cdots \times (\delta_{x_k} - \mathrm{P}) \times \mathrm{P}^{m-k} h$$

for $x_i \in S$ and $0 \le k \le m$. Note that $\pi_0 h = \mathrm{P}^m h$ and that, for $k > 0$, $\pi_k h$ is a completely degenerate function of k variables. For h integrable these projections induce a decomposition of the U–statistic

$$U_n(h) := U_n^{(m)}(h) := U_n^{(m)}(h, \mathrm{P}) := \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m})$$

into a sum of U–statistics of orders $k \leq m$ which are orthogonal if $\mathrm{P}^m h^2 < \infty$ and whose kernels are completely degenerate, namely, the *Hoeffding decomposition*:

$$U_n^{(m)}(h) = \sum_{k=0}^{m} \binom{m}{k} U_n^{(k)}(\pi_k h) \tag{2.13}$$

(here the superindex P and the subindex m of $\pi_{k,m}^{\mathrm{P}}$ are not displayed; they will be dropped whenever no confusion is possible). This decomposition follows easily by expanding

$$h(x_1, \ldots, x_m) = \delta_{x_1} \times \cdots \times \delta_{x_m} h = \big((\delta_{x_1} - \mathrm{P}) + \mathrm{P}\big) \times \cdots \times \big((\delta_{x_m} - \mathrm{P}) + \mathrm{P}\big) h$$

into terms of the form $(\delta_{x_{i_1}} - \mathrm{P}) \times \cdots \times (\delta_{x_{i_k}} - \mathrm{P}) \times \mathrm{P}^{m-k} h$. It is very simple to check that h symmetric is P–degenerate of order $r-1$ iff $r = \min\{k > 0 : \pi_{k,m}^{\mathrm{P}} h \not\equiv 0\}$. Therefore, h is degenerate of order $r - 1 \geq 0$ iff its Hoeffding expansion, except for the constant term, starts at term r, that is,

$$U_n(h) - \mathrm{P}^m h = \sum_{k=r}^{m} \binom{m}{k} U_n^{(k)}(\pi_k h). \tag{2.14}$$

Hoeffding's decomposition is a basic tool in the analysis of U–statistics.

We are interested in the behavior of $\|U_n(h) - \mathrm{P}^m h\|_{\mathcal{H}} := \sup_{h \in \mathcal{H}} |U_n(h) - \mathrm{P}^m h|$ for possibly uncountable families $\mathcal{H}$ of symmetric functions $h : S^m \to \mathbb{R}$. Whereas the *measurability* requirements for decoupling are minimal, randomization requires (or at least would not be useful without) the possibility of using Fubini's theorem on expressions of the form $\sup_{h \in \mathcal{H}} |\sum \varepsilon_{i_1} \cdots \varepsilon_{i_k} h(X_{i_1}, \ldots, X_{i_m})|$, whose integrals one wants to compute by first integrating over the ε's and then over the X's or viceversa. In particular these expressions should be measurable. If the class $\mathcal{H}$ of measurable functions is countable there are no measurability problems. A quite general situation for which one can work without measurability problems, as if the class were countable, is when $\mathcal{H}$ is *image admissible Suslin* that is, when there is a map from a Polish space Y onto $\mathcal{H}$, T, such that the composition of T and the evaluation map, $(y, x_1, \ldots, x_m) \to T(y)(x_1, \ldots, x_m)$, is jointly measurable (Dudley, 1984). Often the classes of functions of interest are parametrized by G_δ subsets $\Theta \subset \mathbb{R}^d$ and the evaluation map is jointly measurable in the arguments and the parameter, thus the usefulness of the image admissible Suslin concept. If $\mathcal{H}$ is image admissible Suslin, so are the classes $\{\pi_k h : h \in \mathcal{H}\}$ (e.g., Arcones and Giné, 1993). For simplicity, image admissible Suslin classes of functions will simply be denoted as *measurable* classes.

Also, we will assume that all the classes of functions $\mathcal{H}$ considered in this subsection admit everywhere finite measurable envelopes H.

Notation: The symbol $\simeq$ between two expressions means two sided inequality up to multiplicative constants that depend only on the order m of the U–process and on the exponent p. Likewise, the symbols $\lesssim$ and $\gtrsim$ are used for one sided inequalities up to multiplicative constants.

The following lemma for functions of one variable is well known and easy to prove:

2.5. LEMMA. *Let $\mathcal{H}$ be a measurable class of P–centered functions $h : S \to \mathbb{R}$ such that, for some $p \geq 1$, the envelope H of the class satisfies* $\mathrm{P}H^p < \infty$. *Then, for all* $n < \infty$,

$$2^{-p}\mathbb{E}\|\sum_{i=1}^{n}\varepsilon_i h(X_i)\|_{\mathcal{H}}^p \leq \mathbb{E}\|\sum_{i=1}^{n} h(X_i)\|_{\mathcal{H}}^p \leq 2^p\mathbb{E}\|\sum_{i=1}^{n}\varepsilon_i h(X_i)\|_{\mathcal{H}}^p$$

The *randomization theorem* to be stated immediately below can be considered as an extension of this lemma where the P–centering hypothesis on $h(x)$ is replaced by a P–degeneracy hypothesis on $h(x_1, \ldots, x_m)$. This theorem is stated in full generality although only the cases $r = m$ (the completely degenerate case), $r = 1$ and $r = 2$ are used below. The proof for $r = m$ is a straightforward consequence of the decoupling theorems (Theorem 2.1) and of Lemma 2.5 above. The proof for general r is equally easy but more complicated (see our forthcoming book).

2.6. THEOREM. *For $1 \leq r \leq m$ and $p \geq 1$, let $\mathcal{H}$ be a measurable class of real functions defined on S^m consisting of* P*–centered,* P*–degenerate functions of order at least $r - 1$ such that* $\mathrm{P}^m H^p < \infty$. *Then,*

$$\begin{aligned}
\mathbb{E}\| \sum_{(i_1,\ldots,i_m)\in I_n^m} h(X_{i_1},\ldots,X_{i_m})\|_{\mathcal{H}}^p& \\
&\simeq \mathbb{E}\|\sum_{(i_1,\ldots,i_m)\in I_n^m} \varepsilon_{i_1}\cdots\varepsilon_{i_r} h(X_{i_1},\ldots,X_{i_m})\|_{\mathcal{H}}^p \\
&\simeq \mathbb{E}\|\sum_{(i_1,\ldots,i_m)\in I_n^m} \varepsilon_{i_1}^1\cdots\varepsilon_{i_r}^r h(X_{i_1},\ldots,X_{i_m})\|_{\mathcal{H}}^p \\
&\simeq \mathbb{E}\|\sum_{(i_1,\ldots,i_m)\in I_n^m} \varepsilon_{i_1}^1\cdots\varepsilon_{i_r}^r h(X_{i_1}^1,\ldots,X_{i_m}^m)\|_{\mathcal{H}}^p. \qquad (2.15)
\end{aligned}$$

PROOF for $r = m$. Let us use, for simplicity of notation, the abbreviations $\mathbf{i}$ for the multiindex $(i_1, \ldots, i_m)$, $\mathbf{X}_\mathbf{i}$ for the vector $(X_{i_1}, \ldots, X_{i_m})$, $\mathbf{X}_\mathbf{i}^{\mathrm{dec}}$ for $(X_{i_1}^1, \ldots, X_{i_m}^m)$, $\varepsilon_\mathbf{i}$ for the product $\varepsilon_{i_1}\cdots\varepsilon_{i_m}$ and $\varepsilon_\mathbf{i}^{\mathrm{dec}}$ for the product $\varepsilon_{i_1}^1\cdots\varepsilon_{i_m}^m$. Since in the present case (h canonical) we have $\mathbb{E}h(X_1, x_2, \ldots, x_m) = 0$ for all $x_2, \ldots, x_m \in S$, letting $\mathbb{E}_r$ denote integration with respect to only the variables ε^r, X^r, Lemma 2.5 gives

$$\begin{aligned}
\mathbb{E}\|\sum_{\mathbf{i}\in I_n^m} h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\|_{\mathcal{H}}^p &= \mathbb{E}\mathbb{E}_1\|\sum_{i_1=1}^{n}\Big(\sum_{(i_2,\ldots,i_m):\mathbf{i}\in I_n^m} h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\Big)\|_{\mathcal{H}}^p \\
&\simeq \mathbb{E}\mathbb{E}_1\|\sum_{\mathbf{i}\in I_n^m}\varepsilon_{i_1}^1 h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\|_{\mathcal{H}}^p \\
&= \mathbb{E}\mathbb{E}_2\|\sum_{i_2=1}^{n}\Big(\sum_{(i_1,i_3,\ldots,i_m):\mathbf{i}\in I_n^m}\varepsilon_{i_1}^1 h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\Big)\|_{\mathcal{H}}^p \\
&\simeq \mathbb{E}\mathbb{E}_2\|\sum_{\mathbf{i}\in I_n^m}\varepsilon_{i_1}^1\varepsilon_{i_2}^2 h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\|_{\mathcal{H}}^p \\
&\simeq \cdots \simeq \mathbb{E}\|\sum_{\mathbf{i}\in I_n^m}\varepsilon_\mathbf{i}^{\mathrm{dec}} h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\|_{\mathcal{H}}^p.
\end{aligned}$$

Now the equivalences (2.15) for $r = m$ follow by several applications of the decoupling Theorem 2.1.

□

c) Randomization of tail probabilities. Let F be a vector space, let the function $h : S^m \to F$ be symmetric in its entries and let $I = \{(i_1, \ldots, i_m) \in \mathbb{N}^m : i_j \neq i_k \text{ for all } j \neq k\}$. For finite sets $A \subset \mathbb{N}$ we set

$$S_A = \sum_{\mathbf{i} \in I \cap A^m} h(\mathbf{x_i}).$$

The following elementary lemma (basically an inclusion–exclusion principle) provides decoupling and randomization of tail probabilities.

2.7. LEMMA. *Let A_i, $i = 0, \ldots, m$, be $m+1$ finite disjoint sets of integers, $A_i \neq \emptyset$ if $i \neq 0$, and let $A = \cup_{i=0}^m A_i$. Then,*

$$m! \sum_{\mathbf{i} \in A_1 \times \cdots \times A_m} h(\mathbf{x_i}) = (-1)^m S_{A_0} + \sum_{k=1}^{m} (-1)^{m-k} \sum_{1 \le i_1 < \ldots < i_k \le m} S_{A_0 \cup A_{i_1} \cup \ldots \cup A_{i_k}}. \tag{2.16}$$

This lemma for $A_0 = \emptyset$ was observed by Giné and Zinn (1994) and for general A_0 by Zhang (1996). See these references (or our forthcoming book) for its proof. An almost immediate consequence of it is the following *one sided* decoupling and randomization inequality for tail probabilities of U-processes (Giné and Zinn, 1994).

2.8. THEOREM. *Let $\mathcal{H}$ be a measurable class of real functions on S^m, symmetric in their entries. Then,*
(a) For natural numbers $n_0 < n$, if D_j are subsets of $\{n_0 + 1, \ldots, n\}$, $j = 1, \ldots, m$, and $M = n_0 + \sum_{j=1}^m |D_j|$, then, for all $t > 0$,

$$\Pr\Big\{\Big\| \sum_{\mathbf{i} \in D_1 \times \cdots \times D_m} h(X_{i_1}^1, \ldots, X_{i_m}^m) \Big\|_{\mathcal{H}} \ge \frac{2^m t}{m!}\Big\}$$
$$\le 2^m \max_{n_0 < k \le M} \Pr\Big\{\Big\| \sum_{\mathbf{i} \in I_k^m} h(X_{i_1}, \ldots, X_{i_m}) \Big\|_{\mathcal{H}} \ge t\Big\}; \tag{2.17}$$

(b) for natural numbers $n_0 < n$ and all $t > 0$,

$$\Pr\Big\{\Big\| \sum_{n_0 < i_1, \ldots, i_m \le n} \varepsilon_{i_1}^1 \cdots \varepsilon_{i_m}^m h(X_{i_1}^1, \ldots, X_{i_m}^m) \Big\|_{\mathcal{H}} \ge \frac{2^{2m} t}{m!}\Big\}$$
$$\le 2^{2m} \max_{n_0 < k \le nm} \Pr\Big\{\Big\| \sum_{\mathbf{i} \in I_k^m} h(X_{i_1}, \ldots, X_{i_m}) \Big\|_{\mathcal{H}} \ge t\Big\}. \tag{2.18}$$

PROOF assuming Lemma 2.7. Inequality (2.17) is trivially true if $D_j = \emptyset$ for some j. So, assuming that D_j is not empty for any j, we take $A_0 = \{1, \ldots, n_0\}$, $A_j =$

$\{1+n_0+\sum_{i=1}^{j-1}|D_i|,\dots,n_0+\sum_{i=1}^{j}|D_i|\}$, $j=1,\dots,m$, which are disjoint, and note that, by permutation of the factors in the infinite product of $(S,\mathcal{S},P)$,

$$\Pr\Big\{\|\sum_{\mathbf{i}\in D_1\times\cdots\times D_m} h(X_{i_1}^1,\dots,X_{i_m}^m)\|\geq\frac{2^m t}{m!}\Big\}$$
$$=\Pr\Big\{\|\sum_{\mathbf{i}\in A_1\times\cdots\times A_m} h(X_{i_1},\dots,X_{i_m})\|\geq\frac{2^m t}{m!}\Big\}.$$

Then, part (a) follows by direct application of Lemma 2.7.

Part (b) follows from part (a) and Fubini's theorem because

$$\sum_{n_0\leq i_1,\dots,i_m\leq n}\varepsilon_{i_1}^1\cdots\varepsilon_{i_m}^m h(X_{i_1}^1,\dots,X_{i_m}^m)$$

is a linear combination with coefficients ±1 of 2^m terms of the form

$$\sum_{\mathbf{i}\in D_1\times\cdots\times D_m} f(X_{i_1}^1,\dots,X_{i_m}^m)$$

with $D_j=\{n_0<i\leq n:\varepsilon_i^j=1\}$ or $D_j=\{n_0<i\leq n:\varepsilon_i^j=-1\}$.

□

It should be noted that *there is no converse to inequality (2.18)* in general, even for $m=1$. For instance, if X is such that $\Pr\{X>t\}\simeq c_1t^{-1}(\log t)^{-1}(\log\log t)^{-2}$ and $\Pr\{X<-t\}\simeq c_2t^{-1}(\log t)^{-1}(\log\log t)^{-2}$ as $t\to\infty$ and $c_1\neq c_2$ (and one can find c_i's such that this random variable is even centered), then $\sum_{i=1}^n X_i=O_P\big(n(\log n\log\log n)^{-1}\big)$ whereas $\sum_{i=1}^n\varepsilon_iX_i=O_P\big(n(\log n)^{-1}(\log\log n)^{-2}\big)$. To see this just note that X is in the domain of atraction of a 1–stable law with centerings that upset the normings, and εX is in the domain of attraction of a 1–stable law with centerings equal to zero and with the same normings (see, e.g., Giné, Mason and Götze, 1997).

Theorem 2.8 is useful for proving 'converse limit theorems', that is, for deducing integrability properties of h under the assumption that the U–statistic with kernel h satisfies a limit theorem such as the clt or the lil.

It does not seem that Theorem 2.8 follows from decoupling of tail probabilities (Theorem 2.2); at any rate, Theorem 2.8 is much more elementary than Theorem 2.2.

3. Limit theorems for U–statistics. If h is integrable, then the U–statistics (1.1) based on the kernel h form a reverse martingale, hence, they converge a.s. and in L_1; the limit is a constant by the Hewitt–Savage zero–one law and this constant is necessarily $\mathbb{E}h(X_1,\dots,X_m)$ by L_1 convergence. The law of large numbers was first proved by Hoeffding (1948), but this slick argument belongs to Berk (1966). Giné and Zinn (1992) gave an example of a U–statistic with a kernel not in L_1 that converges a.s. and the question of finding necessary and sufficient conditions on h for the U–stastistics $U_n(h)$ to converge (possibly after centering) a.s. or in probability to a constant, is open. For the case $h(x_1,\dots,x_m)=x_1\cdots x_m$ see Cuzick, Giné and Zinn (1995) and Zhang (1996). Recent developments on the exact estimation of moments of U–statistics (Klass and Nowicki, 1996) allow for some optimism, but it is too early to tell.

Similar comments apply to the law of the iterated logarithm, except that it was not known until very recently (Arcones and Giné, 1995) that finiteness of the second moment of the kernel implies the lil for the corresponding U–statistic in the completely degenerate case and for all m. The proof of this result does rely heavily on decoupling (Theorem 2.1). Here again, h being in L_2 is not a necesary condition for the lil in the canonincal case (Giné and Zhang, 1996), and necessary and sufficient conditions are not known.

On the other hand, the clt is completely solved. Sufficiency of square integrability of the kernel for a completely degenerate U–statistic of order m (for any m) to satisfy the clt was proved by Rubin and Vitale (1980) and necessity by Giné and Zinn (1994). Decoupling (Theorem 2.8) plays a basic role in the proof of necessity.

Only the clt and the lil will be described here. As a consequence of Hoeffding's decomposition (2.13), (2.14), it is clear that, at least under some integrability for h, the clt (resp. the lil) for the completely degenerate or canonical case give the clt (resp. the lil) in general. So, only canonical kernels will be considered.

a) The central limit theorem. Let X_i be i.i.d. centered random variables, with finite second moment equal to 1. Then, the clt and the lln for sums of i.i.d. random variables gives

$$\frac{1}{n}\sum_{(i,j)\in I_n^2} X_i X_j = \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\Big)^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_d g^2 - 1,$$

where g is $N(0,1)$. This is the clt for the U–statistics with kernel $h(x,y) = xy$, which is degenerate if $\mathbb{E}X_1 = 0$. This simple example is very appropriate because canonical kernels are just limits in L_2 of linear combinations of products $\phi(x_1)\cdots\phi(x_m)$, ϕ P–centered. Extrapolating, the example suggests that a canonical U–statistic of m variables, multiplied by $n^{m/2}$, should converge in law to an element of a Gaussian chaos of order m. This is the content of the direct clt for canonical U–statistics, which we now describe for completeness and also for use in the next section.

Let $L_2^c(S,\mathcal{S},P)$ be the space of real valued P–centered, P–square integrable functions on S. Let G_P be an isonormal Gaussian process on $L_2^c(S,\mathcal{S},P)$, that is, a centered Gaussian process with parameter set $L_2^c(S,\mathcal{S},P)$ such that $\mathbb{E}G_P(f)G_P(g)$ $= \int fg dP$. If $\{\phi_i\}_{i\in I}$ is an orthonormal basis of $L_2^c(S,\mathcal{S},P)$ and if $\{g_i\}_{i\in I}$ is a family of independent $N(0,1)$ random variables, then the equation

$$G_P\Big(\sum_{i\in I} a_i\phi_i\Big) = \sum_{i\in I} a_i g_i, \quad \sum_{i\in I} a_i^2 < \infty,$$

produces such a process. By identifying random variables which are a.s. equal, G_P becomes a linear isometry from $L_2^c(S,\mathcal{S},P)$ onto the Hilbert space of jointly normal random variables generated by $\{G_P(f)\}$ (or, isomorphically, by the g_i's). Then, the finite dimensional central limit theorem simply asserts that the finite dimensional distributions of the processes $\big\{\frac{1}{n^{1/2}}\sum_{i=1}^{n} f(X_i) : f \in L_2^c(S,\mathcal{S},P)\big\}$, converge in law to the finite dimensional distributions of $\{G_P(f) : f \in L_2^c(S,\mathcal{S},P)\}$, that is, for every finite set of functions $f_1,\dots,f_k$ in $L_2^c(S,\mathcal{S},P)$,

$$\Big(\frac{1}{n^{\frac12}}\sum_{i=1}^n f_1(X_i),\dots,\frac{1}{n^{\frac12}}\sum_{i=1}^n f_k(X_i)\Big) \to_{\mathcal{L}} \Big(G_P(f_1),\dots,G_P(f_k)\Big), \tag{3.1}$$

with convergence of up to second moments (of any norm) as well. The central limit theorem for canonical U–statistics may be viewed as the extension of the isometry G_P to an isometry K_P from the Hilbert space of all P–canonical square integrable kernels onto a Gaussian chaos Hilbert space (precisely, the Gaussian chaos corresponding to G_P) in such a way that the finite dimensional distributions of properly normalized U–statistics converge to the corresponding finite dimensional distributions of the process K_P. (I learned this way of seeing the clt for U–statistics from Bretagnolle, 1983.)

Let $L_2^{c,k}(S,\mathcal{S},P)$ ($L_2^{c,k}(P)$ for short) denote the Hilbert space of P–canonical functions of k variables. It follows easily from basic Hilbert space and measure theory that if $\{\phi_i\}_{i\in I}$ is an orthonormal basis for $L_2^c(S,\mathcal{S},P)$, then the following set of functions is an orthonormal basis for $L_2^{c,k}(S,\mathcal{S},P)$:

$$\Big\{\frac{1}{\binom{k}{r_j:j\in I}^{\frac12}}\sum_{\mathbf{i}:j(\mathbf{i})=r_j}\phi_{i_1}(x_1)\cdots\phi_{i_k}(x_k) : \mathbf{i}=(i_1,\dots,i_k)\in I^k\Big\}, \tag{3.2}$$

where $I^k := \{(i_1,\dots,i_k) : i_r \in J^k,\ i_r \neq i_s \text{ if } r\neq s,\ r,s=1,\dots,k\}$, for any $j\in I$ and $\mathbf{i}\in I^k$, $j(\mathbf{i}) = \sum_{\ell=1}^k I_{i_\ell=j}$ is the number of occurrences of j in the multiindex $\mathbf{i}=(i_1,\dots,i_m)$, and $\binom{k}{r_j:j\in I}$ denotes the combinatorial number $\binom{k}{m_1,\dots,m_n}$ if $\{r_j : j\in I\} = \{m_1,\dots,m_n\}$. So, if h is a P–canonical kernel of k variables, then

$$h(x_1,\dots,x_k) = \sum_{\mathbf{i}\in I^k} a_{\mathbf{i}}\phi_{i_1}(x_1)\cdots\phi_{i_k}(x_k), \tag{3.3}$$

in the L_2 sense, with coefficients

$$a_{\mathbf{i}} := a_{i_1\dots i_k} = \mathbb{E}\big[h(x_1,\dots,x_k)\prod_{r=1}^k \phi_{i_r}(x_r)\big],$$

which are *symmetric* in their indices. Given a version of G_{P}, a version of K_{P} can be constructed as follows: K_{P} is linear and

$$K_P\Big(\frac{1}{\binom{k}{r_j:j\in I}^{\frac12}}\sum_{\mathbf{i}:j(\mathbf{i})=r_j}\phi_{i_1}(x_1)\cdots\phi_{i_k}(x_k)\Big) = \prod_{j\in I}\frac{1}{\sqrt{r_j!}}H_{r_j}\big(G_P(\phi_j)\big), \tag{3.4}$$

where H_r is the Hermite polynomial of degree k and leading coefficient 1 [concretely, H_k is defined by the relation $\exp(ux-u^2/2) = \sum_{k=0}^\infty H_k(x)u^k/k!$]. Therefore, if h has the expansion (3.3), then

$$K_P(h) = \frac{1}{\sqrt{k!}}\sum_{\mathbf{i}\in I^k} a_{\mathbf{i}}\prod_{j\in I} H_{j(\mathbf{i})}\big(G_P(\phi_j)\big). \tag{3.5}$$

We call K_{P} the *isonormal Gaussian chaos process* associated to the Gaussian process G_{P} (and will shortly explain why). Then, the Rubin and Vitale (1980) central limit theorem can be stated as follows:

3.1. THEOREM. *For arbitrary natural numbers r_ℓ, $1 \le \ell \le k < \infty$, let h_ℓ be P–square integrable, P–canonical kernels in r_ℓ variables (that is, $h_\ell \in L_2^{c,r_\ell}(P)$), and let K_P be an isonormal Gaussian chaos process on $\oplus_{r=1}^{\infty} L_2^{c,r}(P)$. Then,*

$$\left(\binom{n}{r_1}^{\frac{1}{2}} U_n(h_1), \dots, \binom{n}{r_k}^{\frac{1}{2}} U_n(h_k)\right) \to_{\mathcal{L}} \left(K_P(h_1), \dots, K_P(h_k)\right) \tag{3.6}$$

as $n \to \infty$, with convergence of up to second moments of the norm.

In fact, this limit theorem admits an extension to finite numbers of functions in $\oplus_{k=1}^{\infty} L_2^{c,k}(\mathrm{P})$. For a single function $h \in \oplus_{k=1}^{\infty} L_2^{c,k}(\mathrm{P})$ the result is that, if $h = \sum_{k=1}^{\infty} h_k$ with $h_k \in L_2^{c,k}(P)$, then

$$\sum_{k=1}^{\infty} \binom{n}{k}^{\frac{1}{2}} U_n(h_k) \to_{\mathcal{L}} K_P(h)$$

(Dynkin and Mandelbaum, 1983).

Let $(\Omega, \Sigma, \Pr)$ be the probability space where the isonormal process G_P is defined, and let $\sigma(G_P)$ be the sub–σ–algebra of Σ generated by the random variables $\{G_P(\psi) : \psi \in L_2^c(P)\}$. Then $L_2(G_P) := L_2(\Omega, \sigma(G_P), \Pr)$ is the Hilbert space of square integrable G_{P} measurable functions. Let $\mathcal{P}_k(G_P)$ ($\mathcal{P}_k$ for short) be the Hilbert subspace of $L_2(G_P)$ generated by the polynomials of degree at most k in the variables $G_P(\psi)$, $\psi \in L_2^c(P)$, and let $\mathcal{H}_k(G_P)$ ($\mathcal{H}_k$ for short) be the orthogonal complement of $\mathcal{P}_{k-1}$ in $\mathcal{P}_k$, that is,

$$\mathcal{H}_k = \mathcal{P}_k \ominus \mathcal{P}_{k-1}.$$

It turns out that K_{P}, extended as the identity on constants, is an isometry from the Hilbert subspace of $L_2(S^{\mathbb{N}}, \mathrm{P}^{\mathbb{N}})$ generated by the constants and the canonical kernels of all orders, $\mathbb{R} \oplus \left(\oplus_{k=1}^{\infty} L_2^{c,k}(P)\right)$ (note that all these spaces are orthogonal in $L_2(S^{\mathbb{N}}, \mathrm{P}^{\mathbb{N}})$), onto $L_2(G_{\mathrm{P}}) = \oplus_{k=0}^{\infty} \mathcal{H}_k(\mathrm{P})$, such that

$$K_P(L_2^{c,k}(P)) = \mathcal{H}_k(P), \quad k \in \mathbb{N}, \tag{3.7}$$

In other words, the orthogonal decomposition into canonical kernels of different orders induces, via K_{P}, the chaos decomposition of $L_2(G_{\mathrm{P}})$. This justifies the name given to the process K_{P}. We note that it is possible to *simultaneously* prove the clt for U–statistics and the chaos decomposition of $L_2(G_{\mathrm{P}})$, quite economically. See our forthcomming book for details. For a similar abbreviated account of the same theory see Bretagnolle (1983). Dynkin and Mandelbaum (1983) contains another derivation of the same facts.Theorem 3.1 was first proved for $m = 2$ by Serfling (1980) and Gregory (1977).

Theorem 3.1 has the following converse (Giné and Zinn, 1994):

3.2. THEOREM. *Let $h : S^k \to \mathbb{R}$ be a measurable symmetric function on $(S, \mathcal{S})$ and let X, X_i, $i \in \mathbb{N}$, be i.i.d. S–valued random variables with probability law P. If the sequence of random variables*

$$\Big\{\frac{1}{n^{k/2}} \sum_{I_n^k} h(X_{i_1}, \dots, X_{i_k})\Big\}_{n=k}^{\infty} \tag{3.8}$$

is stochastically bounded, then $Eh^2(X_1, \dots, X_k) < \infty$ and, moreover, h is P–canonical.

Here is a sketch of the proof. By Theorem 2.8 on decoupling and randomization, stochastic boundedness of the sequence (3.8) implies stochastic boundedness of the sequence of decoupled and randomized U–statistics

$$\Big\{\frac{1}{n^{k/2}} \sum_{1 \le i_1, \dots, i_k \le n} \varepsilon_{i_1}^1 \cdots \varepsilon_{i_k}^k h(X_{i_1}^1, \dots, X_{i_k}^k)\Big\}_{n=k}^{\infty}.$$

It then follows from this and properties of Rademacher multilinear forms that the sequence

$$\Big\{\frac{1}{n^k} \sum_{1 \le i_1, \dots, i_k \le n} h^2(X_{i_1}^1, \dots, X_{i_k}^k)\Big\}$$

is also stochastically bounded. But then, by positivity, so is the family of variables

$$\Big\{\frac{1}{n^k} \sum_{1 \le i_1, \dots, i_k \le n} (h^2 I_{h^2 \le c})(X_{i_1}^1, \dots, X_{i_k}^k) : n \in \mathbb{N}, c > 0\Big\}.$$

Now, this and the law of large numbers for U–statistics applied to the bounded kernels $h^2 I_{h^2 \le c}$ imply that the numbers $\mathbb{E}\big[(h^2 I_{h^2 \le c})(X_1, \dots, X_k)\big]$ are bounded uniformly in c, hence, that $\mathbb{E}h^2(X_1, \dots, X_k) < \infty$. This, the direct clt and Hoeffding's decomposition yield that h is P–canonical. [The property of Rademacher multilinear forms used here is that their fourth moment is dominated by a universal constant times the square of their second moment, which is elementary, in fact very easy to check in the decoupled case; this then allows use of the Paley–Zygmund argument (Kahane, 1968, page 6), conditionally on the X's, to obtain tightness of the sums of squares.]

We complete this section with the observation that Theorem 3.1, the central limit theorem for U–statistics in several dimensions, can be used in conjunction with Theorem 2.2, decoupling of tail probabilities for U–statistics, to produce a comparison theorem for tail probabilities of Gaussian polynomials and their decoupled versions. The result is as follows. For ease of notation we set $|\mathbf{i}| := \max_\ell \mathbf{i}_\ell$.

Given a sequence $\{g_i : i \in \mathbb{N}\}$ of i.i.d. $N(0,1)$ random variables and a polynomial $Q_{(m)}$ of degree m in the variables g_i, and with coefficients in a Banach space B, with expansion

$$Q_{(m)} = \sum_{k=0}^{m} \sum_{\max_{r \le k} |i_r| \le N} a_{i_1, \dots, i_k} \prod_{j \in \mathbb{N}} H_{j(i_1, \dots, i_k)}(g_j),$$

where the coefficients a_{i} are symmetric in their indices (which we can assume without any loss of generality), its decoupled version is defined as

$$Q_{(m)}^{dec} = \sum_{k=0}^{m} (m-k)! \sum_{\max_{r\le k}|i_r|\le N} a_{i_1,\dots,i_k} \sum_{\mathrm{j}\in I_m^k} g_{i_1}^{(j_1)} \cdots g_{i_k}^{(j_k)}$$

by Arcones and Giné (1993a), where $\{g_i^{(j)} : i \in \mathbb{N}\}$, $j = 1,\cdots,m$, are m independent copies of the sequence $\{g_i\}$. With this definition, we have:

3.3. THEOREM. *For each $m \in \mathbb{N}$ there exists $C_m \in (0,\infty)$ such that, if B is a Banach space, $Q_{(m)}$ is a Gaussian polynomial of degree m in an orthogaussian sequence $\{g_i\}$, with coefficients in B, and $Q_{(m)}^{dec}$ is its decoupled version, then*

$$\frac{1}{C_m} \Pr\Big\{\|Q_{(m)}^{dec}\| > C_m t\Big\} \le \Pr\Big\{\|Q_{(m)}\| > t\Big\} \le C_m \Pr\Big\{\|Q_{(m)}^{dec}\| > \frac{t}{C_m}\Big\}.$$

Since the constant C_m is independent of N, the theorem extends to the whole Gaussian chaos of order m (for each m). This is a generalization (Arcones and Giné, 1993, and Giné, 1997) of a theorem of Kwapień (1987) for homogeneous polynomials in $\{g_i\}$ of degree at most one in each g_i.

The proof consists in observing that $Q_{(m)}$ is the limit in law of a U-statistic with values in the finite dimensional space generated by the (finite number of) coefficients a_{i}, and that $Q_{(m)}^{dec}$ is the limit in law of the corresponding decoupled U-statistics, so that the theorem follows by taking limits in the inequality of Theorem 2.2. [This simple proof would not be possible without Theorem 2.2; Kwapień (1987) developed very effective and elegant tools to prove the version of Theorem 3.3 for homogeneous tetrahedral polynomials and the version of Corollary 2.3 above for expected values of convex functions and symmetric variables, and some of these tools made their way into the proof of Theorem 2.2.]

b) The law of the iterated logarithm. As with the clt, in order to guess the natural norming in the lil for degenerate U-statistics, it is instructive to begin with the simplest example, namely the kernel $h(x,y) = xy$ and random variables X_i i.i.d. with $EX_i = 0$ and $EX_i^2 = 1$. Then, the Hartman-Wintner lil for sums of i.i.d. square integrable random variables and the law of large numbers readily show that

$$\begin{aligned}
&\limsup_n \frac{1}{2n\log\log n} \sum_{I_n^2} X_i X_j \\
&\quad = \limsup_n \Big[\Big(\frac{1}{(2n\log\log n)^{\frac{1}{2}}} \sum_{i=1}^n X_i\Big)^2 - \frac{1}{2n\log\log n}\sum_{i=1}^n X_i^2\Big] = 1 \quad \text{a.s.}
\end{aligned}$$

In fact, by Strassen's lil (e.g. Ledoux and Talagrand, 1991, page 206), for almost every ω, the set of limit points of the sequence $\big\{\sum_{I_n^2} X_i(\omega)X_j(\omega)/2n\log\log n\big\}_{n=1}^{\infty}$ is precisely the interval $[0,1]$. This is just a particular case of a more general statement: the kernel xy is replaced by a general square integrable P-canonical kernel in m variables, the norming $2n\log\log n$ is replaced by $a_n = (2n\log\log n)^{m/2}$, and the limit set $[0,1]$ becomes the set $\big\{\mathbb{E}\big[h(X_1,\dots,X_m)g(X_1)\cdots g(X_m)\big] : \mathbb{E}g^2(X_1) \le 1\big\}$.

Decoupling and randomization, together with the hypercontractivity property of Rademacher chaos, will be seen to provide an elegant path towards this result.

Contrary to the case of sums of i.i.d. random variables, square integrability of the kernel is not a necessary condition for the lil when $m \geq 2$. However, it is necessary when h is restricted to be of a particular type, and there is a necessary condition for the LIL in terms of integrability of h which differs from square integrability only by a power of $\log\log|h|$.

To ease notation, given $h : S^m \to \mathbb{R}$, we set

$$\alpha_n(h) := \frac{1}{(2n\log\log n)^{\frac{m}{2}}} \sum_{(i_1,\ldots,i_m)\in I_n^m} h(X_{i_1},\ldots,X_{i_m}). \tag{3.9}$$

The lil for canonical U–statistics is then as follows:

3.4. THEOREM. *Let X, X_i, $i \in \mathbb{N}$, be i.i.d. random variables with values in a measurable space $(S, \mathcal{S})$ and common law P. Let $h_j : S^m \to \mathbb{R}$ be P–canonical functions with $\mathbb{E}h_j^2 < \infty$, $j = 1, \ldots, d$. Then, with probability one, the sequence*

$$\big\{(\alpha_n(h_1),\ldots,\alpha_n(h_d))\big\}_{n=1}^{\infty} \tag{3.10}$$

is relatively compact in $\mathbb{R}^d$ and its limit set is

$$K := \Big\{\mathbb{E}\big[g(X_1)\cdots g(X_m)\big(h_1(X_1,\ldots,X_m),\ldots,h_d(X_1,\ldots,X_m)\big)\big] : \mathbb{E}g^2(X) \leq 1\Big\}. \tag{3.11}$$

This theorem is due to Dehling (1989) for $m = 2$ and to Arcones and Giné (1995) for general m. Dehling and Utev (to appear) gives a sketch of a proof of Theorem 3.4 for general m and $d = 1$, different from ours.

The main point in our proof consists in obtaining the following intermediate proposition (the bounded lil):

3.5. PROPOSITION. *Let $(S, \mathcal{S}, P)$ be a probability space, let X_i, $i \in \mathbb{N}$, be i.i.d. random variables with values in S and law P, and let $h : S^m \to \mathbb{R}$ be a P–canonical kernel such that $\mathbb{E}h^2 < \infty$. Then, for every $0 < p < 2$, there exists a constant $C_{m,p} < \infty$ depending only on m and p such that*

$$\mathbb{E}\sup_{n\in\mathbb{N}}\Bigg[\frac{1}{(2n\log\log n)^{m/2}}\Bigg|\sum_{(i_1,\ldots,i_m)\in I_n^m} h(X_{i_1},\ldots,X_{i_m})\Bigg|\Bigg]^p \leq C_{m,p}\big(\mathbb{E}h^2(X_1,\ldots,X_m)\big)^{\frac{p}{2}}. \tag{3.12}$$

(3.3) and polarization imply that the set of finite linear combinations of functions of the form

$$h_m^{\psi}(x_1,\ldots,x_m) := \psi(x_1)\cdots\psi(x_m), \quad \int \psi d\mathrm{P} = 0, \ \int \psi^2 d\mathrm{P} \leq 1 \tag{3.13}$$

is dense in $L_2^{c,m}(\mathrm{P})$. Then, Proposition 3.5 reduces the proof of Theorem 3.4, by means of a standard approximation argument, to the lil for kernels of the form

$$h = \sum_{r=1}^{k} c_r h_m^{\psi_r}. \tag{3.14}$$

The lil (i.e., Theorem 3.4) for U–statistics with kernels of the form (3.14) is an immediate consequence of Strassen's lil for sums of $\mathbb{R}^d$ valued i.i.d. random variables. Here is how it works for $m = 2$ and $d = 1$. Strassen's lil (e.g., Ledoux and Talagrand, 1991) asserts that if $\mathbf{Y}$, $\mathbf{Y}_i$, $i \in \mathbb{N}$, are i.i.d. random vectors in $\mathbb{R}^k$, then, the sequence

$$\left\{\frac{1}{(2n\log\log n)^{\frac{1}{2}}}\sum_{i=1}^{n}\mathbf{Y}_i(\omega)\right\}_{n=1}^{\infty}$$

is relatively compact for almost every $\omega \in \Omega$ if and only if $\mathbb{E}\mathbf{Y} = 0$ and $\mathbb{E}|\mathbf{Y}|^2 < \infty$, and then

$$\lim\ \mathrm{set}\left\{\frac{1}{(2n\log\log n)^{\frac{1}{2}}}\sum_{i=1}^{n}\mathbf{Y}_i(\omega)\right\} = K_{\mathbf{Y}}\quad \text{a.s.}$$

where $K_{\mathbf{Y}}$ is the subset of $\mathbb{R}^k$ defined by

$$K_{\mathbf{Y}} = \left\{\mathbb{E}\left[\left(g(\mathbf{Y})\right)\mathbf{Y}\right] : g \text{ real, measurable, and } \mathbb{E}g^2(\mathbf{Y}) \le 1\right\}.$$

Hence, if $\psi_i(X_1)$, $i = 1, \ldots, k$, are centered and square integrable, we have

$$\begin{aligned}\lim\ \mathrm{set}&\left\{\frac{1}{(2n\log\log n)^{\frac{1}{2}}}\sum_{j=1}^{n}\left(\psi_1(X_j),\ldots,\psi_k(X_j)\right)\right\}\\ &= \left\{\mathbb{E}\left[g(X)\left(\psi_1(X),\ldots,\psi_k(X)\right)\right] : \mathbb{E}g^2(X) \le 1\right\},\end{aligned}$$

where we take this statement to mean that, moreover, the sequence in question is relatively compact with probability one, and where X is a random variable with law P. Then, applying the continuous function $\lambda(x_1, \ldots, x_k) = \sum_{r=1}^{k} c_r x_r^2$ to both terms, we obtain

$$\begin{aligned}\mathrm{limset}&\left\{\frac{1}{2n\log\log n}\sum_{r=1}^{k}c_r\Big(\sum_{j=1}^{n}\psi_r(X_j)\Big)^2\right\}\\ &= \left\{\sum_{r=1}^{k}c_r\left[\mathbb{E}\psi_r(X)g(X)\right]^2 : \mathbb{E}g^2(X) \le 1\right\}.\end{aligned} \tag{3.15}$$

Now, with h as in (3.14) and $m = 2$,

$$\begin{aligned}\frac{1}{2n\log\log n}\sum_{r=1}^{k}c_r\Big(\sum_{j=1}^{n}\psi_r(X_j)\Big)^2 &= \frac{1}{2n\log\log n}\sum_{1\le i,j\le n}h(X_i,X_j)\\ &= \frac{1}{2n\log\log n}\sum_{I_n^2}h(X_i,X_j) + \frac{1}{2n\log\log n}\sum_{i=1}^{n}h(X_i,X_i),\end{aligned}$$

and the last summand tends to zero a.s. by the law of large numbers since $\mathbb{E}|h(X_1, X_1)| \le \sum |c_r|\mathbb{E}\psi_r^2(X_1) < \infty$. Moreover,

$$\sum_{r=1}^{k} c_r \left[\mathbb{E}\psi_r(X)g(X)\right]^2 = \sum_{r=1}^{k} c_r \mathbb{E}\left[\psi_r(X_1)\psi_r(X_2)g(X_1)g(X_2)\right]$$
$$= \mathbb{E}\left[h(X_1, X_2)g(X_1)g(X_2)\right].$$

Hence, (3.15) becomes

$$\operatorname{limset}\left\{\frac{1}{2n\log\log n}\sum_{I_n^2} h(X_i, X_j)\right\} = \left\{\mathbb{E}\left[h(X_1, X_2)g(X_1)g(X_2)\right] : \mathbb{E}g^2(X) \le 1\right\},$$

that is, the lil in Theorem 3.4 for the simple function h given by (3.14). The proof for $m > 2$ is slightly more involved, and Newton's identities help to account for the sums with repeated X_i's in the analogue of the identity below (3.15).

We next see how to obtain the basic inequality (3.12) in Proposition 3.5 by indicating several steps.

Step 1: Decoupling and randomization. Let K be a natural number and let $0 < p < 2$. Letting $\tilde{h}_r = (0, \overset{r-1}{\ldots}, 0, h/a_r, h/a_{r+1}, \ldots, h/a_{2^K}) \in \ell_{2^K}^\infty$, it is easy to see that

$$\max_{n \le 2^K} \frac{1}{a_n}\Big|\sum_{\mathbf{i} \in I_n^m} h(\mathbf{X}_\mathbf{i})\Big| = \Big\|\sum_{\mathbf{i} \in I_{2^K}^m} \tilde{h}_{|\mathbf{i}|}(\mathbf{X}_\mathbf{i})\Big\|.$$

So, we can apply Theorem 2.1 for $\ell_{2^K}^\infty$-valued kernels (which can be viewed as a family of real valued kernels, as indicated immediately below the proof of Theorem 2.1) and obtain

$$\mathbb{E}\max_{n \le 2^K}\left|\frac{1}{a_n}\sum_{\mathbf{i} \in I_n^m} h(\mathbf{X}_\mathbf{i})\right|^p \le C\mathbb{E}\max_{n \le 2^K}\left|\frac{1}{a_n}\sum_{\mathbf{i} \in I_n^m} \varepsilon_\mathbf{i}^{\mathrm{dec}} h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\right|^p$$

for some constant $C < \infty$ depending only on m and p.

Step 2: Blocking. Blocking is an essential part of the proof of the lil for sums of i.i.d. random variables, and it is achieved via maximal inequalities (Lévy or Ottaviani). Once the statistic is decoupled and randomized, we can apply Lévy's inequalities conditionally on all but one of the m sequences $\{\varepsilon_i^j, X_i^j : i \in \mathbb{N}\}$, repeatedly, and obtain

$$\mathbb{E}\max_{n \le 2^K}\left|\frac{1}{a_n}\sum_{\mathbf{i} \in I_n^m} \varepsilon_\mathbf{i}^{\mathrm{dec}} h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\right|^p \le 2^m\mathbb{E}\max_{k \le K-1}\left|\frac{1}{a_k^*}\sum_{\mathbf{i} \in I_{2^{k+1}}^m} \varepsilon_\mathbf{i}^{\mathrm{dec}} h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\right|^p,$$

where $a_k^* := a_{2^k}$. We are now prepared for application of a basic property of Rademacher multilinear forms.

Step 3a: A maximal inequality. Let ψ on $\mathbb{R}_+ \cup \{0\}$ be a *Young modulus* that is, a real function such that $\psi(0) = 0$ and ψ is convex and strictly increasing to infinity.

Then, the space $L_\psi(\Omega, \Sigma, \Pr)$ of all the random variables ξ defined on Ω such that $\mathbb{E}\psi(|\xi|/c) < \infty$ for some $0 < c < \infty$, equipped with the norm

$$\|\xi\|_\psi = \inf\{c > 0 : \mathbb{E}\psi(|\xi|/c) \le 1\},$$

is a Banach space (cf. Krasnoselsky and Rutitsky, 1961). For instance, if $\psi(x) = x^p$, $1 \le p < \infty$, then $\|\xi\|_\psi = \|\xi\|_p$. We are more intersted in Young functions of exponential type, ψ_α, $0 < \alpha < \infty$, which are defined as follows:

$$\psi_\alpha(x) = \begin{cases} \exp(x^\alpha) - 1, & \text{if } \alpha \ge 1 \\ \tau_\alpha(x) - \alpha \exp\left(\frac{1-\alpha}{\alpha}\right), & \text{if } 0 < \alpha < 1, \end{cases}$$

where $\tau_\alpha(x)$ denotes $\exp(x^\alpha)$ if $x \ge \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{\alpha}}$, and it denotes the (ordinate of) tangent line to the function $y = \exp(x^\alpha)$ at the point $\left(\left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{\alpha}}, \exp\left(\frac{1-\alpha}{\alpha}\right)\right)$ if $0 \le x \le \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{\alpha}}$. (The complication in the definition of ψ_α for $\alpha < 1$ is due to the fact that the function $y = \exp(x^\alpha)$ is not convex near zero.) Note that for all $p > 0$ and all $\alpha > 0$ there is $c_{p,\alpha} < \infty$ such that

$$\|\xi\|_p \le c_{p,\alpha} \|\xi\|_{\psi_\alpha}.$$

We can now state a useful maximal inequality (Arcones and Giné, 1995), valid for Young moduli slightly more general than ψ_α, but not for power moduli.

3.6. PROPOSITION. *Let ψ be a Young modulus such that*

$$\limsup_{x \wedge y \to \infty} \frac{\psi^{-1}(xy)}{\psi^{-1}(x)\psi^{-1}(y)} < \infty \tag{3.16}$$

and

$$\limsup_{x \to \infty} \frac{\psi^{-1}(x^2)}{\psi^{-1}(x)} < \infty. \tag{3.17}$$

Then, there exists a finite constant C_ψ such that, for every sequence of random variables $\{\xi_k : k \in \mathbb{N}\}$,

$$\left\| \sup_k \frac{|\xi_k|}{\psi^{-1}(k)} \right\|_\psi \le C_\psi \sup_k \|\xi_k\|_\psi. \tag{3.18}$$

This inequality is good for us because of the following property of Rademacher chaos variables:

Step 3b: Integrability (hypercontractivity) of Rademacher sums. If $Y = \sum a_i \varepsilon_i$, it is classical (Bonami, 1970) that

$$(\mathbb{E}|Y|^p)^{1/p} \le (p-1)^{1/2} (\sum a_i^2)^{1/2}, \quad p \ge 2.$$

Conditionally applying this inequality to the decoupled Rademacher m–linear form

$$Z = \sum_{i_1, \ldots, i_m} a_{i_1 \ldots i_m} \varepsilon^1_{i_1} \cdots \varepsilon^m_{i_m},$$

we obtain

$$(\mathbb{E}|Z|^p)^{1/p} \le (p-1)^{m/2}(\sum a^2_{i_1\dots i_m})^{1/2}, \quad p \ge 2.$$

(This inequality is also true for undecoupled Rademacher multilinear forms, but it requires a little more work and we do not need it here.) Then, developing the exponential, one gets

$$\|Z\|_{\psi_{2/m}} \le C_m(\sum a^2_{i_1\dots i_m})^{1/2} \tag{3.19}$$

for some universal constant C_m. Combining (3.18) and (3.19), we obtain that, if Z_k is a sequence of decoupled Rademacher m–linear forms, then

$$\left\|\sup_k \frac{|Z_k|}{(\log k)^{m/2}}\right\|_\psi \le c_m \sup_k\bigl(\mathbb{E}|Z_k|^2\bigr)^{1/2} \tag{3.20}$$

Step 3c: Applying inequality (3.20). If we apply inequality (3.20) to the right hand side of the inequality from Step 2, we obtain (recall that the $\psi_{2/m}$ norm dominates a constant times the L_p norm)

$$\begin{aligned}\mathbb{E}_\varepsilon \max_{k\le K-1}\left|\frac{1}{a_k^*}\sum_{\mathbf{i}\in I^m_{2^{k+1}}} \varepsilon_{\mathbf{i}}^{\mathrm{dec}} h(\mathbf{X}_{\mathbf{i}}^{\mathrm{dec}})\right|^p &\le 3^{\frac{mp}{2}}\mathbb{E}_\varepsilon \max_{k\le K}\left|\frac{1}{(\log k)^{\frac{m}{2}}}\frac{1}{2^{\frac{km}{2}}}\sum_{i\in I^m_{2^k}} \varepsilon_{\mathbf{i}}^{\mathrm{dec}} h(\mathbf{X}_{\mathbf{i}}^{\mathrm{dec}})\right|^p \\ &\le C\max_{k\le K}\Bigl(\frac{1}{2^{km}}\sum_{\mathbf{i}\in I^m_{2^k}} h^2(\mathbf{X}_{\mathbf{i}}^{\mathrm{dec}})\Bigr)^{\frac{p}{2}}\end{aligned}$$

for a constant $C < \infty$ depending only on p and m, where $\mathbb{E}_\varepsilon$ denotes integration with respect to the ε variables only. So, integrating with respect to the X variables, we have

$$\mathbb{E}\max_{k\le K-1}\left|\frac{1}{a_k^*}\sum_{\mathbf{i}\in I^m_{2^{k+1}}} \varepsilon_{\mathbf{i}}^{\mathrm{dec}} h(\mathbf{X}_{\mathbf{i}}^{\mathrm{dec}})\right|^p \le C\mathbb{E}\max_{k\le K}\Bigl(\frac{1}{2^{km}}\sum_{\mathbf{i}\in I^m_{2^k}} h^2(\mathbf{X}_{\mathbf{i}}^{\mathrm{dec}})\Bigr)^{\frac{p}{2}}.$$

Thus, we have reduced the lil to a law of large numbers since the variable at the right is basically an average.

Step 4. Doob's maximal inequality. Now it is an exercise to check that Doob's maximal inequality for reverse martingales (recall h^2 is integrable) bounds the last expected value by $(2/(2-p))\bigl(\mathbb{E}h^2(\mathbf{X})\bigr)^{p/2}$, and Proposition 3.5 obtains by combining the four steps.

Pisier (1975) has an analogous approach to the lil for sums of i.i.d. Banach valued random variables. The proof sketched here is from Arcones and Giné (1995).

The lil for degenerate U–statitics is still unfinished. Next we coment on some recent developments. The first comment is that the condition $\mathbb{E}h^2 < \infty$ in Theorem 3.4 is the best possible *moment* condition for the lil: it can be shown (Giné and Zhang, 1996) that this condition is necessary for kernels of the form (3.14) (with $k < \infty$). The best *necessary integrability conditions* for the bounded lil so far are the following (Giné and Zhang, loc cit):

3.7. THEOREM. *Suppose we have*

$$\limsup_{n\to\infty} \frac{1}{(2n\log\log n)^{\frac{m}{2}}}\Big|\sum_{(i_1,\ldots,i_m)\in I_n^m} h(X_{i_1},\ldots,X_{i_m})\Big| < \infty \quad \text{a.s.} \tag{3.21}$$

Then

$$\sup_{n\ge 1} \frac{\mathbb{E}\min\left\{h^2(X_1,\ldots,X_m), n(\log\log n)^{m-1}\right\}}{(\log\log n)^{m-1}} < \infty. \tag{3.22}$$

In particular, h is P–canonical and, for $m \ge 2$,

$$\mathbb{E}h^2(\log\log|h|)^{2-m}g(\log\log|h|) < \infty \tag{3.23}$$

for all bounded non–negative monotone decreasing integrable functions g on $\mathbb{R}^+$.

The proof of this theorem relies heavily on decoupling. The integrability condition (3.22) for $m = 1$ reduces to $\mathbb{E}h^2 < \infty$, so that Theorem 3.7 recovers in particular Strassen's converse lil. The following example shows that (3.23) cannot, in general, be improved.

3.8. EXAMPLE. Let I_r be the indicator function of the interval $\big(1-\frac{1}{2^{r-1}}, 1-\frac{1}{2^r}\big]$, for all $r \in \mathbb{N}$, and let

$$h\big(x,\alpha),(y,\beta)\big) = \sum_{r=1}^{\infty} \frac{2^r}{\sqrt{r}} I_r(x)I_r(y)\alpha\beta,$$

defined on S^2, $S = [0,1]\times\{-1,1\}$. Let $P = \lambda\times\big(\delta_{-1}+\delta_1\big)/2$, where λ is Lebesgue measure. Then h is P–canonical and it can be proved, using truncation, binomial probabilities and a better exponential inequality for Rademacher chaos (Ledoux and Talagrand, 1991), that this kernel does satisfy the bounded lil, i.e., that

$$\limsup_{n\to\infty} \frac{1}{n\log\log n}\Big|\sum_{(i,j)\in I_n^2} h\big((X_i,\varepsilon_i),(X_j,\varepsilon_j)\big)\Big| < \infty \quad \text{a.s.}$$

However, not only $\mathbb{E}h^2(X_1,X_2) = \infty$, but in fact, $\int h^2 g(\log\log|h|)dP^2 = \infty$ for all bounded, decreasing functions g such that $\int_0^\infty g(t)dt = \infty$. So, this example shows that for P–canonical U–statistics, (i) the condition $\mathbb{E}h^2 < \infty$ is not necessary for the lil, and (ii) the necessary integrability conditions given in Theorem 3.7 are best possible, at least for $m = 2$. We refer to Giné and Zhang (1996) for details.

It has recently been shown (Goodman, 1996) that

3.9. THEOREM. *Let $h(x,y)$ be a measurable symmetric P–canonical kernel and let X, X_i, $i\in\mathbb{N}$, be i.i.d.(P). Let $H(x) := \mathbb{E}h(x,X)$. Then, the conditions*

$$\mathbb{E}\Big[\frac{H(X)}{\log\log H(X)}\Big] < \infty$$

and

$$\sigma := \sup_{\mathbb{E}g^2(X)\le 1} \mathbb{E}\big[h(X_1,X_2)g(X_1)g(X_2)\big] < \infty$$

imply

$$\limsup_{n\to\infty} \frac{1}{2n\log\log n}\Big|\sum_{I_n^2} h(X_i, X_j)\Big| < 4\sqrt{2\sigma} \quad \text{a.s.}$$

This is the best relatively satisfactory lil under conditions weaker than $\mathbb{E}h^2(X_1, X_2) < \infty$. For another lil that does not require finiteness of the second moment of h, see Giné and Zhang, loc. cit.

We should remark that, by Hoeffding's decomposition, the lil with norming $(2n\log\log n)^{-1/2}$ for non-degenerate square integrable U–statistics reduces to the lil for sums of the i.i.d. random variables $P^{m-1}h(X_i)$ together with Marcinkiewicz type laws of large numbers for degenerate U–statistics, as noted in Sefling (1971).

4. Limit theorems for U–processes. We are concerned in this section with the law of large numbers, the central limit theorem and the law of the iterated logarithm for U–processes. The three types of results reduce to control of quantities of the form

$$\Pr\big\{\sup_{\mathcal{H}} |U_n(h)| > c_n\big\} \quad \text{or} \quad \mathbb{E}\big[\sup_{\mathcal{H}} |U_n(h)|\big],$$

where $\mathcal{H}$ is a family of kernels and Pr, $\mathbb{E}$, must be replaced by $\Pr^*$, $\mathbb{E}^*$ (outer probability, outer expectation), when $\sup |U_n(h)|$ is not measurable. This is immediately clear for the lln and the lil, and Theorem 2.1.3 in the lectures on the bootstrap in this volume makes it also clear for the clt. We will estimate these quantities via metric entropy bounds after decoupling and randomization, as we now explain. (This is in analogy with empirical processes, with the added ingredient of decoupling.)

We recall that, as in (3.19), if

$$Z = \sum_{1\le i_1<\ldots<i_m\le n} a_{i_1,\ldots,i_m}\varepsilon_{i_1}\cdots\varepsilon_{i_m}$$

is a real homogeneous polynomial of degree m in n independent Rademacher variables, then Bonami's result gives

$$\|Z\|_{\psi_{2/m}} \le \big(\mathbb{E}Z^2\big)^{1/2} = \big(\sum a_{\mathbf{i}}^2\big)^{1/2} \tag{4.1}$$

for a constant $C_m < \infty$ that only depends on m (see e.g. Kwapień and Woyczynski, 1992, or our forthcoming book). Then, if Z_k, $k \le N$, is a family of such polynomials, Proposition 3.6 gives the following maximal inequality:

$$\|\max_{k\le N} |Z_k|\|_{\psi_{2/m}} \le C_m[\log N]^{m/2}\max_k\big(\mathbb{E}Z_k^2\big)^{1/2}, \quad N > 1. \tag{4.2}$$

The norm at the left side can be replaced by the L_p norm for any p, with a change in the constant, that we will continue denoting as C_m. The maximal inequality (4.2) is really all that is needed (besides decoupling and randomization) to prove the law of large numbers. To prove the other two limit theorems, one combines inequality (4.2) with a measure of the size of the class $\mathcal{H}$, namely, its metric entropy for certain distances, particularly if $\mathcal{H}$ is Vapnik–Červonenkis, to obtain the pertinent maximal inequalities. The key to this is the following well known theorem, basically due to Dudley (1967), with formal but important improvements by other authors (Pisier, Fernique). For a detailed proof, see Ledoux and Talagrand (1991) or our forthcoming book. First, some definitions:

The *covering number* $N(T,d,\varepsilon)$, $\varepsilon>0$, of a metric or pseudometric space (T,d) is the smallest number of open balls of radius at most ε and centers in T required to cover T, that is,

$$N(T,d,\varepsilon):=\min\{n : \text{there exist } t_1,\ldots,t_n\in T \text{ such that } T\subseteq \cup_{i=1}^n B(t_i,\varepsilon)\}. \tag{4.3}$$

A process $X(t)$, $t\in T$, (T,d) a metric or pseudometric space, is *separable* if there exist a countable set $T_0\subseteq T$ and a set $\Omega_0\subseteq\Omega$ with $\Pr\Omega_0=0$ such that for all ω not in Ω_0, $t\in T$ and $\varepsilon>0$, $X(t,\omega)$ is in the closure of the set $\{X(s,\omega): s\in T_0\cap B(t,\varepsilon)\}$. It is well known and easy to see that if (T,d) is a separable metric or pseudometric space and X is continuous in probability for d, then X admits a separable version. With these definitions we have:

4.1. THEOREM. *Let (T,d) be a pseudometric space of diameter D and let ψ be a Young modulus satisfying conditions (3.16) and (3.17) in Proposition 3.6. Suppose that*

$$\int_0^D \psi^{-1}\big(N(t,d,\varepsilon)\big)d\varepsilon<\infty, \tag{4.4}$$

and let $X(t)$, $t\in T$, be a stochastic process satisfying

$$\|X(t)-X(s)\|_\psi\le d(s,t),\ \ s,t\in T. \tag{4.5}$$

Then, any separable version $\tilde X$ of X has almost all its sample paths in $C_u(T,d)$ and, moreover,

$$\Big\|\sup_{t\in T}|\tilde X(t)|\Big\|_\psi\le\|\tilde X(t_0)\|_\psi+K\int_0^D\psi^{-1}\big(N(T,d,\varepsilon)\big)d\varepsilon, \tag{4.6}$$

and

$$\Big\|\sup_{\substack{d(s,t)<\delta\\ s,t\in T}}|\tilde X(t)-\tilde X(s)|\Big\|_\psi\le K\int_0^\delta\psi^{-1}\big(N(T,d,\varepsilon)\big)d\varepsilon \tag{4.7}$$

for all $\delta>0$ and a finite constant K that depends only on ψ.

Since, for X separable,

$$\Big\|\sup_{t\in T}|X(t)|\Big\|_\psi=\sup_{\substack{S\subseteq T\\ S\text{ finite}}}\Big\|\max_{t\in S}|X(t)\Big\|_\psi$$

and likewise for $\big\|\sup_{d(s,t)<\delta,\, s,t\in T}|X(t)-X(s)|\big\|_\psi$, the sups in (4.6) and (4.7) can be replaced by maxima over finite sets. The key estimate here, of which all the 'chaining' proofs are (more or less complicated) variations of, is the following. Let us assume $X(t_0)=0$ and T finite. For each $k=0,1,\ldots$ let $\{t_1^k,\ldots,t_{N_k}^k\}=T_k$ be the centers of $N_k:=N(T,d,2^{-k})$ open balls of radius at most 2^{-k} and centers in T covering T. Note that T_0 consists of one point, which we may take to be t_0. For each $k\ge0$ let $\pi_k: T\to T_k$ be a function satisfying $d(t,\pi_k(t))<2^{-k}$ for all $t\in T$, which obviously exists. Moreover, T being finite, there is k_T such that, if $k\ge k_T$ and $s\in T$, then

$d(\pi_k(t),t)=0$; this implies, by (4.5), that $X(t)=X(\pi_k(t))$ a.s. Then, for $t\in T$ we have:

$$X(t)=\sum_{r=1}^{k_T}\big(X(\pi_k(t))-X(\pi_{k-1}(t))\big)\ \text{ a.s.}$$

and, since

$$d(\pi_k(t),\pi_{k-1}(t))\le d(\pi_k(t),t)+d(t,\pi_{k-1}(t))<3\cdot 2^{-k},$$

the maximal inequality (3.18) of Proposition 3.6 (better, (4.2) for ψ) and hypothesis (4.5) give, using (3.17) in the last step,

$$\begin{aligned}\Big\|\max_{t\in T}|X(s)|\Big\|_\psi &\le \sum_{k=1}^{k_T}\Big\|\max_{\substack{t\in T_k,s\in T_{k-1}\\ d(s,t)<3\cdot 2^{-k}}}|X(t)-X(s)|\Big\|_\psi\\ &\le 3C_\psi\sum_{k=1}^{k_S}2^{-k}\psi^{-1}(N_kN_{k-1})\\ &\le K\sum_{k=1}^{k_T}2^{-k}\psi^{-1}(N_k),\end{aligned}$$

that is, inequality (4.6).

There is quite a large class of collections of kernels that have good metric entropy properties for the $L_p(\mathrm{P})$ distance, uniformly in P. These families of functions go by the names of 'Euclidean classes', 'polynomial classes' or 'Vapnik–Červonenkis classes'. A class of sets $\mathcal{C}$ is Vapnik–Červonenkis (Dudley, 1977) if there is an $n<\infty$ such that $\mathcal{C}$ does not shatter any subsets of cardinality n ($\mathcal{C}$ shatters a finite set A, if all the subsets of A can be obtained by intersection of A with sets in $\mathcal{C}$). A class of functions $\mathcal{H}$ is VC–subgraph if the subgraphs of all the functions in $\mathcal{H}$ form a VC class of sets. By a combinatorial lemma due to Sauer, Vapnik and Cervonenkis, and Shelah, independently, and a result of Dudley, extended by Pollard, the VC subgraph classes of functions satisfy the following entropy bound:

4.2. THEOREM. *If $\mathcal{H}$ is a VC–subgraph class of measurable functions on a measurable space $(S,\mathcal{S})$ with an everywhere finite envelope H (H is any measurable function such that $H(s)\ge\sup_{h\in\mathcal{H}}|h(s)|$) and $p\ge 1$, then there are constants $K<\infty$ depending on $\mathcal{H}$ and $d>0$ depending on p and $\mathcal{H}$, such that*

$$N(\mathcal{H},L_p(\mathrm{P}),\|H\|_{\mathrm{P},p}\varepsilon)\le K\Big(\frac{1}{\varepsilon}\Big)^d, \tag{4.8}$$

where P is any probability measure on $(S,\mathcal{S})$ and $\|H\|_{\mathrm{P},p}$ denotes the $L_p(\mathrm{P})$ norm of H.

See e.g., van der Vaart and Wellner (1996, Theorem 2.6.7), Dudley (1984) or Pollard (1984). See Dudley (1988) for other classes of functions that satisfy this theorem or similar bounds that make Theorem 4.1 applicable to them. These references provide many examples of such classes of functions.

Next we describe results on the lln, clt and lil for U–processes.

a) The law of large numbers. Let $\mathcal{H}$ be a class of symmetric functions on S^m with an everywhere finite envelope H, and let P be a probability mesure on $(S, \mathcal{S})$. We ask whether

$$\|U_n - \mathrm{P}^m\|_{\mathcal{H}} := \sup_{h\in\mathcal{H}} |U_n(h) - \mathrm{P}^m h| \to 0 \quad a.s.$$

For $m = 1$ this is the law of large numbers for empirical processes, that is, a generalization to classes of functions of the Glivenko–Cantelli theorem (Vapnik and Červonenkis, 1971,1981; Giné and Zinn, 1984; Talagrand, 1987). The functions being in general non–degenerate, which corresponds to the case $r = 1$ in the randomization theorem (Theorem 2.6), we will only be able to randomize by one Rademacher factor. This will justify the following definition of a random distance on $\mathcal{H}$. For X_i i.i.d.(P) and $f, g \in \mathcal{H}$, we set

$$\tilde{e}_{n,1}(f,g) = \frac{1}{n}\sum_{i_1=1}^{n}\left|\frac{(n-m)!}{(n-1)!}\sum_{(i_2,\ldots,i_m):\mathbf{i}\in I_n^m}(f-g)(X_{i_1},\ldots,X_{i_m})\right|.$$

Actually, here, as in Section 2, we assume the variables X_i to be the coordinates $S^{\mathbb{N}} \to S$, and when we introduce new auxiliary variables, such as Rademacher randomizers, we assume them to be defined also as coordinates in another factor of the general probability space so that, in particular, they are independent of the X variables.

The following is the U–statistic analogue of the VC law of large numbers for empirical processes.

4.3. THEOREM. *Let $\mathcal{H}$ be a measurable (:=image admissible Suslin) class of symmetric kernels on $(S^m, \mathcal{S}^m)$ with everywhere finite P^m–integrable envelope. Then the following statements are equivalent:*

i) $\|U_n - \mathrm{P}^m\|_{\mathcal{H}} \to 0 \quad a.s.$

ii) $\dfrac{1}{n}\log N\big(\mathcal{H}, \tilde{e}_{n,1}, \varepsilon\big) \to 0$ *in pr* for all* $\varepsilon > 0$.

PROOF (Sketch). The strong law i) is equivalent to

$$\mathbb{E}\|U_n - \mathrm{P}^m\|_{\mathcal{H}} \to 0$$

because, under the conditions of the theorem, $\|U_n - \mathrm{P}^m\|_{\mathcal{H}}$, $n \geq m$, is a reversed submartingale. By the randomization theorem (Theorem 2.6), this is equivalent to

$$\mathbb{E}\Big\|\frac{(n-m)!}{n!}\sum_{I_n^m}\varepsilon_{i_1}h(X_{i_1},\ldots,X_{i_m})\Big\|_{\mathcal{H}} \to 0$$

(note $\|\mathrm{P}^m h\|_{\mathcal{H}}\mathbb{E}\big|\sum_{i_1=1}^n \varepsilon_{i_1}/n\big| \to 0$). The implication i) $\Rightarrow$ ii) now follows from a Sudakov type minorization inequality for suprema of linear combinations of Rademacher variables due to Carl and Pajor, 1988 (see Ledoux and Talagrand, 1991, Corollary 4.14).

To prove that ii) implies i) we observe first that, by standard truncation techniques, it suffices to prove

$$\lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}\Big\|\frac{(n-m)!}{n!}\sum_{I_n^m}\varepsilon_{i_1}h(X_{i_1},\ldots,X_{i_m})\Big\|_{\mathcal{H}_M}=0, \tag{4.9}$$

where $\mathcal{H}_M=\{hI_{H\le M}:h\in\mathcal{H}\}$. For simplicity, set $\tilde N_{n,1}(\mathcal{H}_M,\varepsilon):=N(\mathcal{H}_M,\tilde e_{n,1},\varepsilon)$. It is easy to shown (see Arcones and Giné, 1993, page 1511 and Lemma 2.20 in Giné and Zinn, 1984) that condition ii) implies that for all $\varepsilon>0$ there exists $M_0(\varepsilon)<\infty$ such that

$$\mathbb{E}\Big[\frac{1}{n}\log\tilde N_{n,1}(\mathcal{H}_M,\varepsilon)\Big]^r\to 0 \tag{4.10}$$

for all $r<\infty$.

For ω fixed, let $\mathcal{H}_M^*$ be a subset of $\mathcal{H}_M$ of cardinality $\tilde N_{n,1}(\mathcal{H}_M,\varepsilon)$, ε–dense in $\mathcal{H}_M$ for the distance $\tilde e_{n,1}(\omega)$. Then, by the triangle inequality,

$$\begin{aligned}\mathbb{E}_\varepsilon&\Big\|\frac{(n-m)!}{n!}\sum_{I_n^m}\varepsilon_{i_1}h(X_{i_1},\ldots,X_{i_m})\Big\|_{\mathcal{H}_M}\\&\le\varepsilon+\mathbb{E}_\varepsilon\Big\|\frac{(n-m)!}{n!}\sum_{I_n^m}\varepsilon_{i_1}h(X_{i_1},\ldots,X_{i_m})\Big\|_{\mathcal{H}_M^*},\end{aligned}$$

and by inequality (4.2) with $m=1$, we have

$$\begin{aligned}\mathbb{E}_\varepsilon&\Big\|\frac{(n-m)!}{n!}\sum_{I_n^m}\varepsilon_{i_1}h(X_{i_1},\ldots,X_{i_m})\Big\|_{\mathcal{H}_M^*}\\&\le C\frac{(n-m)!}{n!}\big[\log\tilde N_{n,1}(\mathcal{H}_M,\varepsilon)\big]^{1/2}\\&\qquad\times\max_{h\in\mathcal{H}_M^*}\Big[\sum_{i_1=1}^n\Big(\sum_{(i_2,\ldots,i_m):\mathrm{i}\in I_n^m}h(X_{i_1},\ldots,X_{i_m})\Big)^2\Big]^{1/2}\\&\le CM\frac{n^{1/2}}{n-m+1}\big[\log\tilde N_{n,1}(\mathcal{H}_M,\varepsilon)\big]^{1/2}.\end{aligned}$$

Integrating with respect to the X variables in this last inequality, and using (4.10), proves (4.9) and therefore, the law of large numbers i).

□

Since $\tilde e_{n,1}(f,g)\le e_{n,1}(f,g):=U_n(|f-g|)$, and $e_{n,1}$ is the L_1 distance for the (random) uniform probability measure on the points $(X_{i_1},\ldots,X_{i_m})\in S^m$, $1\le i_1<\ldots<i_m\le n$, it follows from Theorems 4.2 and 4.3 that

4.4. COROLLARY. *If $\mathcal{H}$ is a VC class of symmetric funtions on S^m with everywhere finite envelope H, then $\|U_n(h,\mathrm{P})-\mathrm{P}^m h\|_{\mathcal{H}}\to 0$ a.s. for all probability measures* P *on $(S,\mathcal{S})$ for which H is P^m–integrable.*

For example, if X_i are i.i.d. in $\mathbb{R}^2$, with law P, and $\Delta(x,y,z)$ denotes the triangle with vertices x, y, z in $\mathbb{R}^2$, then Corollary 4.4 shows that

$$\sup_{\theta\in\mathbb{R}^2}\left|\frac{1}{\binom{n}{3}}\sum_{1\le i_1<i_2<i_3\le n} I\{\theta\in\Delta(X_{i_1},X_{i_2},X_{i_3})\} - \Pr\{\theta\in\Delta(X_1,X_2,X_3)\}\right| \to 0 \ \text{ a.s.} \tag{4.11}$$

for any probability laws P in $\mathbb{R}^2$, because the corresponding family of subsets of $\mathbb{R}^6$, $C_\theta := \{(x,y,z) : \theta \in \Delta(x,y,z)\}, \theta \in \mathbb{R}^2$, is VC. The process

$$D_n(\theta) := \frac{1}{\binom{n}{3}}\sum_{1\le i_1<i_2<i_3\le n} I\{\theta\in\Delta(X_{i_1},X_{i_2},X_{i_3})\}, \quad \theta\in\mathbb{R}^2,$$

is the *simplicial depth process* of the sample $\{X_i\}_{i=1}^n$. The argmax of this process are the simplicial medians of the sample, and the argmax of the function $D(\theta) := \Pr\{\theta \in \Delta(X_1, X_2, X_3)\}$, if it exists (even if it is not unique) is the population simplicial median (Liu, 1990). It can be shown that, as a consequence of the above limit, the empirical simplicial median is a consistent estimator of the population simplical median. This extends to any dimensions.

Another corollary of Theorem 4.3 is the following extension to U–statistics of the Maurier law of large numbers:

4.5. COROLLARY. *If B is a separable Banach space and if $h : S^m \to B$ satisfies $\mathbb{E}\|h\| < \infty$, then $U_n(h) \to \mathrm{P}^m h$ a.s.*

Except for Corollary 4.4 for $m = 2$, which belongs to Nolan and Pollard (1987), the results described in this subsection were obtained by Arcones and Giné (1993), and we refer to this article for detailed proofs. The Nolan–Pollard article contains a very interesting application of rates of convergence to zero of degenerate U–processes to density estimation. See also Turki–Moalla (1996).

b) The central limit theorem. We refer to the lectures on the bootstrap, Section 2.1, for background on convergence in law in $\ell^\infty(T)$. In the case of empirical processes, T is a class of functions of one variable whereas in the U–processes case, T is a set of functions of several variables. Recall also the definition of the processes G_{P} and K_{P} from Section 3a) above.

Given a kernel $h(x_1, \ldots, x_m)$ not necessarily P–degenerate, we can decompose it as

$$h(x_1,\ldots,x_m) - \mathrm{P}^m h = \sum_{k=1}^m \big((\mathrm{P}^{m-1}h)(x_k) - \mathrm{P}^m h\big) + \bar{h}(x_1,\ldots,x_m)$$

where $\bar{h}$, defined by this same relation, is centered and degenerate of order 1. Then, the central limit theorem for the first term reduces to the clt for the empirical measure over the class of functions of one variable $\{\mathrm{P}^{m-1}h : h \in \mathcal{H}\}$. The fact that the kernel $\bar{h}$ is degenerate of order 1 (with $r = 2$ in Definition 2.4) implies, by Theorem 2.6, that we can randomize the U–process with kernel $\bar{h}$ with two Rademacher factors, and this is easily seen to be equivalent, in this situation, to randomization by two Rademacher factors of the original U–process. In conclusion, we obtain the following corollary of the randomization theorem (Arcones and Giné, 1993):

4.6. THEOREM. *Let $\mathcal{H}$ be a measurable class of symmetric kernels on $(S^m, \mathcal{S}^m)$ with everywhere finite envelope H and let* P *be a probability measure on $(S, \mathcal{S})$. Then, the conditions*

i) the class $\mathrm{P}^{m-1}\mathcal{H} := \{\mathrm{P}^{m-1}h : h \in \mathcal{H}\}$ is P*-Donsker,*

ii) there exists $r > 0$ such that

$$\mathbb{E}\Big\|n^{-m+1/2}\sum_{I_n^m}\varepsilon_{i_1}\varepsilon_{i_2}h(X_{i_1},\ldots,X_{i_m})\Big\|_{\mathcal{H}}^r \to 0,$$

imply that

$$\Big\{n^{1/2}(U_n(h) - \mathrm{P}^m h) : h \in \mathcal{H}\Big\} \to_{\mathcal{L}} \Big\{mG_{\mathrm{P}}(P^{m-1}h - \mathrm{P}^m h) : h \in \mathcal{H}\Big\}. \tag{4.12}$$

Conversely, if $t^2\mathrm{P}^m\{H > t\} \to 0$ as $m \to \infty$ and the clt (4.12) holds, then condition i) holds and the limit in condition ii) holds true for every $0 < r < 2$.

Condition i) has to do with empirical processes, which have been extensively studied, and condition ii) can be checked using the tools for Rademacher chaos decribed at the beginning of the section. For instance, a direct application of Theorem 4.1 with $\psi = \psi_1$, the exponential modulus of order 1, which corresponds to the Rademacher chaos of order 2 in inequalities (4.1), (4.2), shows that the condition

$$n^{-1/2}\mathbb{E}^*\int_0^\infty \log N(\mathcal{H}, e_{n,2}, \varepsilon)d\varepsilon \to 0,$$

where $e_{n,2}(f,g) = \big[U_n((f-g)^2)\big]^{1/2}$, implies condition ii). (A smaller random distance suffices.) In particular, by Theorem 4.2, *measurable VC classes satisfy the clt (4.12) for any* P *for which the envelope is square integable* (again, less than square integrability suffices). The clt for not necesarily degenerate VC classes was obtained, for $m = 2$, by Nolan and Pollard (1988), and by Arcones and Giné (1993) and, independently, by Sherman (1994), for general m.

The clt for degenerate classes is, in a sense, more interesting. Combining the randomization theorem for $r = m$, the maximal inequality (4.2), Theorem 4.1, the characterization of convergence in law in $\ell^\infty(T)$ given as Theorem 2.1.3 in the bootstrap lectures, and Theorem 3.1 for finite dimensional convergence, we obtain the following result (recall $\pi_k^{\mathrm{P}}(h)$ denotes the Hoeffding projection of h, and that $\pi_k^{\mathrm{P}}(h)$ is canonical –see Section 2b)):

4.7. THEOREM. *Let $\mathcal{H}$ be a mesurable VC class of symmetric kernels $S^m \to \mathbb{R}$, with everywhere finite envelope H such that $\mathrm{P}^m H^2 < \infty$. Then,*

$$\left\{\binom{n}{k}^{k/2} U_n^{(k)}(\pi_{k,m}^{\mathrm{P}}(h)) : h \in \mathcal{H}\right\} \to_{\mathcal{L}} \{K_{\mathrm{P}}((\pi_{k,m}^{\mathrm{P}}(h)) : h \in \mathcal{H}\}$$

in $\ell^\infty(\mathcal{H})$, for $k = 1, \ldots, m$, where K_{P} is the isonormal Gaussian chaos process associated to P.

For a detailed proof, see Corollary 5.7 in Arcones and Giné (1993). (See also Nolan and Pollard, 1988, for $m = 2$; Sherman, 1994, Corollary 8, contains a slightly weaker result for general m, as well as an interesting application to generalized regression).

Bach to the simplicial depth process D_n in (4.11), Theorem 4.6 shows that the simplicial depth process satisfies the clt, that is,

$$n^{1/2}(D_n - D) \to_{\mathcal{L}} G$$

in $\ell^\infty(\mathbb{R})$, and where G is the Gaussian process mG_{P}. Chen, Arcones and Giné (1994) use this observation to prove that, if P is angularly symmetric with respect to some center and enjoys certain smoothness properties, then the empirical simplicial median is an asymptotically normal $\sqrt{n}$–consistent estimator of the population simplicial median. This article also contains a general setup for treating M–estimators based on multivariate criterion functions, something considered also by Sherman (1994).

In empirical process theory, besides the VC condition, there is another type of conditions that ensures that some limit theorems hold, namely, 'bracketing conditions'. We will not treat these here because they do not seem to be very adequate in the degenerate situation: see Arcones and Giné (1993) and Turki–Moalla (1996).

c) The law of the iterated logarithm. The lil for non–degenerate U–processes reduces to the lil for the empirical process over the class $\{\mathrm{P}^{m-1}h : h \in \mathcal{H}\}$, well studied (e.g., Ledoux and Talagrand, 1991), and to convergence to zero a.s. of the higher order terms in the Hoeffding decomposition. This is done in Arcones (1993). The lil for VC classes of P–degenerate functions was considered in Arcones and Giné (1995), who obtained the following:

4.8. THEOREM. *Let $\mathcal{H}$ be a measurable VC class of* P*–canonical symmetric functions $S^m \to \mathbb{R}$ with an everywhere finite enevelope H such that $\mathbb{E}H^2 < \infty$. Let $\alpha_n(h)$, $h \in \mathcal{H}$, be defined as in (3.9). Then, for almost every ω, the sequence $\{\alpha_n(h,\omega) : h \in \mathcal{H}\}_{n=1}^\infty$ is relatively compact in $\ell^\infty(\mathcal{H})$ and its limit set is*

$$K = \Big\{\{\mathbb{E}\big[h(X_1,\ldots,X_m)g(X_1)\cdots g(X_m)\big] : h \in \mathcal{H}\} : \mathbb{E}g^2(X) \le 1\Big\}.$$

We succintly comment on the proof (see our article for details). The reduction to an analogue of Proposition 3.5 above is less straightforward than in the finite dimensional case, but still flows along lines similar to the analogous reduction in the finite dimensional case as described in Section 3 (some extra arguments, that we omit, are necessary). Inequality (3.12) in Proposition 3.5 now takes the form

$$\mathbb{E}\sup_n \|\alpha_n(h)\|_{\mathcal{H}}^p \le C_{m,p}\big(\mathbb{E}H^2(X_1,\ldots,X_m)\big)^{p/2} \tag{4.13}$$

for all $0 < p < 2$. The proof of (4.13) follows the same steps as the proof of (3.12) described above, the main difference being that now, instead of inequality (3.19), we must use the following analogue:

4.9. LEMMA. *If $\mathcal{H}$ is a measurable VC class of functions, then*

$$\Big\|\big\|\sum \varepsilon_{i_1}^1 \cdots \varepsilon_{i_m}^m h(x_{i_1}^1, \ldots, x_{i_m}^m)\big\|_{\mathcal{H}}\Big\|_{\psi_{2/m}} \leq c\Big(\sum H^2(x_{i_1}^1, \ldots, x_{i_m}^m)\Big)^{1/2}.$$

PROOF. There is no loss of generality in assuming that $\mathcal{H}$ contains the function 0. Take T in Theorem 4.1 to be $\mathcal{H}$. Take $d^2(h_1, h_2)$ to be $\sum(h_2 - h_1)^2(x_{i_1}^1, \ldots, x_{i_m}^m)$. Then, the process $X(h) = \sum \varepsilon_{i_1}^1 \cdots \varepsilon_{i_m}^m h(x_{i_1}^1, \ldots, x_{i_m}^m)$ satisfies inequality (4.5) with $\psi = \psi_{2/m}$ due to inequality (4.1) (=(3.19)), and the diameter of $\mathcal{H}$ for this pseudodistance is precisely

$$D = 2\Big(\sum H^2(x_{i_1}^1, \ldots, x_{i_m}^m)\Big)^{1/2}.$$

Moreover, $X = \tilde{X}$. Hence, inequality (4.6) in Theorem 4.1 gives

$$\Big\|\big\|\sum \varepsilon_{i_1}^1 \cdots \varepsilon_{i_m}^m h(x_{i_1}^1, \ldots, x_{i_m}^m)\big\|_{\mathcal{H}}\Big\|_{\psi_{2/m}} \leq K \int_0^D \big(\log N(\mathcal{H}, d, \varepsilon)\big)^{m/2} d\varepsilon.$$

But this last integral is dominated by a constant times D because of Theorem 4.2 (as $\mathcal{H}$ is VC), proving the lemma.

□

This description of the proof of Theorem 4.8 constitutes an oversimplification: as we have just mentioned, the reduction to the bounded lil is a little more complicated than in the finite dimensional case and one actually needs also an inequality similar to that of Lemma 4.9 for the increments of the process which utilizes inequality (4.7) in Theorem 4.1. However it conveys the idea that the proof that has been sketched in the previous section for the finite diensional case extends to the infinite dimensional VC case only with formal changes, given Theorem 4.1.

The proof of the lil from the previous section, with only formal changes (an easy analogue of Lemma 4.9), also gives the lil for degenerate $U-$ statistics with kernels h taking values in a separable type 2 Banach space and such that $\mathbb{E}\|h\|^2 < \infty$ (Arcones and Giné, 1995). Previously, Dehling, Denker and Philipp (1986) had proved the lil for kernels h taking values in a Hilbert space and such that $\mathbb{E}\|h\|^{2+\delta} < \infty$ for some $\delta > 0$.

The lil for degenerate U-processes over VC classes does have statistical applications, for instance to determine the a.s. size of the remainder term in smooth statistical functionals. In this direction, it has been applied (Arcones and Giné, 1995) to obtain the exact a.s. size of the remainder term in the linearization of the Lynden–Bell product limit estimator of a distribution function when the data are truncated (and this, in turn, has some interesting consequences regarding the analysis of density estimates based on truncated data).

References

ARCONES, M. A. (1993). The law of the iterated logarithm for U–processes. *J. Mult. Analysis* **47** 139-151.

ARCONES, M. A. AND GINÉ, E. (1992). On the bootstrap of U and V–statistics. *Ann. Statist.* **20** 655-674.

ARCONES, M. AND GINÉ, E. (1993). Limit theorems for U–processes. *Ann. Probab.* **21** 1494–1542.

ARCONES, M. AND GINÉ, E. (1993a). On decoupling, series expansions, and tail behavior of chaos processes. *J. Theoret. Probab.* **6** 101–122.

ARCONES, M. AND GINÉ, E. (1995). On the law of the iterated logarithm for canonical U–statistics and processes. *Stoch. Proc. Appl.* **58** 217–245.

BERK, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

BONAMI, A. (1970). Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier* **20** 335-402.

BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionelles. *Ann. Inst. H. Poincaré* Sect. B. **3**, 256-261.

CARL, A. AND PAJOR, A. (1988). Gelfand numbers of operators in a Hilbert space. *Inventiones Math.* **94**, 479-504.

CHEN, Z., ARCONES, M. AND GINÉ, E. (1994). Estimators related to U–processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.* **22** 1460–1477.

CUZICK, J., GINÉ, E., ZINN, J. (1995). Laws of large numbers for quadratic forms, maxima of products and truncated sums of i.i.d. random variables. *Ann. Probab.* **23** 292–333.

DEHLING, H. (1989). Complete convergence of triangular arrays and the law of the iterated logarithm for degenerate U–statistics. *Stat. Probab. Letters* **7** 319–321.

DEHLING, H., DENKER, M. AND PHILIPP, W. (1986) A bounded law of the iterated logarithm for Hilbert space valued martingales and its application to U–statistics. *Prob. Th. Rel. Fields* **72** 111-131.

DEHLING, H. AND UTEV, S. A. (1994). The law of the iterated logarithm for degenerate U–statistics. Proceedings of the Kolmogorov Seminar, St. Pertersburg. To appear.

DUDLEY, R. M. (1967). The size of compacts subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Anal.* **1**, 290-330.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

DUDLEY, R. M. (1984). A course on empirical processes. *Lect. Notes in Math.* **1097**, 1-142. Springer , New York.

DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326.

DYNKIN, E. B. AND MANDELBAUM, A. (1983). Symmetric statistics, Poisson point processes and multiple Wiener integrals. *Ann. Statist.* **11**, 739-745.

DE LA PEÑA, V. H. (1992). Decoupling and Khintchine inequalities for U–statistics. *Ann. Probab.* **20** 1877–1892.

DE LA PEÑA, V. H. AND GINÉ, E. (1998). *An introduction to decoupling inequalities and applications.* In preparation.

DE LA PEÑA, V. H. AND MONTGOMERY–SMITH, S. (1995). Decoupling inequalities for the tail probabilities of multivariate U–statistics. *Ann. Probab.* **23** 806–816.

DE LA PEÑA, V. H., MONTGOMERY–SMITH, S. AND SZULGA, J. (1994). Contraction and decoupling inequalities for multilinear forms in U–statistics. *Ann. Probab.* **22** 1745–1765.

GINÉ, E. (1997). Lectures on some aspects of the bootstrap. In this volume.

GINÉ, E. (1997). A consequence for random polynomials of a result of de la Peña and Montgomery–Smith. Preprint.

GINÉ, E., GÖTZE, F. AND MASON, D. (1996). When is the student t-statistic asymptotically standard normal? *Ann. Probab.*, to appear.

GINÉ, E. AND ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probability* **12**, 929-989.

GINÉ, E. AND ZINN, J. (1992). Marcinkiewicz type laws of large numbers and convergence of moments for U-statistics. *Probability in Banach Spaces 8* 273–291. Birkhäuser, Boston.

GINÉ, E. AND ZINN, J. (1994). A remark on convergence in distribution of U-statistics. *Ann. Probab.* **22** 117-125.

GINÉ, E. AND ZHANG, C. H. (1996). On integrability in the LIL for degenerate U-statistics. *J. Theoret. Probab.* **9** 385-412.

GREGORY, G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5**, 110-123.

GOODMAN, V. (1996). A bounded LIL for second order U-statistics. Preprint.

HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* 17 34–43.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

KAHANE, J.P. (1968). *Some random series of functions.* D. C. Heath, Lexington, Massachusets.

KLASS, M. AND NOWICKI, K. (1996). Order of magnitude bounds for expectations of Δ_2-functions of non-negative random bilinear forms and generalized U-statistics. *Ann. Probab.*, to appear.

KRASNOSELSKY, M. A. AND RUTITSKY, Y. B. (1961). *Convex functions and Orlicz spaces.* Noordhof, 1961.

KWAPIEŃ, S. (1987). Decoupling for polynomial chaos. *Ann. Probability* **15** 1062-1071.

KWAPIÉN, S. AND WOYCZYNSKI, W. (1992). *Random series and stochastic integrals: single and multiple.* Birkhäuser, Boston.

LEDOUX, M. AND TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, Berlin.

LIU, R. V. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18**, 405-414.

MONTGOMERY-SMITH, S. (1994). Comparison of sums of independent identically dsitributed random variables. *Prob. Math. Statist.* **14** 281–285.

NOLAN, D. AND POLLARD, D. (1987). U-Processes: rates of convergence. *Ann. Statist.* **15**, 780-799.

NOLAN, D. AND POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probability* **16**, 1291-1298.

PISIER, G. (1975). Le théorème limite central et la loi du logarithme itéré dans les espaces de Banach. *Séminaire Maurey-Schwartz, 1975–76*, Exposés III et IV.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

RUBIN, M. AND VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8**, 165-170.

SERFLING, R. J. (1971). The law of the iterated logarithm for U-statistics and related von Mises functionals. *Ann. Math. Statist.* **42** 1794.

SERFLING, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

SHERMAN, R. P. (1991). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22** 439–459.

TALAGRAND, M. (1987). The Glivenko-Cantelli problem. *Ann. Probability* **15** 837-870.

TURKI–MOALLA, KEOKEB. (1996). *Processus empiriques, théorèmes limites et applications statistiques.* Doctoral Thesis, Université de Paris Sud.

VAN DER VAART, A. W. AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, Berlin and New York.

VAPNIK, V. N. AND ČERVONENKIS. A. JA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Prob. Appl.* **16**, 164-280.

VAPNIK, V. N. AND ČERVONENKIS. A. JA. (1981). Necessary and sufficient conditions for the convergence of means to their expectation. *Theory Prob. Appl.* **26**, 532-553.

ZHANG, C.-H. (1996). Strong law of large numbers for sums of products. *Ann. Probab.* **24** 1589–1615

Department of Mathematics
Department of Statistics
University of Connecticut
Storss, CT 06269
gine@uconnvm.uconn.edu

LECTURES ON SOME ASPECTS OF THE BOOTSTRAP

Evarist GINE (*)

(*) Work partially supported by NSF Grants Nos. DMS-9300725 and DMS-9625457 and by the CNRS through the Laboratoire de Mathématiques Appliquées de l'Université Blaise Pascal, Clermont-Ferrand, France.

Table of Contents

Preface 39
Chapter 1: On the bootstrap in $\mathbb{R}$ 41
1.1. Efron's bootstrap of the mean in $\mathbb{R}$ 41
1.1.1. Results 42
1.1.2. About the proofs 46
1.2. The general exchangeable bootstrap of the mean 54
1.3. The bootstrap of the mean for stationary sequences 59
1.4. The bootstrap of U–statistics 64
1.5. A general m out of n bootstrap 76
Chapter 2: On the bootstrap for empirical processes 83
2.1. Background from empirical process theory 83
2.1.1. Convergence in law of sample bounded processes 83
2.1.2. Symmetrization, Lévy type and Hoffmann–Jørgensen inequalities 90
2.1.3. Donsker classes 97
2.2. Poissonization inequalities and Efron's bootstrap in probability ... 103
2.3. The almost sure Efron's bootstrap 113
2.4. The exchangeable bootstrap 119
2.5. Uniformly pregaussian classes of functions and the bootstrap 128
2.6. Some remarks on applications 139
2.6.1. The bootstrap of the median 139
2.6.2. The delta method 143
2.6.3. M–estimators 143
References 147

Preface

Let X, X_i, $i \in \mathbb{N}$, be i.i.d.(P) and let $H_n := H_n(X_1, \ldots, X_n; \mathrm{P})$ be a root (that is, a function of both the data and their common probability law), symmetric in the x entries, whose law under P we would like to estimate. If Q_n approximates P (e.g. in the sense of convergence in distribution) and $\{X_i^*\}_{i=1}^n$ are i.i.d.(Q_n), then, under the appropriate hypotheses, the probability law of $H_n^* := H_n(X_1^*, \ldots, X_n^*; \mathrm{Q}_n)$ may approximate that of H_n. This principle would be useful if the law of H_n^* were easier to obtain or, at least, a large number of independent samples from the law Q_n could be easily produced in order to approximate it. In 1979, in a landmark paper, Efron proposed (among other things) to take $\mathrm{Q}_n = \mathrm{P}_n(\omega)$ with $\mathrm{P}_n(\omega) = \sum_{i=1}^n \delta_{X_i(\omega)}/n$, the empirical distribution, and gave this procedure the name of bootstrap. This makes very good sense because $\mathrm{P}_n(\omega) \to_w \mathrm{P}$ a.s. in great generality, and resampling from $\mathrm{P}_n(\omega)$, in our computer era, is easy. As he pointed out, the empirical distribution is not the only possible candidate for Q_n, particularly if a restricted model is assumed. For instance, suppose the variables X_i are i.i.d. $N(\theta, \sigma^2)$, θ and σ^2 unknown, and take $H_n = \sqrt{n}(\bar{X}_n - \theta)$, where $\bar{X}_n$ is the sample mean. If we take $Q_n = N(\bar{X}_n, s_n^2)$ ($s_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$, the sample variancc), then, estimating the law of H_n by that of H_n^*, which is $N(0, s_n^2)$, amounts to estimating the law of the Student t-statistic, $\sqrt{n}(\bar{X}_n - \theta)/s_n$, by $N(0,1)$. In this sense, the bootstrap has been around for some time in one form or other. Of course, the extraordinary merit of Efron's proposal consists in the formulation of a basic principle that applies in great generality, both in parametric and in non-parametric settings.

In these lectures we propose to study first order consistency of the bootstrap in the simple but important case of the mean, taken in a general sense (including e.g. the Kolmogorov–Smirnov statistic since the distribution function $F(x)$ is the mean of the process $I_{X \le x}$, $x \in \mathbb{R}$). Chapter 1 will be devoted to the bootstrap of the mean in finite dimensions, and it will also include the bootstrap of U–statistics and of very general statistics when the bootstrap sample size is reduced. In Chapter 2 we will consider the bootstrap of empirical processes (the bootstrap of the mean in infinite dimensions).

Already for the simple statistic $\bar{X}_n - \mathbb{E}X$, one encounters several features of the bootstrap that are general. For instance, there are situations when the bootstrap approximation works better than the limit law (we will touch only very briefly on this) although this is not always the case since, in particular, the regular bootstrap of the mean (i.e., mimicking the statistic for the empirical distribution instead of the original) does not work in general, neither a.s. nor in probability. The limits of validity of the bootstrap can be exactly determined in our simple situation. There are ways to modify the regular bootstrap when it does not work, such as reducing the bootstrap sample size. In fact, sampling without replacement m times from the n data, with $m/n \to 0$, works in great generality. In other instances, however, reducing the bootstrap sample size is not the only solution and another appropriate course of action may be to devise more complicated sampling plans that better mimick the original random mechanism; one of the first, very simple, instances of this is the bootstrap of degenerate U–statistics. Another way of describing the regular bootstrap of the mean is that instead of the average of the data, one takes a linear combination of the data with multinomial coefficients; then the question arises as to whether other random coefficients are also appropriate, or even better (like, e.g., in the Bayesian bootstrap); many of these different bootstraps are instances of the 'exchangeable bootstrap', and therefore we will examine this bootstrap for the sample mean. The bootstrap of the mean when the observations are α–mixing instead of independent is also studied; the moving blocks bootstrap that applies to this situation is another, more sophisticated, departure from the regular Efron's bootstrap. All these questions will be treated in Chapter 1, which will conclude with consideration of a bootstrap procedure that applies in great generality, the so called 'm out of n bootstrap without replacement'.

As indicated above, in the second part of these lectures we will study the bootstrap for the empirical process indexed by families of functions as general as possible. We will see that whenever the empirical process satisfies the central limit theorem, the bootstrap works, and conversely (in a sense). This is probably the most general statement that can be made regarding consistency of the bootstrap: both, directly and via the delta method, this result validates the bootstrap for a great wealth of statistics. If one restricts the class of functions, but still remaining within a very general situation, it can also be proved that basically any sensible model based bootstrap works for the empirical process, including the smooth bootstrap, the symmetric bootstrap, the parametric bootstrap, a 'projection onto the model' bootstrap, etc. Finally, as an application of the bootstrap for empirical processes, we will consider the bootstrap of M–estimators.

I thank my wife Rosalind for her constant support and extraordinary patience during the writing of these notes. I thank Dragan Radulović and Jon Wellner for personal comunications that made their way into these notes and for reading parts of a first draft. Thanks also to the organizers of the École d'Été de Saint–Flour for the opportunity to prepare these lectures and thanks as well to all the participants in the course for their comments, their interest and their patience. I would like to mention here that both, these lectures and the lectures on decoupling and U–statistics in this volume have their origin in a short course on these same topics that I gave at the Université de Paris-Sud (Orsay) in 1993. It is therefore a pleasure for me to also extend my gratitude to the Orsay Statistics group.

Saint–Flour, Storrs, 1996 Evarist Giné

Chapter 1: On the bootstrap in $\mathbb{R}$

In this chapter we study the consistency of the bootstrap mostly for the statistic $H_n = na_n^{-1}(\bar{X}_n - \mathbb{E}X)$. We will see that it is not always possible to bootstrap it, not even in probability, and will explore the limits of validity of the bootstrap procedure. We will also see how it can be made consistent by reducing the bootstrap sample size. This will be done in Section 1. A more general bootstrap (the 'exchangeable bootstrap') will be considered in Section 2. Section 3 is devoted to the bootstrap of the mean for mixing observations. The next section is devoted to the bootstrap of U–statistics, as a simple instance of the need to adapt the bootstrap procedure so as to mimic the main features of the original statistic. Finally, we present in Section 5 a bootstrap procedure (the m out of n bootstrap without replacement) which is consistent in very general situations.

1.1. Efron's bootstrap of the mean in $\mathbb{R}$. In this Section we let X, X_i, $i \in \mathbb{N}$, be i.i.d.(P), and set $\mathbf{X} = \{X_i\}_{i=1}^{\infty}$, $\bar{X}_n = \sum_{i=1}^n X_i/n$, $\sigma_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$, $n \in \mathbb{N}$. For each $n \in \mathbb{N}$, the bootstrap variables $X_{n,i}^*$, $i = 1, \ldots, n$, are defined to be conditionally i.i.d. given the sample $\mathbf{X}$, and with conditional law

$$\Pr\{X_{n,i}^* = X_j \big| \mathbf{X}\} = \frac{1}{n}, \quad j = 1, \ldots, n.$$

As is customary, we denote $\Pr(\cdot|\mathbf{X})$ by $\Pr^*(\cdot)$, and so we do with the conditional law ($\mathcal{L}^*$) and the conditional expectation ($\mathbb{E}^*$) given the sample $\mathbf{X}$. For instance, if U_i, $i \in \mathbb{N}$, are i.i.d. uniform on $[0,1]$, independent of $\mathbf{X}$, then a realization of the bootstrap sample is

$$X_{n,i}^* = \sum_{j=1}^n X_j I_{U_i \in A(j,n)}, \quad i = 1, \ldots, n, \quad n \in \mathbb{N},$$

where $A(j,n) = ((j-1)/n, j/n]$. Without loss of generality we can assume the U's and the X's defined as coordinates in a product probability space, the U's depending only on ω', the X's only on ω. The bootstrap sample mean is $\bar{X}_n^* = \sum_{i=1}^n X_{n,i}^*/n$.

Whereas the meaning of the limit

$$\mathcal{L}^*\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big) \to_w \mu \ \text{ a.s.} \tag{1.1}$$

is clear, namely, that for almost every ω, $\mathcal{L}^*\Big(\sqrt{n}(\bar{X}_n^*(\omega) - \bar{X}_n(\omega))\Big) \to_w \mu$, where ω is fixed and the randomness comes from the U's, the meaning of

$$\mathcal{L}^*\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big) \to_w \mu \ \text{ in pr.} \tag{1.1$'$}$$

is less clear and we explain it now. Let d be a distance metrizing convergence in law to μ in $\mathbb{R}$. For example, $d(\nu,\mu) = \sup\big\{\int f d(\mu - \nu) : \|f\|_\infty + \|f\|_{BL} \le 1\big\}$ (with $\|f\|_{BL} = \sup_{x \ne y} |f(y) - f(x)|/|y - x|$), which is in fact a countable sup; or $d_k(\nu,\mu)$, defined as the sup of the same integrals but now over over the k times differentiable functions f such that $\|f\|_\infty + \sum_{i=1}^k \|f^{(i)}\|_\infty \le 1$. Moreover, if μ has no atoms, d

can be taken to be the sup distance between distribution functions. Then, (1.1') simply means that

$$d\Big(\mathcal{L}^*(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)), \mu\Big) \to 0 \text{ in pr.} \tag{1.2}$$

This definition does not depend on the distance used since it is in fact equivalent to (1.1) holding along some subsequence of every subsequence. These definitions extend to any bootstrap functionals H_n^*.

1.1.1. Results. The following theorem asserts that the mean can be bootstrapped a.s. iff $\mathbb{E}X^2 < \infty$. The direct part is due to Bickel and Freedman (1981) and Singh (1981) and the converse to Giné and Zinn (1989).

1.1. THEOREM. *(a) If $\mathbb{E}X^2 = \sigma^2 < \infty$ then*

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^n (X_{n,i}^* - \bar{X}_n)}{\sqrt{n}}\Big) \to_w N(0,\sigma^2) \text{ a.s.} \tag{1.3}$$

(b) Conversely, if there exist random variables $c_n(\omega)$, an increasing sequence $\{a_n\}_{n=1}^\infty$ of positive numbers tending to infinity, and a random probability measure $\mu(\omega)$ non-degenerate with postive probability, such that

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^n X_{n,i}^*(\omega)}{a_n} - c_n(\omega)\Big) \to_w \mu(\omega) \text{ a.s.,} \tag{1.4}$$

then $\sigma^2 := \mathbb{E}X^2 < \infty$, $\sqrt{n}/a_n \to \sqrt{c}$ for some $c > 0$, $\mu = N(0, c\sigma^2)$ a.s. and $c_n(\omega)$ can be taken to be $c_n(\omega) = n\bar{X}_n(\omega)/a_n$

Here is the analogue for the bootstrap in probability:

1.2. THEOREM. *(a) If X is in the domain of attraction of the normal law and the constants a_n are such that $\mathcal{L}\big(\sum_{i=1}^n (X_i - \mathbb{E}X)/a_n\big) \to_w N(0,1)$, then*

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^n (X_{n,i}^* - \bar{X}_n)}{a_n}\Big) \to_w N(0,1) \text{ in pr.} \tag{1.5}$$

(b) Conversely, if

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^n X_{n,i}^*(\omega)}{a_n} - c_n(\omega)\Big) \to_w \mu(\omega) \text{ in pr.} \tag{1.6}$$

with $\mu(\omega)$ non degenerate on a set of positive probability, then there is $\sigma > 0$ such that $\mathcal{L}\big(\sum_{i=1}^n (X_i - \mathbb{E}X)/a_n\big) \to_w N(0,\sigma^2)$, in particular, X is in the domain of atraction of the normal law with admissible norming constants a_n, and $\mu = N(0,\sigma^2)$ a.s.

Part (a) of this theorem was observed by Athreya (1985) and Part (b) by Giné and Zinn (1989).

These two theorems set limits to the validity of Efron's bootstrap in the case of the mean.

Theorems 1.1 and 1.2 also hold for i.i.d. random vectors in $\mathbb{R}^d$: The Cramér–Wold's device (that is, taking linear combinations of the coordinates) reduces the vector case to $\mathbb{R}$.

We should also remark that there is convergence of all bootstrap moments in both (1.3) and (1.5), a.s. in one case, in pr. in the other. In fact, under the hypotheses of Theorem 1.2 (a), we have that for all $t > 0$,

$$\mathbb{E}^* \exp\Big\{t\sum_{i=1}^{n}(X_{n,i}^* - \bar{X}_n)/a_n\Big\} \to \mathbb{E}e^{tZ} \text{ in pr.,} \tag{1.5'}$$

where Z is $N(0,1)$, and the analogous statement holds a.s. if $\mathbb{E}X^2 < \infty$. (Arcones and Giné, 1991; previously, Bickel and Freedman, 1981, had observed that (1.3) holds with convergence of the second bootstrap moments). (1.5') justifies bootstrap estimation of variances and other functionals of the original distribution.

If $\mathbb{E}X^2 < \infty$ then $\sigma_n^2 \to \mathrm{Var}X$ a.s. by the law of large numbers, and Theorem 1.1 (a) gives that

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^n(X_{n,i}^* - \bar{X}_n)}{\sigma_n\sqrt{n}}\Big) \to_w N(0,1) \text{ a.s.} \tag{1.7}$$

Likewise, if $\mathbb{E}X^2 = \infty$ but X is in the domain of atraction of the normal law with norming constants a_n, as in (a) of Theorem 1.2, then Raikov's theorem (e.g. Gnedenko–Kolmogorov's book, or a simple standard argument) easily implies that

$$\lim_{n\to\infty}\sum_{i=1}^{n} X_i^2/a_n^2 = \lim_{n\to\infty}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2/a_n^2 = 1 \text{ in pr.}$$

and therefore, Theorem 1.2 (a) shows

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^n(X_{n,i}^* - \bar{X}_n)}{\sigma_n\sqrt{n}}\Big) \to_w N(0,1) \text{ in pr.} \tag{1.8}$$

It is an exercise to check that σ_n in equations (1.7) and (1.8) can be replaced by σ_n^*.

The two statements above about the studentized bootstrap clt also have converses. Here is the complete statement:

1.3. THEOREM. *(a) $\mathbb{E}X^2 < \infty$ if and only if the studentized bootstrap clt holds a.s., that is, iff (1.7) holds. (b) X is in the domain of attraction of the normal law iff the studentized bootstrap clt holds in probability, that is, iff (1.8) holds.*

Part (a) of this theorem was observed by Csörgő and Mason (1989) and part (b) by Hall (1990).

The exact conditions under which there exist random normings $A_n \to \infty$ and random centerings B_n such that $\{(\bar{X}_n^* - B_n)/A_n\}_{n=1}^{\infty}$ converges in law conditionally on $\mathbf{X}$, a.s. or in probability, have been determined respectively by Sepanski (1993) and Hall (1990). Besides normal limits, only Poisson limits are possible and then, the relevant side of the tail is slowly varying at infinity (so, in this case, X is not even in the Feller class). We do not discuss the Poisson limit situation, which corresponds to $\mathbb{E}|X| = \infty$ and does not relate to the bootstrap of the mean.

We will not discuss either p–stable domains of attraction with $p \le 1$, for the same reason. Regarding domains of attraction the following two results essentially tell the story: to have consistency of the bootstrap in this case, we must reduce the bootstrap sample size.

1.4. THEOREM. *Suppose that X is in the domain of atraction of a non–degenerate p-stable law μ, $1 < p < 2$, concretely, assume $\mathcal{L}\big(\sum_{i=1}^n (X_i - \mathbb{E}X)/a_n\big) \to_w \mu$ for some constants $a_n \nearrow \infty$. Let $m_n \nearrow \infty$. Then,*

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^{m_n}(X_{n,i}^* - \bar{X}_n)}{a_{m_n}}\Big) \to_w \mu \ \text{ in pr.} \tag{1.9}$$

if and only if

$$\frac{m_n}{n} \to 0,$$

The direct part of this theorem is due to Athreya (1985) and the converse was observed in Arcones and Giné, (1989). As with Theorems 1.1 and 1.2, conditional weak convergence in (1.9) can be strengthened to coditional convergence of bootstrap moments, but here only short of the p–th moment, that is, we have

$$\mathbb{E}^*\Big|\frac{\sum_{i=1}^{m_n}(X_{n,i}^* - \bar{X}_n)}{a_{m_n}}\Big|^\alpha \to \int |x|^\alpha d\mu(x) \ \text{ in pr.,} \quad 0 < \alpha < p \tag{1.9'}$$

(Arcones and Giné, 1991).

1.5. THEOREM. *Let X be in the domain of atraction of a non–degenerate p-stable law μ, $1 < p \le 2$, concretely, assume $\mathcal{L}\big(\sum_{i=1}^n (X_i - \mathbb{E}X)/a_n\big) \to_w \mu$ for some constants $a_n \nearrow \infty$, and let $m_n \nearrow \infty$ be a regular sequence in the sense that $\liminf_{n\to\infty} m_n/m_{2n} > 0$. Then,*

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^{m_n}(X_{n,i}^* - \bar{X}_n)}{a_{m_n}}\Big) \to_w \mu \ \text{ a.s.} \tag{1.10}$$

if

$$\frac{m_n \log\log n}{n} \to 0, \tag{1.11}$$

and (1.10) does not hold if $\liminf_{n\to\infty}(m_n \log\log n)/n > 0$.

This theorem is due to Arcones and Giné (1989).

Self–normalization is also possible in the previous two theorems. Arcones and Giné (1991) show that for X in the domain of attraction of a p–stable law, $1 < p \le 2$, and for $m_n/n \to 0$,

$$w - \lim_{n\to\infty} \mathcal{L}^*\Bigg[\frac{\sum_{i=1}^{m_n}(X_{n,i}^* - \bar{X}_n)}{\big(\sum_{i=1}^{m_n} X_{n,i}^{*2}\big)^{1/2}}\Bigg] = w - \lim_{n\to\infty} \mathcal{L}\Bigg[\frac{\sum_{i=1}^{n}(X_i - \mathbb{E}X)}{\big(\sum_{i=1}^{n} X_i^2\big)^{1/2}}\Bigg] \ \text{ in pr.} \tag{1.12}$$

Their proof also shows that this holds a.s. if the sequence m_n is regular and satisfies (1.11). Deheuvels, Mason and Shorack (1992) have another approach to a result similar to (1.12), but with the norming in the bootstrap quantity depending on the X's, as in (1.8). They also prove the following theorem for the bootstrap of the maximum of i.i.d. uniform random variables, one of the first examples of failure of Efron's bootstrap with $m_n = n$ (Bickel and Freedman, 1981). Let X be uniform on $(0, \theta)$. Then, it is easy to see that

$$\mathcal{L}\Big(\frac{n\big(\theta - \max_{1\le i\le n}|X_i|\big)}{\theta}\Big) \to_w \mu, \tag{1.13}$$

where μ is the exponential distribution with unit parameter. Here is the Deheuvels et al. bootstrap version of this limit.

1.6. THEOREM. *Let X and μ be as in (1.13). Then, if $m_n/n \to 0$ we have*

$$\mathcal{L}^*\Big(\frac{m_n\big(\max_{1\le i\le n}|X_i| - \max_{1\le i\le m_n}|X^*_{n,i}|\big)}{\max_{1\le i\le n}|X_i|}\Big) \to_w \mu \text{ in pr.,} \tag{1.14}$$

and if $(m_n \log\log n)/n \to 0$ then the limit in (1.14) holds a.s.

Back to Theorem 1.3 (a), one may ask how good is $\mathcal{L}^*\big(\sum_{i=1}^n (X^*_{n,i} - \bar{X}_n)/\sigma_n\sqrt{n}\big)$ as an approximation to $\mathcal{L}\big(\sum_{i=1}^n (X_i - \mathbb{E}X)/\sigma\sqrt{n}\big)$. This has been thoroughly studied, starting with Singh (1981), who showed that if $\mathbb{E}|X|^3 < \infty$ and X is skewed, the bootstrap approximation may be better than the normal approximation. Hall (1988) shows that in case $\mathbb{E}|X|^3 = \infty$ the bootstrap approximation can actually do worse. We will present here a weaker and simpler result on direct comparison of the bootstrap and the original distributions which also indicates how the bootstrap improves on the normal approximation for skewed random variables with finite third moment. D. Radulović showed this to me and I thank him for allowing me to report on his arguments in these lectures. For probability measures μ, ν on $\mathbb{R}$, define

$$d_4(\mu,\nu) = \sup\Big\{\Big|\int f d(\mu-\nu)\Big| : \|f\|_\infty \le 1, \|f^{(i)}\|_\infty \le 1,\ 1 \le i \le 3\Big\},$$

a distance that metrizes weak convergence of probability measures on $\mathbb{R}$. It is easy to see, using a Lindeberg type argument, that if $\mathbb{E}|X|^3 < \infty$ and, without loss of generality, $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$, then

$$d_4\Big[\mathcal{L}\Big(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\Big), N(0,1)\Big] = O\big(n^{-1/2}\big), \tag{1.15}$$

and that this cannot in general be improved if $\mathbb{E}X^3 \neq 0$. For the bootstrap, we have:

1.7. PROPOSITION. *If $\mathbb{E}|X|^3 < \infty$ then*

$$d_4\Big[\mathcal{L}^*\Big(\frac{\sum_{i=1}^n (X^*_{n,i} - \bar{X}_n)}{\sigma_n\sqrt{n}}\Big), \mathcal{L}\Big(\frac{\sum_{i=1}^n (X_i - \mathbb{E}X)}{\sigma\sqrt{n}}\Big)\Big] = o\big(n^{-1/2}\big) \quad a.s. \tag{1.16}$$

1.1.2. About the proofs. It is worth noting that as long as the norming constants a_{m_n} tend to infinity, the triangular array $\{X_{n,i}^* : i = 1, \ldots, m_n\}_{n=1}^{\infty}$ is a.s. infinitesimal:

$$\max_{i \le m_n} \Pr^*\{|X_{n,i}^*| > \delta a_{m_n}\} = \Pr^*\{|X_{n,1}^*| > \delta a_{m_n}\} = \frac{1}{n}\sum_{j=1}^{n} I_{|X_j| > \delta a_{m_n}},$$

and, by the law of large numbers, the limsup of this last sum is a.s. bounded by $\mathbb{E}|X|I_{|X|>c}$ for all $c > 0$; letting $c \to \infty$ along a countable sequence gives the a.s. infinitesimality. Then, the proofs of Theorems 1.1 to 1.5 consist in i) showing bootstrap convergence by just checking the conditions for the general normal (or stable) convergence criterion for infinitesimal arrays and ii) applying the converse part of this criterion to infer properties on the distribution of X from bootstrap convergence. This program provides relatively simple proofs, except for Theorem 1.5, where one must proceed, roughly speaking, as in the proof of the LIL. Similar comments apply to Theorem 1.6. (1.12) just follows from the bootstrap of a stable convergence theorem in $\mathbb{R}^2$, and its proof is not different from that of the direct parts of Theorems 1.4 and 1.5. Finally, Proposition 1.7 is proved by a Lindeberg type argument applied to a simple coupling between the bootstrap and the original statistics. We will give complete proofs of Theorem 1.1 and Proposition 1.7, and then indicate parts of proofs of the other results.

1.7. PROOF OF THEOREM 1.1 a). We will prove that if $\mathbb{E}X^2 < \infty$ and if $m_n \to \infty$ then

$$\mathcal{L}^*\Big(\frac{\sum_{i=1}^{m_n}(X_{n,i}^* - \bar{X}_n)}{\sqrt{m_n}}\Big) \to_w N(0, \sigma^2) \;\text{ a.s.} \tag{1.3'}$$

By the general criterion for normal convergence (e.g., Araujo and Giné, 1980, Cor. 2.4.8, p.63), it suffices to prove

$$m_n \Pr^*\{|X_{n,1}^*| > \delta m_n^{1/2}\} \to 0 \;\text{ a.s.} \tag{1.17}$$

for all $\delta > 0$,

$$\mathrm{Var}^*\big(X_{n,1}^* I_{|X_{n,1}^*| \le m_n^{1/2}}\big) \to \mathrm{Var}X \;\text{ a.s.} \tag{1.18}$$

and

$$m_n^{1/2}\mathbb{E}^* X_{n,1}^* I_{|X_{n,1}^*| > m_n^{1/2}} \to 0 \;\text{ a.s.} \tag{1.19}$$

(Then, one makes the set of measure 1 where convergence takes place in (1.17) independent of δ just by taking a countable dense set of δ's, which is all that is needed.) The basic observation is that, since $\mathbb{E}X^2 < \infty$, the law of large numbers yields

$$\frac{1}{n}\sum_{j=1}^{n} |X_j|^p I_{|X_j| > \delta m_n^{1/2}} \to 0 \;\text{ a.s.},\; 0 < p \le 2,\; \delta > 0. \tag{1.20}$$

(Replace $\delta m_n^{1/2}$ by c rational and take limits first as $n \to \infty$ and then as $c \to \infty$.) Then, (1.17) and (1.19) follow immediately because

$$m_n \Pr^*\{|X_{n,1}^*| > \delta m_n^{1/2}\} = \frac{m_n}{n}\sum_{j=1}^{n} I_{|X_j| > \delta m_n^{1/2}} \le \frac{\delta^{-2}}{n}\sum_{j=1}^{n} X_j^2 I_{|X_j| > \delta m_n^{1/2}}$$

and

$$m_n^{1/2}\left|\mathbb{E}^* X_{n,1}^* I_{|X_{n,1}^*|>m_n^{1/2}}\right| \le \frac{\delta^{-1}}{n}\sum_{j=1}^n X_j^2 I_{|X_j|>m_n^{1/2}}.$$

As for (1.18), using (1.20) and the law of large numbers once more, we obtain:

$$\begin{aligned}\mathrm{Var}^*\left(X_{n,1}^* I_{|X_{n,1}^*|\le m_n^{1/2}}\right) &= \frac{1}{n}\sum_{j=1}^n X_j^2 I_{|X_j|\le m_n^{1/2}} - \left[\frac{1}{n}\sum_{j=1}^n X_j I_{|X_j|\le m_n^{1/2}}\right]^2 \\ &\simeq \frac{1}{n}\sum_{j=1}^n X_j^2 - \left[\frac{1}{n}\sum_{j=1}^n X_j\right]^2 \to \mathrm{Var}X \quad \text{a.s.}\end{aligned}$$

□

Part (a) of Theorem 1.3 follows from Part (a) of Theorem 1.1 because $\sigma_n^2 \to \mathrm{Var}X$ a.s. by the law of large numbers.

Next we prove part (b) of Theorems 1.1 and 1.3. This requires some preparation. The first lemma tells us that the limit in (1.4) must be normal.

1.8. LEMMA. *If the limit (1.4) holds a.s. then the Lévy measure of the limit $\mu(\omega)$ is zero a.s. and*

$$\sum_{j=1}^n I_{|X_j|>\lambda a_n} = 0 \ \text{ eventually a.s. and } n\Pr\{|X|>\lambda a_n\}\to 0 \tag{1.21}$$

for all $\lambda > 0$.

PROOF. By a.s. infinitesimality, $\mu(\omega)$ is a.s. infinitely divisible. Let $\pi(\omega)$ be its Lévy measure. First we will show that π is non-random (a constant measure on a set of probability 1). By the converse clt (e.g. Araujo and Giné, 1980, Chapter 2), with probability one

$$n\mathcal{L}^*\left(X_{n,1}^*(\omega)\right)\big|_{|x|>\delta} \to_w \pi(\omega)\big|_{|x|>\delta} \tag{1.22}$$

for all $\delta = \delta(\omega)$ such that $\pi(\omega)\{\delta, -\delta\} = 0$. Since we cannot control the continuity points of the possibly uncountable number of masures $\pi(\omega)$, we must smooth these measures out. For each $\delta \in \mathbb{Q}^+$, let h_δ be a bounded, even, continuous function identically zero on $[0, \delta/2]$ and identically one on $[\delta, \infty)$, and set $\pi_\delta(dx, \omega) = h_\delta(x)\pi(dx, \omega)$. Then, (1.22) implies

$$\sum_{j=1}^n h_\delta(X_i/a_n)\delta_{X_i/a_n} \to_w \pi_\delta \ \text{ a.s.}$$

Let $\mathcal{F}$ be a countable measure determining class of real bounded continuous functions on $\mathbb{R}$ (e.g. $\mathcal{F} = \{\cos tx, \sin tx : t \in \mathbb{Q}\}$). Then, the previous limit gives that, on a set of measure one,

$$\sum_{j=1}^n h_\delta(X_i/a_n) f(X_i/a_n) \to_w \int f\,d\pi_\delta \ \text{ a.s. for all } f\in\mathcal{F} \text{ and } \delta\in\mathbb{Q}^+.$$

Since $a_n \to \infty$ and the summands are bounded, the sum of the first k terms at the left hand side is a.s. eventually zero for all k, and therefore, the variables $\int f d\pi_\delta$ are all measurable for the tail σ-algebra of the sequence $\mathbf{X}$. Since there are only a countable number of them, there is a common set of probability one where they are all constant, by the zero–one law. Since $\mathcal{F}$ is measure determining, $\pi_\delta(\omega)$ is a constant measure on this set for all δ. Hence, there is a Lévy measure π such that $\pi_\delta(\omega)\big|_{|x|>\delta} = \pi\big|_{|x|>\delta}$ a.s. for all $\delta > 0$. Let $\bar{\pi} = \pi \circ |x|^{-1}$. (1.22) then becomes

$$\sum_{j=1}^{n} I_{|X_j|>\lambda a_n} \to \bar{\pi}(\lambda, \infty) \tag{1.23}$$

for all $\lambda > 0$ of continuity for $\bar{\pi}$, in particular on a countable dense set D of $\mathbb{R}^+$. (1.23) implies that $\bar{\pi}(\lambda, \infty)$ takes on only non–negative integer values for all $\lambda \in D$. Suppose $\bar{\pi}(\lambda, \infty) = r \neq 0$ for some $\lambda > 0$. Then,

$$\sum_{j=1}^{n} I_{|X_j|>\lambda a_n} = r \quad \text{eventually a.s.,}$$

which in particular implies

$$\lim_{n\to\infty} \Pr\Big\{\sum_{j=1}^{n} I_{|X_j|>\lambda a_n} = r\Big\} = 1. \tag{1.24}$$

On the other hand, there is enough uniform integrability in (1.23) to have convergence of expected values (e.g., by Hoffmann–Jørgensen's inequality: see Lemma 1.12, Chapter 2) so that

$$n \Pr\{|X| > \lambda a_n\} \to r$$

and therefore,

$$\lim_{n\to\infty} \Pr\Big\{\sum_{j=1}^{n} I_{|X_j|>\lambda a_n} = e^{-r}\frac{r^r}{r!} < 1,$$

contradiction with (1.24). Hence, $r = 0$, that is, $\pi = 0$ and the limits (1.21) hold true.

□

1.9. LEMMA. *If the limit (1.4) holds a.s. then, the standard deviation $\sigma(\omega)$ of the normal component of the limit measure $\mu(\omega)$ in (1.4) is a.s. a constant σ different from zero. If $\mathbb{E}X^2 < \infty$, then $n/a_n^2 \to \sigma^2/\mathrm{Var}X$, whereas if $\mathbb{E}X^2 = \infty$, then*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i^2}{a_n^2} = \sigma^2 \quad \text{a.s.} \tag{1.25}$$

PROOF. The first limit in (1.21) gives that for all $\lambda > 0$ and $p \in \mathbb{R}$,

$$\sum_{j=1}^{n} |X_j|^p I_{|X_j|>\lambda} = 0 \quad \text{eventually a.s.} \tag{1.26}$$

and therefore, we can 'untruncate' in the necessary condition for the clt in terms of truncated variances, which then becomes

$$\lim_{n\to\infty}\left[\frac{\sum_{j=1}^n X_j^2(\omega)}{a_n^2}-\left(\frac{\sum_{j=1}^n X_j(\omega)}{a_n\sqrt{n}}\right)^2\right]=\sigma^2(\omega)\quad \text{a.s.} \tag{1.27}$$

Since $a_n\to\infty$, (1.27) shows that $\sigma^2(\omega)$ is a tail random variable, thus a.s. constant, say σ^2. Since $\mu(\omega)$ is not degenerate with positive probability and $\pi=0$ a.s., it follows that $\sigma\neq 0$. And of course, the fact that $\mu(\omega)$ exists implies that $\sigma<\infty$. If $\mathbb{E}X^2<\infty$, the conclusion of the lemma follows from (1.27) and the law of large numbers. Let us now assume $\mathbb{E}X^2=\infty$. Then, we claim

$$\lim_{n\to\infty}\frac{\left[\sum_{j=1}^n |X_j|/n\right]^2}{\sum_{j=1}^n X_j^2/n}=0\quad \text{a.s.} \tag{1.28}$$

(1.28) follows from the Paley–Zygmund argument applied to the empirical measure, that is, from the following self–evident inequalities,

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n|X_j| &\le a+\frac{1}{n}\sum_{j=1}^n|X_j|I_{|X_j|>a}\\ &\le a+\Big(\frac{1}{n}\sum_{j=1}^n X_j^2\Big)^{1/2}\Big(\frac{1}{n}\sum_{j=1}^n I_{|X_j|>a}\Big)^{1/2},\end{aligned}$$

upon dividing by $\left(\sum_{j=1}^n X_j^2/n\right)^{1/2}$ and then taking limits first as $n\to\infty$ and then as $a\to\infty$ (use the Kolmogorov law of large numbers, both for finite and infinite expectations, when taking these limits). Inequality (1.28) was stated as a lemma (with other powers, besides 1 and 2) in Giné and Zinn (1989) and J. Chen informed us some time after publication that he and H. Rubin already had noticed this inequality, with a different proof, in Chen and Rubin (1984). Zinn and I hereby acknowledge their priority (we did not have the opportunity to publish an acknowledgement before). Back to the proof, it is clear that (1.25) follows from (1.27) and (1.28).

□

Had we assumed the sequence a_n to be regular (precisely, a_n/n monotone, and $a_n/n^{1/r}$ monotone increasing for some $r<2$), a refinement of Feller of the Marcinkiewicz law of large numbers (e.g., Stout, 1974, page 132) would automatically imply that (1.25) with $0<\sigma^2<\infty$ and $\mathbb{E}X^2=\infty$ can not both be true at the same time, which would conclude the proof of Theorem 1.1, part b, by contradiction. But without any conditions, an extra argument is needed. Here it is:

1.10. Conclusion of the proof of Theorem 1.1 b). Assume $\mathbb{E}X^2=\infty$. By (1.26), we can truncate the variables X_j at a_n in (1.25), Lemma 1.9, and then take expectations of the resulting sums –by Hoffmann–Jørgensen's inequality (see Chapter 2, Lemma 1.12 below), on account of the boundedness of the summands– to obtain

$$\frac{n}{a_n^2}\mathbb{E}X^2I_{|X|\le a_n}\to\sigma^2\neq 0. \tag{1.29}$$

This inequality implies, by the monotonicity of the a_n's, that

$$\limsup_{n\to\infty} \frac{n}{a_n^2} \max_{k\le n} \frac{a_k^2}{k} < \infty.$$

Since $n/a_n^2 \to 0$, (1.29) also implies that there exists $r_n \to \infty$ such that

$$\delta_n := \frac{n}{a_n^2} \max_{k\le r_n} a_k^2 \to 0.$$

Also, since by the first limit in (1.21), Lemma 1.8, $|X_n|/a_n < 1$ a.s., the Borel–Cantelli lemma gives $\sum_{n=1}^{\infty} \Pr\{|X| > a_n\} < \infty$, and therefore,

$$\sum_{k=r_n}^{\infty} k \Pr\{a_{k-1} < |X| \le a_k\} \to 0.$$

Using (1.29) once more, together with the last three limits, we obtain

$$\begin{aligned}
\sigma^2 &\le \lim_{n\to\infty} \frac{n}{a_n^2} \sum_{k=1}^{n} a_k^2 \Pr\{a_{k-1} < |X| \le a_k\} \\
&\le \lim_{n\to\infty} \delta_n + \limsup_{n\to\infty} \frac{n}{a_n^2} \left[\max_{k\le n} \frac{a_k^2}{k}\right] \sum_{k=r_n}^{\infty} k \Pr\{a_{k-1} < |X| \le a_k\} \to 0.
\end{aligned}$$

This contradicts $\sigma^2 \neq 0$ and therefore we must have $\mathbb{E}X^2 < \infty$ and $a_n \simeq \sqrt{n}$.

□

1.11. PROOF OF THEOREM 1.3 a). It has already been observed above that $\mathbb{E}X^2 < \infty$ implies (1.7). Assume (1.7) holds. Since $\mathbb{E}^*\left[(X_{n,1}^* - \bar{X}_n)/\sigma_n\sqrt{n}\right]^2 = 1/n$, the random variable at the left of (1.7) is conditionally a.s. the n-th row sum of an infinitesimal triangular array of independent variables. Then, a.s. asymptotic normality implies (by the converse clt)

$$n\Pr^*\left\{\frac{|X_{n,1}^* - \bar{X}_n|}{\sigma_n\sqrt{n}} > \varepsilon\right\} \to 0 \ \text{ a.s.},$$

for all $\varepsilon > 0$, that is,

$$\sum_{j=1}^{n} I_{|X_j - \bar{X}_n|/\sigma_n\sqrt{n} > \varepsilon} \to 0 \ \text{ a.s.}$$

for all $\varepsilon > 0$. This shows

$$\max_{1\le j\le n} \frac{|X_j - \bar{X}_n|}{\sigma_n\sqrt{n}} \to 0 \ \text{ a.s.} \tag{1.30}$$

If $\mathbb{E}X^2 < \infty$ there is nothing to prove. If $\mathbb{E}X^2 = \infty$ then we can combine (1.28) with (1.30) to conclude that

$$\max_{1\le j\le n} \frac{|X_j|}{\sqrt{\sum_{i=1}^{n} X_i^2}} \to 0 \ \text{ a.s.}$$

But this can only happen if $\mathbb{E}X^2 < \infty$ by a result of Kesten (1971).

□

1.12. REMARK ON THE PROOFS OF THEOREM 1.2 AND THEOREM 1.3 b). By a subsequence argument, Theorem 1.2 a) follows by showing that the necessary conditions for normal convergence (the analogues of (1.17)-(1.19)) hold in probability. Then, to show that these limits hold, one uses the fact that, by the converse clt, $n\Pr\{|X| > \delta a_n\} \to 0$ and $na_n^{-2}\mathbb{E}X^2 I_{|X|\leq a_n} \to 1$. For instance the analogue of (1.17) that must be proved in this case is

$$\sum_{i=1}^{n} I_{|X|>\delta a_n} \to 0 \text{ in pr.}$$

and for this, one just notices that

$$\mathbb{E}\sum_{i=1}^{n} I_{|X|>\delta a_n} = n\Pr\{|X| > \delta a_n\} \to 0.$$

It is not difficult to complete the proof. To prove the converse (part b), working as in the proof of Theorem 1.1 along subsequences, one obtains that, assuming $\mathbb{E}X^2 = \infty$, $\sum_{i=1}^{n} X_i^2/a_n^2 \to \sigma^2 \neq 0$ in probability. Then, the converse weak law of large numbers gives, in particular, $n\Pr\{|X| > \delta a_n\} \to 0$ and $na_n^{-2}\mathbb{E}X^2 I_{|X|\leq\delta a_n} \to \sigma^2$ for all $\delta > 0$ which, by the direct clt, implies that X is in the domain of atraction of the normal law, with norming constants a_n. Theorme 1.3 b), direct part, follows from 1.2 a) and Raikov's theorem, as observed above. As for part b, the studentized bootstrap clt in pr. gives, following the proof of Theorem 1.3 a) along subsequences, that the limit (1.30) holds in probability, a condition that is known to be necessary and suficient for X to belong to the domain of attraction of the normal law (O'Brien, 1980).

1.13. CONVERGENCE OF MOMENTS IN THEOREM 1.1. We will only show that if $\mathrm{Var}X = 1$ and Z is $N(0,1)$, then

$$\mathbb{E}^* \exp\Big\{t\Big|\sum_{i=1}^{n}(X_{n,i}^* - \bar{X}_n)/\sqrt{n}\Big|\Big\} \to \mathbb{E}\exp\{t|Z|\} \quad a.s. \tag{1.3'}$$

for all $t > 0$ (since (1.5') has a similar proof). We assume without loss of generality that $\mathbb{E}X = 0$ and that Z is independent of $\mathbf{X}$. Let $\{\varepsilon_i\}$ be a Rademacher sequence independent of $\{X_{n,i}^*\}$. Then, convexity, the properties of Rademacher variables and the facts that $\max_{j\leq n} X_j^2/n \to 0$ a.s. and $\sum_{j=1}^{n} X_j^2/n \to 1$ a.s., give

$$\begin{aligned}
\mathbb{E}^* \exp\Big\{t\Big|\frac{\sum_{i=1}^{n}(X_{n,i}^* - \bar{X}_n)}{\sqrt{n}}\Big|\Big\} &\leq \mathbb{E}^* \exp\Big\{2t\Big|\frac{\sum_{i=1}^{n}\varepsilon_i X_{n,i}^*}{\sqrt{n}}\Big|\Big\} \\
&\leq 2\mathbb{E}^* \exp\Big\{4t^2\frac{\sum_{i=1}^{n} X_{n,i}^{*2}}{n}\Big\} \\
&= 2\Big[\frac{1}{n}\sum_{j=1}^{n}\exp\Big\{\frac{4t^2X_j^2}{n}\Big\}\Big]^n \to 2e^{4t^2}.
\end{aligned}$$

That is, the sequence of bootstrap exponential moments is a.e. a bounded sequence. By the bootstrap clt and the continuous mapping theorem, there is a.s. weak convergence of the conditional laws of

$$\exp\left\{t\left|\frac{\sum_{i=1}^{n}(X_{n,i}^{*}-\bar{X}_n)}{\sqrt{n}}\right|\right\}$$

to the law of $\exp\{t|Z|\}$ and, therefore, (1.3') follows by a.s. uniform integrability.

□

1.14. REMARKS ON THEOREMS 1.4 AND 1.5, THE STABLE CASE. A simple sufficiency proof of Theorem 1.4 consists in routinely checking the conditions of the clt for triangular arrays, as in the normal case. Just to show how the limit $m_n/n \to 0$ gets into the picture, we will check one of these three conditions, concretely

$$m_n \Pr^*\{X_{n,i}^* > \delta a_{m_n}\} \to c\delta^p \text{ in pr.} \tag{1.31}$$

for all $\delta > 0$, assuming that X is in the doa of a stable law with norming constants a_n, thus, in particular, assuming the necessary condition $n \Pr\{|X| > \delta a_n\} \to c\delta^p$ for all $\delta > 0$. The expected value of the left side of (1.31) satisfies

$$\mathbb{E}\Big[m_n \Pr^*\{X_{n,i}^* > \delta a_{m_n}\}\Big] = \mathbb{E}\Big[\frac{m_n}{n}\sum_{j=1}^{n} I_{|X_j|>\delta a_{m_n}}\Big] = m_n \Pr\{|X| > \delta a_{m_n}\} \to c\delta^p$$

whereas its variance tends to zero:

$$\begin{aligned}
\mathbb{E}\Big(\frac{m_n}{n}\sum_{j=1}^{n}(I_{|X_j|>\delta a_{m_n}} - \Pr\{|X| > \delta a_{m_n}\})\Big)^2 & \\
&= \frac{m_n^2}{n}\mathbb{E}\Big(I_{|X_j|>\delta a_{m_n}} - \Pr\{|X| > \delta a_{m_n}\}\Big)^2 \\
&= \frac{m_n^2}{n}(1 - \Pr\{|X| > \delta a_{m_n}\}) \Pr\{|X| > \delta a_{m_n}\} \\
&\simeq \Big(\frac{m_n}{n}\Big) m_n \Pr\{|X| > \delta a_{m_n}\} \to 0
\end{aligned}$$

beacuse $m_n/n \to 0$ and $m_n \Pr\{|X| > \delta a_{m_n}\}$ is bounded. This proves (1.31). The remaining conditions for stable convergence are proved similarly. For the converse, just note that, if $m_{n'}/n' \to c > 0$ for some subsequence $\{n'\}$, then the argument in the second part of the proof of Lemma 1.8 shows that the Lévy measure of the limit must be zero, which is not the case for a stable limit. The proof of Theorem 1.5 is more involved. Here we describe only part of the direct proof. The statement to be proved corresponding to (1.31), is

$$\frac{m_n}{n}\sum_{j=1}^{n}(I_{|X_j|>\delta a_{m_n}} - \Pr\{|X| > \delta a_{m_n}\}) \to 0 \text{ a.s.}$$

If the sequence m_n is regular, it turns out that we can block, symmetrize and apply a maximal inequality as in the usual proof of the lil, to conclude that it suffices to show

$$\sum_{n=1}^{\infty} \Pr\Big\{ \frac{m_{2^n}}{2^n} \Big| \sum_{j=1}^{2^n} (I_{|X_i|>\delta a_{2^n}} - \Pr\{|X| > \delta a_{m_n}\}) \Big| > \varepsilon \Big\} < \infty,$$

and Prohorov's exponential inequality shows that this is the case if $(m_n \log\log n)/n \to 0$. We omit the details.

The proof of Theorem 1.6 is also omitted, and we turn now our attention to the proof of Proposition 1.7. This proof will illustrate how the bootstrap conditional distributions keep some of the skewness of the original distributions (this is not the case, obviously, for the normal approximation).

1.15. PROOF OF PROPOSITION 1.7. We assume without loss of generality that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$ (besides the crucial hypothesis $\mathbb{E}|X|^3 < \infty$). By translation invariance of the family of functions in the definition of d_4, we have

$$d_4\Big[\mathcal{L}^*\Big(\frac{\sum_{i=1}^n (X_{n,i}^* - \bar{X}_n)}{\sigma_n\sqrt{n}}\Big), \mathcal{L}\Big(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\Big)\Big] \le n d_4\Big[\mathcal{L}^*\Big(\frac{X_{n,1}^* - \bar{X}_n}{\sigma_n\sqrt{n}}\Big), \mathcal{L}\Big(\frac{X}{\sqrt{n}}\Big)\Big]. \tag{1.32}$$

Let now f be as in the definition of d_4, that is, $\|f\|_\infty \le 1$ and $\|f^{(i)}\|_\infty \le 1$, $1 \le i \le 4$. The first and second conditional moments of $(X_{n,1}^* - \bar{X}_n)/(\sigma_n\sqrt{n})$ are respectively 0 and $1/n$, just as the first and secon moments of $X/\sqrt{n}$. McLaurin's development of f then gives

$$\begin{aligned}\Big|\mathbb{E}^* f\Big(\frac{X_{n,1}^* - \bar{X}_n}{\sigma_n\sqrt{n}}\Big) - \mathbb{E}f\Big(\frac{X}{\sqrt{n}}\Big)\Big| &\le \frac{1}{6\sigma_n^3 n^{3/2}}\Big|\mathbb{E}^*(X_{n,1}^* - \bar{X}_n)^3 - \sigma_n^3\mathbb{E}X^3\Big| \\ &\quad + \frac{1}{6\sigma_n^3 n^{3/2}}\mathbb{E}^*\Big[|f'''(\eta_1) - f'''(0)||X_{n,1}^* - \bar{X}_n|^3\Big] \\ &\quad + \frac{1}{6\sigma_n^3 n^{3/2}}\mathbb{E}\Big[|f'''(\eta_2) - f'''(0)||X|^3\Big] \\ &:= I_n + II_n + III_n,\end{aligned} \tag{1.33}$$

η_i, $i = 1, 2$, being random variables respectively between 0 and $(X_{n,1}^* - \bar{X}_n)/(\sigma_n\sqrt{n})$, and between 0 and $X/\sqrt{n}$. Now, since $\sigma_n \to 1$ a.s., $\sup_{\|f'''\|_\infty \le 1} |f'''(\eta_i) - f'''(0)| \le 2$ and $\sup_{\|f^{(4)}\|_\infty \le 1} |f'''(\eta_i) - f'''(0)| \le |\eta_i| \to 0$ a.s., it follows that

$$n^{3/2} \sup_{\|f'''\|_\infty \le 1, \|f^{(4)}\|_\infty \le 1} III_n \to 0 \tag{1.35}$$

and, by the law of large numbers, that

$$n^{3/2} \sup_{\|f'''\|_\infty \le 1, \|f^{(4)}\|_\infty \le 1} II_n \to 0 \quad \text{a.s.} \tag{1.36}$$

I_n is the crucial term in (1.33). In the analoguous proof of normal approximation, the term I_n would just be $\mathbb{E}X^3/(6n^{3/2})$ but here it is of a smaller order because

$$(\sigma_n^3 - 1)\mathbb{E}X^3 \to 0 \quad \text{a.s.}$$

and, by the law of large numbers,

$$\left|\mathbb{E}^*(X^*_{n,1}-\bar{X}_n)^3-\mathbb{E}X^3\right|=\left|\frac{1}{n}\sum_{j=1}^{n}(X_i^3-\mathbb{E}X^3)-3\bar{X}_n\frac{\sum_{i=1}^n X_i^2}{n}+2\bar{X}_n^3\right|\to 0 \text{ a.s.},$$

which gives

$$n^{3/2}\sup_{\|f'''\|_\infty\le 1,\|f^{(4)}\|_\infty\le 1} I_n\to 0 \text{ a.s.} \tag{1.37}$$

combining the estimates (1.35)–(1.37) with (1.34) and then with (1.35), gives

$$d_4\left[\mathcal{L}^*\Big(\frac{\sum_{i=1}^n(X^*_{n,i}-\bar{X}_n)}{\sigma_n\sqrt{n}}\Big),\mathcal{L}\Big(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\Big)\right]=o\big(n^{-1/2}\big),$$

proving the proposition.

□

1.2. The general exchangeable bootstrap of the mean. For each $n\in\mathbb{N}$, let $w_n=(w_n(1),\dots,w_n(n))$ be a vector of n exchangeable random variables independent from the sequence $\{X_i\}$ and satisfying the following conditions:

E1. $w_n(j)\ge 0$ for all n and j, and $\sum_{j=1}^n w_n(j)=1$;

E2. Var $w_n(1)=O(n^{-2})$.

E3. $\max_{1\le j\le n}\sqrt{n}|w_n(j)-1/n|\to_P 0$.

E4. $n\sum_{j=1}^n\big(w_n(j)-1/n\big)^2\to_P c^2>0$.

Define

$$\bar{X}_n^*=\sum_{j=1}^n w_n(j)X_j$$

and take this as the bootstrap of the mean $\bar{X}_n$. Newton and Mason (1992) proved the following theorem.

2.1. THEOREM. *If $\mathbb{E}X^2=\sigma^2<\infty$ and the weights w_n are independent from the sample $\{X_i\}$ and satisfy conditions E.1 to E.4, then*

$$\mathcal{L}\Big(\sqrt{n}(\bar{X}_n^*-\bar{X}_n)\Big|\mathbf{X}\Big)\to_w N\big(0,c^2\sigma^2\big) \text{ a.s.}$$

Their original proof is based on a clt for exchangeable random variables due to Hájek. Here, following Arenal and Matrán (1996), we will deduce it simply from the usual Lindeberg clt for independent random variables, as follows: we will prove first convergence of the laws of the bootstrap variables conditioned on the weights, from which unconditional convergence will follow, and then we will show that unconditional convergence implies convergence of these laws conditioned on the sample. We may assume, without loss of generality that our random variables are defined on a product probability space, that the X's depend on ω and the weights on ω', so that the conditional laws given the weights or given the sample have a very specific meaning. Since $BL(\mathbb{R})$ is separable, the d_{BL} distance between one such conditional law and a fixed probability measure on $\mathbb{R}$ is measurable. The theorem will follow from a series of simple lemmas. The assumptions in the lemmas are the same as those in the theorem (although some of the lemmas require less).

2.2. LEMMA.

$$\mathcal{L}\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big|w_n\Big) \to_w N(0, c^2\sigma^2) \text{ in probability}$$

in the sense that

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big|w_n\Big), N(0, c^2\sigma^2)\Big) \to 0 \text{ in probability.}$$

PROOF. By conditions E3 and E4, for every subsequence there is a further subsequence, call it n', such that

$$\Big(\max_{1\le j\le n'} \sqrt{n'}|w_{n'}(j) - 1/n'|, n'\sum_{j=1}^{n'}\big(w_{n'}(j) - 1/n'\big)^2\Big) \to (0, c^2) \text{ a.s.}$$

Let ω' be a sample point for which this convergence takes place. Then the random variables

$$Y_{n',i} = \frac{(w_{n'}(j,\omega') - 1/n')X_i}{\sigma\sqrt{\sum_{j=1}^{n'}\big(w_{n'}(j,\omega') - 1/n'\big)^2}}, \quad i = 1, \dots, n', \quad n' \in \{n'\},$$

form a triangular array of random variables which are i.i.d. by rows, and satisfy

$$\sum_{i=1}^{n} \mathrm{Var}_X\, Y_{n,i} = 1 \quad\text{and}\quad \lim_{n'\to\infty}\sum_{i=1}^{n'} \mathbb{E}_X Y_{n',i}^2 I_{|Y_{n',i}|>\varepsilon} = 0 \text{ for all } \varepsilon > 0.$$

Then, by Lindeberg's clt, $\sum_{i=1}^{n'} Y_{n',i}$ converges in law to $N(0,1)$. Hence,

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{n'}(\bar{X}_{n'}^* - \bar{X}_{n'})\Big|w_{n'}\Big), N(0, c^2\sigma^2)\Big) \to 0 \text{ a.s.}$$

and the lemma follows.

□

2.3. LEMMA.

$$\mathcal{L}(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)) \to_w N(0, c^2\sigma^2).$$

PROOF. By Lemma 2.2,

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big|w_n\Big), N(0, c^2\sigma^2)\Big) \to 0 \text{ in probability.}$$

Since these random variables are bounded by 2, convergence takes place in L_1 as well, which gives

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big), N(0, c^2\sigma^2)\Big) \le \mathbb{E}\Big[d_{BL}\Big(\mathcal{L}\Big(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big|w_n\Big), N(0, c^2\sigma^2)\Big)\Big] \to 0.$$

□

Some preparation is necessary in order to condition with respect to the X's in the above lemma.

2.4. LEMMA. *The sequence of conditional laws*

$$\left\{\mathcal{L}\left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big|\mathbf{X}\right)\right\}_{n=1}^{\infty}$$

is tight with probability one.

PROOF. The set Ω_0 where this sequence is tight is

$$\Omega_0 = \bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=1}^{\infty}\left\{\omega : \Pr_w\left(\sqrt{n}|\bar{X}_n^*(\omega) - \bar{X}_n(\omega)| \geq N\right| < \frac{1}{k}\right\}.$$

Now, taking into account that, by exchangeability and E1, $\mathrm{Cov}(w_n(i), w_n(j)) = -(\mathrm{Var}\ w_n(1))/(n-1)$, and using E2, we have

$$\begin{aligned}
&\Pr_w\left(\sqrt{n}|\bar{X}_n^*(\omega) - \bar{X}_n(\omega)| \geq C\right) \\
&\quad\leq \frac{n}{C^2}\Big[\sum_{i=1}^{n} X_i^2(\omega)\mathrm{Var}\ w_n(i) + 2\sum_{i<j\leq n} X_i(\omega)X_j(\omega)\mathrm{Cov}\big(w_n(i), w_n(j)\big)\Big] \\
&\quad\leq \frac{K}{C^2}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i^2(\omega) + \Big|\frac{2}{n(n-1)}\sum_{i<j\leq n} X_i(\omega)X_j(\omega)\Big|\Big] \\
&\quad\leq \frac{3K}{C^2}\frac{1}{n}\sum_{i=1}^{n} X_i^2(\omega) \to \frac{3K\mathbb{E}X^2}{C^2}
\end{aligned}$$

for all ω on a set of probability 1 independent of C by the law of large numbers, and for some $K < \infty$ (and independent of ω). But the ω's for which there is convergence belong to Ω_0 as can be seen by enlarging C if necessary.

□

Let us say that two sequences of probability laws are *weakly equivalent* if they have the same subsequential limit laws, along the same subsequences.

2.5. LEMMA. *There is a set Ω_1 of probability one such that, for all $\omega \in \Omega_1$, the sequences*

$$\left\{\mathcal{L}\left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big|\mathbf{X}\right)(\omega)\right\}_{n=1}^{\infty}$$

are all weakly equivalent.

PROOF. Let $F_n(t,\omega)$ be the empirical distribution function corresponding to $\{X_i\}_{i=1}^n$, and let F be the distribution function of the law of X. Let

$$\Omega_1 = \left\{\omega : \frac{1}{n}\sum_{i=1}^{n} X_i^2(\omega) \to \mathbb{E}X^2,\ \|F_n(\omega, t) - F(t)\|_\infty \to 0\right\}.$$

Then $\Pr \Omega_1 = 1$ and therefore it suffices to prove that

$$n\mathbb{E}_w\Big(\bar{X}_n^*(\omega_2) - \bar{X}_n(\omega_2) - \bar{X}_n^*(\omega_1) + \bar{X}_n(\omega_1)\Big)^2 \to 0$$

for all $\omega_1, \omega_2 \in \Omega_1$. By exchangeability, the random vectors

$$\big(\bar{X}_n^*(\omega_1) - \bar{X}_n(\omega_1), \bar{X}_n^*(\omega_2) - \bar{X}_n(\omega_2)\big) = \sum_{i=1}^n \big(w_n(i) - \frac{1}{n}\big)\big(X_i(\omega_1), X_i(\omega_2)\big)$$

have the same conditional laws, given ω_1 and ω_2, as

$$\sum_{i=1}^n \big(w_n(i) - \frac{1}{n}\big)\big(X_{(n,i)}(\omega_1), X_{(n,i)}(\omega_2)\big),$$

where $X_{(n,1)}, \ldots, X_{(n,n)}$ are the order statistics of $X_1, \ldots, X_n$. Hence, proceeding as in the proof of Lemma 2.4,

$$\begin{aligned}
&n\mathbb{E}_w\Big(\bar{X}_n^*(\omega_2) - \bar{X}_n(\omega_2) - \bar{X}_n^*(\omega_1) + \bar{X}_n(\omega_1)\Big)^2 \\
&\quad\le n\mathbb{E}_w\Big(\sum_{i=1}^n \big(w_n(i) - \frac{1}{n}\big)\big(X_{(n,i)}(\omega_2) - X_{(n,i)}(\omega_1)\big)\Big)^2 \\
&\quad\le \frac{3K}{n}\sum_{i=1}^n \big(X_{(n,i)}(\omega_2) - X_{(n,i)}(\omega_1)\big)^2 \\
&\quad= 3K\int_0^1 \big(G_n(\omega_2, t) - G_n(\omega_1, t)\big)^2 dt,
\end{aligned}$$

where $G_n(\omega, t)$ is the left–continuous inverse (w.r.t. the t variable) of $F_n(\omega, t)$ (i.e., the quantile empirical process). Now, it is classical that, for $\omega_i \in \Omega_1$, $G_n(\omega_i, t) \to G(t)$ for almost every $t \in (0,1)$, where G is the quantile function for the law of X. Also, for $\omega_i \in \Omega_1$,

$$\int_0^1 G_n^2(\omega_i, t)dt = \frac{1}{n}\sum_{j=1}^n X_j^2(\omega_i) \to \mathbb{E}X^2 = \int_0^1 G^2(t)dt$$

(recall that $G(t)$, as a function on $([0,1], \mathcal{B}, \lambda)$, has the law of X). Therefore, by (generalized) dominated convergence,

$$\int_0^1 \big(G_n(\omega_2, t) - G_n(\omega_1, t)\big)^2 dt \to \int_0^1 \big(G(t) - G(t)\big)^2 dt = 0,$$

and the lemma follows.

□

The next lemma shows that bootstrap limit distributions are not random.

2.6. LEMMA. *Suppose that for some subsequence n' and almost every ω,*

$$\mathcal{L}\Big(\sqrt{n'}(\bar{X}^*_{n'} - \bar{X}_{n'})\Big|\mathbf{X}\Big)\Big)(\omega) \to_w \mu(\omega).$$

Then, there exists a probability measure μ such that $\mu(\omega) = \mu$ a.e.

PROOF. Let $\mathcal{F}$ be a countable measure–determining set of bounded Lipschitz functions on $\mathbb{R}$ (e.g., $\mathcal{F} = \{\cos tx, \sin tx : t \in \mathbb{Q}\}$). Then, since by E3 (E2 suffices), $\sqrt{n'}\big|\sum_{j=1}^k (w_{n'}(j) - 1/n')X_j(\omega)\big| \to 0$ in conditional probability given the X's for each $k < \infty$, we have that, for each f in $\mathcal{F}$,

$$\mathbb{E}_w f(\sqrt{n'} \sum_{j=k+1}^{n} (w_{n'}(j) - 1/n')X_j(\omega)) \to \int f(x)\mu(\omega, dx).$$

Hence, this last integral, which is measurable, is a tail random variable for every f. Therefore, there are constants $\mu(f)$ such that $\mu(f) = \int f(x)\mu(\omega, dx)$ for all $f \in \mathcal{F}$ and ω in a set of measure 1, thus showing that the measure $\mu := \mu(\omega)$ for a (any) fixed ω in this set satifies the conclusion of the lemma.

□

PROOF OF THEOREM 2.1. By Lemmas 2.4, 2.5 and 2.6, for every subsequence of the natural numbers there is a further subsequence $\{n'\}$ such that, for almost every ω, the conditional laws $\mathcal{L}\Big(\sqrt{n'}(\bar{X}^*_{n'} - \bar{X}_{n'})\Big|\mathbf{X}\Big)(\omega)$ converge weakly to a non–random probability measure μ that may depend on the subsequence $\{n'\}$. Hence, for each f bounded and continuous,

$$\mathbb{E}_w f\Big(\sqrt{n'}(\bar{X}^*_{n'}(\omega) - \bar{X}_{n'}(\omega))\Big) \to \int f d\mu \quad \text{a.s.}$$

and, by bounded convergence,

$$\mathbb{E} f\Big(\sqrt{n'}(\bar{X}^*_{n'} - \bar{X}_{n'})\Big) \to \int f d\mu.$$

But by Lemma 2.3 (unconditional convergence),

$$\mathbb{E} f\Big(\sqrt{n'}(\bar{X}^*_{n'} - \bar{X}_{n'})\Big) \to \int f dN(0, c^2\sigma^2).$$

Hence, $\mu = N(0, c^2\sigma^2)$. This shows

$$\mathcal{L}\Big(\sqrt{n}(\bar{X}^*_n - \bar{X}_n)\Big|\mathbf{X}\Big) \to N(0, c^2\sigma^2) \quad \text{a.s.}$$

□

This generalized bootstrap contains, by appropriate choice of weights, the regular Efron's bootstrap, the under– or over– sampled bootstrap, the bootstrap without replacement, the Bayesian bootstrap, etc. Checking E3–E4 sometimes requires ingenuity; for example, for the regular bootstrap one takes $w_n(j) = \sum_{i=1}^n \delta_{U_i}(A_j)/n$, where U_i are i.i.d. uniform on $(0,1)$ and A_j, $j = 1, \ldots, n$ is a partition of the interval into sets of mass $1/n$ each; then, one uses e.g. Bernstein's inequality to prove E3, and Poissonization to prove E4.

It should be clear that this theorem extends, by Cramér–Wold, to i.i.d. random vectors of $\mathbb{R}^d$. Praestgaard and Wellner (1993) extended it to empirical processes (more on this below).

1.3. The bootstrap of the mean for stationary sequences. The strong or α-mixing coefficient between two σ-fields $\mathcal{A}$ and $\mathcal{B}$ in $(\Omega, \Sigma, \Pr)$ is defined as

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{(A,B)\in\mathcal{A}\times\mathcal{B}} \left|\Pr(A\cap B) - \Pr(A)\Pr(B)\right|,$$

and constitutes one of several standard measures of dependence. If $\{X_i\}_{i=-\infty}^{\infty}$, is a strictly stationary sequence of random variables and $\mathcal{F}_m = \sigma(X_i : i \le m)$ and $\mathcal{F}^n = \sigma(X_i : i \ge n)$, then the *mixing coefficients* α_n of the sequence $\{X_i\}$ are defined as

$$\alpha_n = \alpha(\mathcal{F}_0, \mathcal{F}^n),$$

and the sequence is said to be *strongly mixing* or *α-mixing* if $\lim_{n\to\infty} \alpha_n = 0$. The bootstrap has no reason to work if it does not mimic, in some essential way, the random mechanism that produces the sample (in Efron's bootstrap, independent sampling from $\mathcal{L}(X)$ is mimicked by independent sampling from P_n). In the α–mixing case sample blocks carry dependence information, more of it the larger their size b is, and, sampling from the set of these blocks (instead of from the sample, as in the i.i.d. case), if there are many, produces a distribution close to that of $(X_1, \ldots, X_b)$. So, here is the stationary or Kunsch bootstrap procedure: Given the sample $X_1, \ldots, X_n$, and the block size $b := b(n)$, we let $B_{i,b} = \{X_i, \ldots, X_{i+b-1}\}$ be the b–size block of observations starting at X_i, $i = 1, \ldots, n-b+1$. We sample with replacement $k := k(n) = [n/b]$ of these blocks, say $B_{i_1}, \ldots, B_{i_k}$, and then the bootstrap sample is constructed from the samples in the blocks, that is

$$X_{n,1}^* = X_{i_1}, \ldots, X_{n,b}^* = X_{i_1+b-1}, X_{n,b+1}^* = X_{i_2}, \ldots, X_{n,kb}^* = X_{i_k+b-1}.$$

Formally, $i_1, \ldots, i_k$ are i.i.d. random varibles, independent from the sequence $\{X_i\}$ and uniformly distributed over the set of integers $\{1, \ldots, n-b+1\}$. (Another basically equivalent procedure takes the indices uniformly distributed on $\{1, \ldots, n\}$, and defines the last b blocks formally in the same way but taking X_{n+r} to be X_r –i.e., as if the data were in a circle). We will call the variables $X_{n,i}^*$, $i = 1, \ldots, kb$, the MBB sample with block size b (at stage n).

The theorem that follows, due to D. Radulović (1996), constitutes a considerable strengthening of the original results of Kunsch (1989) and Liu and Singh (1992). We let $S_n = \sum_{i=1}^n X_i$ and $\sigma_n^2 = \mathrm{Var} S_n$ (so, the sums start at X_1 even though the sequence X_i runs over $\mathbb{Z}$). We resume the notation $\mathcal{L}^*$, $\Pr^*$, $\mathbb{E}^*$, etc. to denote respectively conditional law, probability, expectation, etc. given the sample.

3.1. THEOREM. *Let $\{X_i\}_{i=-\infty}^{\infty}$, be a strictly stationary strong mixing sequence of square integrable real valued r.v.'s such that*

$$\mathcal{L}\Big(\frac{S_n - n\mathbb{E}X_1}{\sigma_n}\Big) \to_w N(0,1) \tag{3.1}$$

Let $X^*_{n,i}$, $i = 1, \ldots, kb$, $n \in \mathbb{N}$, be the MBB samples based on $\{X_i\}$, with block size $b := b(n)$, such that

$$b(n) \to \infty \text{ and } \frac{b(n)}{n} \to 0. \tag{3.2}$$

Then, setting $\sigma^*_n := \sqrt{Var^*(\sum_{i=1}^{k(n)b(n)} X^*_{n,i})}$, we have

$$\mathcal{L}^*(H^*_n) := \mathcal{L}^*\Big(\frac{1}{\sigma^*_n}\sum_{i=1}^{k(n)b(n)} (X^*_{n,i} - \mathbb{E}^* X^*_{n,i})\Big) \to_w N(0,1) \text{ in pr.} \tag{3.3}$$

and

$$\mathcal{L}^*(\tilde{H}^*_n) := \mathcal{L}^*\Big(\frac{1}{\sqrt{k(n)}\sigma_{b(n)}}\sum_{i=1}^{k(n)b(n)} (X^*_{n,i} - \mathbb{E}^* X^*_{n,i})\Big) \to_w N(0,1) \text{ in pr.} \tag{3.4}$$

Moreover, if condition (3.1) is replaced by

$$\mathcal{L}\Big(\frac{S_n - n\mathbb{E}X_1}{\sqrt{n}}\Big) \to N(0,\sigma^2) \text{ and } \Big\{\frac{(S_n - n\mathbb{E}X_1)^2}{n}\Big\}_{n=1}^{\infty} \text{ is uniformly integrable,} \tag{3.1'}$$

(and (3.2) is kept unchanged) then

$$\mathcal{L}^*(\hat{H}^*_n) := \mathcal{L}\Big(\frac{\sum_{i=1}^{k(n)b(n)} (X^*_{n,i} - \mathbb{E}^* X^*_{n,i})}{\sqrt{n}}\Big) \to_w N(0,\sigma^2) \text{ in pr.} \tag{3.5}$$

In case

$$\frac{\sigma_n^2}{n} \to \sigma^2 \text{ and } \sigma^2 > 0,$$

it is clear that the regular clt,

$$\mathcal{L}\Big(\frac{S_n - n\mathbb{E}X_1}{\sqrt{n}}\Big) \to N(0,\sigma^2), \tag{3.6}$$

is equivalent to condition (3.1) and also to condition (3.1').

The best result on the central limit theorem for stationary sequences to present belongs to Doukhan, Massart and Rio (1994): letting $\alpha(t) = \alpha_{[t]}$ and letting Q be the right continuous quantile function of X_1, the condition

$$\int_0^1 \alpha^{-1}(t) Q^2(t) dt < \infty \tag{3.7}$$

implies

$$\frac{\sigma_n^2}{n} \to \sigma^2 \text{ and } \mathcal{L}\Big(\frac{S_n - n\mathbb{E}X_1}{\sqrt{n}}\Big) \to_w N(0,\sigma^2). \tag{3.8}$$

The limits (3.8) are also implied by the Ibragimov and Linnik (1971) sufficient condition for the clt, namely that there exists some $\delta > 0$ such that

$$\mathbb{E}|X_1|^{2+\delta} < \infty \text{ and } \sum_{k=1}^{\infty} \alpha_k^{\frac{\delta}{2+\delta}} < \infty. \tag{3.9}$$

Condition (3.9) is stronger than (3.7).

The main step in the proof of Theorem 3.1 consists in deriving a sort of Raikov Theorem (i.e. lln for squares) associated to the clt (3.1): with it we can control the relevant truncated bootstrap moments and thus derive the bootstrap clt from general principles (the criterion for convergence of row sums of infinitesimal arrays to a normal law already used above). The main tool is the basic covariance inequality of Davidov (1968).

3.1. LEMMA. *Let $\{\xi_n\}_{n=1}^{\infty}$ be a uniformly integrable sequence of real random variables. For each $n \in \mathbb{N}$, let $\xi_{n,i}$, $i = 1, \ldots, n$, be a strictly stationary set of random variables individually distributed as ξ_n. Suppose there exist constants $a_{n,i}$ satisfying*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_{n,i} = 0 \tag{3.10}$$

and such that

$$\left|\mathrm{Cov}\left(Y_1 I_{|Y_1|\le M}, Y_i I_{|Y_i|\le M}\right)\right| \le M^2 a_{n,i} \tag{3.11}$$

for all $M < \infty$ and for all $\sigma(\xi_{n,i})$ measurable random variables Y_i. Then,

$$\frac{1}{n}\sum_{i=1}^{n} (\xi_{n,i} - \mathbb{E}\xi_{n,i}) \to 0 \ \text{ in pr.} \tag{3.12}$$

PROOF. Let $Y_{n,i} = \xi_{n,i} - \mathbb{E}\xi_{n,i}$. Since the sequence $\{Y_{n,1}\}_{n=1}^{\infty}$ is uniformly integrable, it follows that

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Y_{n,i}I_{|Y_{n,i}|>M_n} - \mathbb{E}Y_{n,i}I_{|Y_{n,i}|>M_n}\right)\right| \le 2\mathbb{E}\left|Y_{n,1}\right|I_{|Y_{n,1}|>M_n} \to 0$$

whenever $M_n \to \infty$. For $M_n \to \infty$ to be chosen below, set $\tilde{Y}_{n,i} = Y_{n,i} I_{|Y_{n,i}|\le M_n}$. It then suffices to prove that

$$\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_{n,i} - \mathbb{E}\tilde{Y}_{n,i}) \to 0 \ \text{ in probability.}$$

Stationarity of the set $\tilde{Y}_{n,1} \ldots, \tilde{Y}_{n,n}$ for each n, together with (3.11) (note $\tilde{Y}_{n,i}$ is $\sigma(\xi_{n,i})$ measurable), gives

$$\begin{aligned}
\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_{n,i} - \mathbb{E}\tilde{Y}_{n,i})\right| > \varepsilon\right\} &\le \frac{1}{n^2\varepsilon^2} \sum_{1\le i,j\le n} \mathrm{Cov}(\tilde{Y}_{n,i}, \tilde{Y}_{n,j}) \\
&\le \frac{2}{n\varepsilon^2} \sum_{j\le n} \left|\mathrm{Cov}(\tilde{Y}_{n,1}, \tilde{Y}_{n,j})\right| \\
&\le \frac{2M_n^2}{n\varepsilon^2} \sum_{j=1}^{n} a_{n,j}
\end{aligned}$$

for all $\varepsilon > 0$. Now, choosing $M_n = \left(\sum_{j=1}^n a_{n,j}/n\right)^{-1/4}$ makes this probability tend to zero by (3.10).

□

In Theorem 3.1 we can assume, without loss of generality, that $\mathbb{E}X_i = 0$, and we do so in what follows. For each $n \in \mathbb{N}$ and $i = 1, \ldots, N(n) := n - b(n) + 1$, we let $Z_{n,i}$ be the sum of the data in block B_i, that is,

$$Z_{n,i} = X_i + \ldots + X_{i+b(n)-1}, \quad n \in \mathbb{N}, \quad i = 1, \ldots, N(n). \tag{3.13}$$

The previous lemma gives the following corollary, a kind of Raikov's theorem associated to the clt in (3.1):

3.2. COROLLARY. *Under the hypotheses of Theorem 3.1 and assuming* $\mathbb{E}X = 0$, *we have:*

i)

$$\frac{1}{N(n)} \sum_{i=1}^{N(n)} \left(\frac{Z_{n,i}}{\sigma_{b(n)}}\right)^2 \to 1 \text{ in pr.,} \tag{3.14}$$

ii) for every $\delta > 0$ *and* $0 \le p \le 2$,

$$\frac{k(n)}{N(n)} \sum_{i=1}^{N(n)} \left|\frac{Z_{n,i}}{\sqrt{k(n)}\sigma_{b(n)}}\right|^p I_{|Z_{n,i}| > \delta\sigma_{b(n)}\sqrt{k(n)}} \to 0 \text{ in pr.,} \tag{3.15}$$

iii)

$$\frac{1}{N(n)} \sum_{i=1}^{N(n)} \frac{Z_{n,i}}{\sigma_{b(n)}} \to 0 \text{ in pr.,} \tag{3.16}$$

and, for all $\delta > 0$,

$$\frac{1}{N(n)} \sum_{i=1}^{N(n)} \frac{Z_{n,i}}{\sigma_{b(n)}} I_{|Z_{n,i}| \le \delta\sigma_{b(n)}\sqrt{k(n)}} \to 0 \text{ in pr.,} \tag{3.17}$$

PROOF. To prove that the limit (3.14) holds, we apply Lemma 3.1 with $\xi_{N(n),i} = Z_{n,i}^2/\sigma_{b(n)}^2$. First we note that, since $b(n) \to \infty$, the clt (3.1) implies that the variables $Z_{n,1}/\sigma_{b(n)}$ converge in law to a standard normal variable. Then, the second moments of these variables being all equal to 1, this gives that the sequence $\{Z_{n,1}^2/\sigma_{b(n)}^2\}_{n=1}^\infty$ is uniformly integrable. Also, if X is $\sigma(Z_{n,1}^2)$ measurable and Y is $\sigma(Z_{n,i}^2)$ then, by the definition of α_n, and since $\sigma(Z_{n,1}^2) \subset \mathcal{F}_{b(n)}$ and $\sigma(Z_{n,i}^2) \subset \mathcal{F}^i$, we have that $\alpha(X,Y) \le \alpha_{(i-b(n))\vee 0}$ and that

$$\frac{1}{N(n)} \sum_{i=1}^{N(n)} \alpha_{(i-b(n))\vee 0} \le \frac{b(n)}{n - b(n) + 1} + \frac{1}{n - b(n) + 1} \sum_{i=b(n)+1}^{n-b(n)+1} \alpha_i \to 0 \tag{3.18}$$

(since $\alpha_n \to 0$ and $b(n)/n \to 0$). Davidov's (1968) inequality (actually a particular case, Theorem 17.2.1 in Ibragimov and Linnik, 1971), to the effect that

$$\mathrm{Cov}(X,Y) \le 4\alpha(X,Y)\|X\|_\infty\|Y\|_\infty, \tag{3.19}$$

shows that the sequence $\{Z_{n,1}^2/\sigma_{b(n)}^2\}_{n=1}^\infty$ satisfies condition (3.11) with $a_{N,i} = 4\alpha_{(i-b(n))\vee 0}$. (3.18) gives condition (3.10) for these constants, and therefore, our sequence satisfies Lemma 3.1. Its conclusion, the limit (3.12), translates exactly into the limit (3.14). (Note that Lemma 3.1 also holds, with only the obvious changes, if the ξ variables are indexed by a sequence $N(n) \to \infty$ of integers, instead of by $\mathbb{N}$.)

For every $\delta > 0$ and $0 \le p \le 2$, the array

$$\zeta_{N(n),i} = k(n)\Big|\frac{Z_{n,i}}{\sqrt{k(n)}\sigma_{b(n)}}\Big|^p I_{|Z_{n,i}|>\delta\sigma_{b(n)}\sqrt{k(n)}}, \quad i = 1, \dots, N(n), \quad n \in \mathbb{N},$$

also satisfies the hyptheses of Lemma 3.1 with respect to the same array of coefficients $a_{N,i} = \alpha_{(i-b(n))\vee 0}$ as above: (3.11) certainly checks, and $\{\zeta_{N(n),1}\}_{n=1}^\infty$ is uniformly integrable because $\zeta_{N(n),1} \le \delta^{p-2}\xi_{N(n),1}$ for all n. Now, the limit in (3.15) follows from Lemma 3.1 because the uniform integrability of the variables $Z_{n,1}^2/\sigma_{b(n)}^2$ implies

$$\lim_{n\to\infty} k(n)\mathbb{E}\Big|\frac{Z_{n,1}}{\sqrt{k(n)}\sigma_{b(n)}}\Big|^p I_{|Z_{n,1}|>\delta\sigma_{b(n)}\sqrt{k(n)}}$$
$$\le \delta^{p-2}\lim_{n\to\infty}\mathbb{E}\Big|\frac{Z_{n,1}}{\sqrt{k(n)}\sigma_{b(n)}}\Big|^2 I_{|Z_{n,1}|>\delta\sigma_{b(n)}\sqrt{k(n)}} = 0.$$

The rest of the statements follow in the same way, once we observe that the sequence $\{Z_{n,1}/\sigma_{b(n)}\}_{n=1}^\infty$ is uniformly integrable (since the sequence of its squares is) and $\mathbb{E}Z_{n,1} = 0$.

□

Now the proof of Theorem 3.1 becomes a routine check of the classical conditions for normal convergence:

PROOF OF THEOREM 3.1. For each n, let $Z_{n,i}^*$, $i = 1, \dots, k(n)$, be an array of random variables which, conditionally on the sample $\{X_i\}$, are i.i.d. with (conditional) law

$$\Pr{}^*\{Z_{n,i}^* = Z_{n,j}\} = \frac{1}{N(n)}, \quad j = 1, \dots, N(n).$$

With this definition, we have

$$\mathcal{L}^*(\tilde{H}_n^*) := \mathcal{L}^*\Big(\frac{1}{\sqrt{k(n)}\sigma_{b(n)}}\sum_{i=1}^{k(n)}(Z_{n,i}^* - \mathbb{E}^*Z_{n,i}^*)\Big).$$

Hence, by previous arguments and the general criterion on convergence to the normal law of infinitesimal arrays, the proof of (3.4) reduces to checking that the following three limits hold for every $\delta > 0$:

$$\sum_{i=1}^{k(n)}\Pr{}^*\Big\{\Big|\frac{Z_{n,i}^*}{\sqrt{k(n)}\sigma_{b(n)}}\Big| > \delta\Big\} \to 0 \text{ in pr.},$$

$$\sum_{i=1}^{k(n)} \mathrm{Var}^*\Big(\frac{Z_{n,i}^*}{\sqrt{k(n)}\sigma_{b(n)}} I_{|Z_{n,i}^*|\le \delta\sigma_{b(n)}\sqrt{k}}\Big) \to 1 \ \text{ in pr.}$$

and

$$\sum_{i=1}^{k(n)} \mathbb{E}^*\Big(\frac{Z_{n,i}^*}{\sqrt{k(n)}\sigma_{b(n)}} I_{|Z_{n,i}^*|> \delta\sigma_{b(n)}\sqrt{k}}\Big) \to 0 \ \text{ in pr..}$$

By the definition of $Z_{n,i}^*$, these three conditions can be written as:

$$\frac{k(n)}{N(n)} \sum_{i=1}^{N(n)} I_{|Z_{n,i}|>\delta\sigma_{b(n)}\sqrt{k}} \to 0 \ \text{ in pr.,} \tag{3.20}$$

$$\frac{1}{N(n)} \sum_{i=1}^{N(n)} \Big(\frac{Z_{n,i}}{\sigma_{b(n)}}\Big)^2 I_{|Z_{n,i}|\le\delta\sigma_{b(n)}\sqrt{k}} - \Big(\frac{1}{N(n)} \sum_{i=1}^{N(n)} \frac{Z_{n,i}}{\sigma_{b(n)}} I_{|Z_{n,i}|\le\delta\sigma_{b(n)}\sqrt{k}}\Big)^2 \to 1 \ \text{ in pr.} \tag{3.21}$$

and

$$\frac{k(n)}{N(n)} \sum_{i=1}^{N(n)} \frac{Z_{n,i}}{\sqrt{k(n)}\sigma_{b(n)}} I_{|Z_{n,i}|>\delta\sigma_{b(n)}\sqrt{k}} \to 1 \ \text{ in pr.} \tag{3.22}$$

Now, (3.20) and (3.22) are just (3.15) in Corollary 3.2 respectively for $p = 0$ and $p = 1$. The first and second terms at the left of (3.21) converge in probability respectively to 1 and to 0 by (3.14) and (3.17) in Corollary 3.2. Thus, we have proved that the limit (3.4) holds.

Given (3.4), proving (3.3) reduces to showing that

$$\frac{(\sigma_n^*)^2}{k(n)\sigma_{b(n)}^2} \to 1 \ \text{ in probability.}$$

Since $(\sigma_n^*)^2 = k(n)\mathrm{Var}^*(Z_{n,1}^*)$ by conditional independence of the $Z_{n,i}^*$ variables, we have

$$\frac{(\sigma_n^*)^2}{k(n)\sigma_{b(n)}^2} = \frac{1}{N(n)} \sum_{i=1}^{N(n)} \Big(\frac{Z_{n,i}}{\sigma_{b(n)}}\Big)^2 - \Big(\frac{1}{N(n)} \sum_{i=1}^{N(n)} \frac{Z_{n,i}}{\sigma_{b(n)}}\Big)^2,$$

which tends to 1 in probability by (3.14) and (3.16) in Corollary 3.2.

□

Since the MBB procedure produces a triangular array of conditionally row–wise independent random variables, somehow with weaker dependence than original sample, it is conceivable that the MBB works in cases when the original clt does not. In fact this is the case, as shown by an example in Peligrad (1996, Remark 2.1).

1.4. The bootstrap of U–statistics. Degenerate U–statistics, together with the maximum of i.i.d. uniform variables, were among the first examples for which the regular Efron's bootsrap (that is, sampling n times from the empirical measure P_n) was seen not to work. Bretagnolle (1983) discovered that reduction of bootstrap sample size makes the bootstrap consistent. These statistics constitute also an early example of the fact that one can often modify the bootstrap procedure so that it better simulates the original random mechanism. This may require, however, some information about the main features of the problem at hand. In the case of U–stastistics, a basic feature is the degree of degeneracy as it determines the O_P size of the statistic. Arcones and Giné (1992) proposed to empirically degenerate the U–statistic to the same order as the original before bootstrapping: in this case no reduction of bootstrap sample size is necessary. We illustrate both ideas in a simple example. Let X_i be i.i.d., $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$. Then,

$$\frac{1}{n}\sum_{(i,j)\in I_n^2} X_iX_j \to_d Z^2 - 1,$$

where Z is $N(0,1)$ and $I_n^2 = \{(i,j) : 1 \le i,j \le n, i \ne j\}$. This statistic is degenerate of order 1. Let us write the statistic in the form

$$\frac{1}{n}\sum_{(i,j)\in I_n^2} X_iX_j = \left[\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right]^2 - \frac{1}{n}\sum_{i=1}^n X_i^2.$$

A 'naive' application of the bootstrap gives

$$\frac{1}{n}\sum_{(i,j)\in I_n^2} X_{n,i}^*X_{n,j}^* = \left[\frac{\sum_{i=1}^n X_{n,i}^*}{\sqrt{n}}\right]^2 - \frac{1}{n}\sum_{i=1}^n (X_{n,i}^*)^2. \tag{4.1}$$

Now, the law of large numbers for X^2 bootstraps with no problem (the lln bootstraps):

$$\begin{aligned}\mathbb{E}^*\left[\frac{\sum_{i=1}^n (X_{n,i}^*)^2}{n} - 1\right]^2 &\le 2\mathbb{E}^*\left[\frac{\sum_{i=1}^n (X_{n,i}^*)^2}{n} - \mathbb{E}^*(X_{n,1}^*)^2\right]^2 + 2\left[\frac{\sum_{j=i}^n X_j^2}{n} - 1\right]^2 \\ &\le \frac{4}{n}\mathbb{E}^*(X_{n,1}^*)^4 + 2\left[\frac{\sum_{j=i}^n X_j^2}{n} - 1\right]^2 \\ &= \frac{4}{n^2}\sum_{j=1}^n X_j^4 + 2\left[\frac{\sum_{j=i}^n X_j^2}{n} - 1\right]^2 \to 0 \text{ a.s.}\end{aligned}$$

by the Marcinkiewicz and Kolmogorov laws of large numbers. But the clt part at the right of identity (4.1) does not converge to a normal law because *the centering* $\sqrt{n}\bar{X}_n$ *is missing.* Bretagnolle's solution was: reduce the sample size to make the missing centering go to zero (a.s. or in pr.) Obviously, taking the bootstrap sample size in (4.1) to be m_n (instead of n) turns the centering of the clt part into $\sqrt{m_n}\bar{X}_n$, which tends to zero in pr. if $m_n/n \to 0$ (by the clt), and tends to zero a.s. if $(m_n \log\log n)/n \to 0$ (by the lil). And this is what happens in general for the bootstrap of (the clt for) degenerate U–statistics: it works in pr if $m_n/n \to 0$ and it works a.s. if $(m_n \log\log n)/n \to 0$ (the rate $m_n(\log n)^{1+\delta}/n \to 0$ was first used for the a.s. bootstrap but, as Arcones and Giné (1989) observed, $(m_n \log\log n)/n \to 0$ is the appropriate rate for the a.s. bootstrap in many situations –basically, those in which one can invoke some kind of lil). There is another logical solution to the above problem since, after all, one cannot ignore the centering in the bootstrap of the mean even if $\mathbb{E}X = 0$: just add the centering or, what is the same, reason this way: what makes the norming constants to be n instead of $\sqrt{n}$ in the clt for X_iX_j is that the X's are centered, so we are in fact dealing with the statistic $\sum_{I_n^2}(X_i - \mathbb{E}X)(X_j - \mathbb{E}X)/n$, which naturally bootstraps as $\sum_{I_n^2}(X_{n,i}^* - \bar{X}_n)(X_{n,j}^* - \bar{X})/n$. And this works since

$$\begin{aligned}\frac{1}{n}\sum_{(i,j)\in I_n^2}&(X_{n,i}^* - \bar{X}_n)(X_{n,j}^* - \bar{X}_n)\\ &= \left[\frac{\sum_{i=1}^n (X_{n,i}^* - \bar{X}_n)}{\sqrt{n}}\right]^2 - \frac{1}{n}\sum_{i=1}^n (X_{n,i}^* - \bar{X}_n)^2 \to_d Z^2 - 1 \text{ a.s.} \quad (4.2)\end{aligned}$$

by the bootstrap clt and lln. This bootstrap (the 'degenerate bootstrap') has the disadvantage of requiring knowledge that we may not have (we may not know $\mathbb{E}X = 0$), but it is useful in testing (we will elaborate on this below).

We let $(S, \mathcal{S})$ be a measurable space and P a probability measure on it, and let X, X_i be i.i.d. S-valued random variables with law P (i.e., the X's do not have to be real).

4.1. Definition. *A* P^m*-integrable function of* m *variables,* $f : S^m \to \mathbb{R}$*, symmetric in its entries, is* P*-degenerate of order* $r-1$*,* $1 < r \le m$*, if*

$$\int f(x_1, \ldots, x_m)d\mathrm{P}^{m-r+1}(x_r, \ldots, x_m) = \int f d\mathrm{P}^m \quad \text{for all } x_1, \ldots, x_{r-1} \in S$$

whereas

$$\int f(x_1, \ldots, x_m)d\mathrm{P}^{m-r}(x_{r+1}, \ldots, x_m)$$

is not a constant function. If f *is* P^m*-centered and is* P*-degenerate of order* $m-1$*, that is, if*

$$\int f(x_1, \ldots, x_m)d\mathrm{P}(x_1) = 0 \quad \text{for all } x_2, \ldots, x_m \in S,$$

then f *is said to be canonical or completely degenerate with respect to* P*. If* f *is not degenerate of any positive order we say it is non–degenerate or degenerate of order zero.*

In this definition the identities are taken in the almost everywhere sense.

With the notation $\mathrm{P}_1 \times \cdots \times \mathrm{P}_m f = \int f d(\mathrm{P}_1 \times \cdots \times \mathrm{P}_m)$, the *Hoeffding projections* of $f : S^m \to \mathbb{R}$ symmetric are defined as

$$\pi_k^{\mathrm{P}} f(x_1, \ldots, x_k) := \pi_{k,m}^{\mathrm{P}} f(x_1, \ldots, x_k) := (\delta_{x_1} - \mathrm{P}) \times \cdots \times (\delta_{x_k} - \mathrm{P}) \times \mathrm{P}^{m-k} f$$

for $x_i \in S$ and $0 \le k \le m$. Note that $\pi_0^{\mathrm{P}} f = \mathrm{P}^m f$ and that, for $k > 0$, $\pi_k^{\mathrm{P}} f$ is a completely degenerate function of k variables. For f integrable these projections induce a decomposition of the U-statistic

$$U_n(f) := U_n^{(m)}(f) := U_n^{(m)}(f, \mathrm{P}) := \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} f(X_{i_1}, \ldots, X_{i_m})$$

into a sum of U-statistics of orders $k \le m$ which are orthogonal if $\mathrm{P}^m f^2 < \infty$ and whose kernels are completely degenerate, namely, the *Hoeffding decomposition*:

$$U_n^{(m)}(f) = \sum_{k=0}^{m} \binom{m}{k} U_n^{(k)}(\pi_k^{\mathrm{P}} f) \tag{4.3}$$

(the subindex m of $\pi_{k,m}^{\mathrm{P}}$ is not displayed; it will be dropped whenever no confusion is possible). This decomposition follows easily by expanding

$$f(x_1, \dots, x_m) = \delta_{x_1} \times \cdots \times \delta_{x_m} f = \big((\delta_{x_1} - \mathrm{P}) + \mathrm{P}\big) \times \cdots \times \big((\delta_{x_m} - \mathrm{P}) + \mathrm{P}\big) f$$

into terms of the form $(\delta_{x_{i_1}} - \mathrm{P}) \times \cdots \times (\delta_{x_{i_k}} - \mathrm{P}) \times \mathrm{P}^{m-k} f$. It is very simple to check that f symmetric is P–degenerate of order $r-1$ iff $r = \min\{k > 0 : \pi_{k,m}^{\mathrm{P}} f \not\equiv 0\}$. Therefore, f is degenerate of order $r - 1 \ge 0$ iff its Hoeffding expansion, except for the constant term, starts at term r, that is,

$$U_n(f) - \mathrm{P}^m f = \sum_{k=r}^{m} \binom{m}{k} U_n^{(k)}(\pi_k^{\mathrm{P}} f). \tag{4.4}$$

Hoeffding's decomposition is a basic tool in the analysis of U-statistics and in particular it will be put to use for the bootstrap. We recall that P_n refers to the empirical measure constructed from the sample $X_1, \dots, X_n$, and that, conditionally on $\mathbf{X}$, the variables $X_{n,i}^*$ are i.i.d. with law P_n. According to the definition, we will write

$$U_n^*(f) := U_n^{(m)*}(f) := U_n^{(m)}(f, \mathrm{P}_n) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} f(X_{n,i_1}^*, \dots, X_{n,i_m}^*).$$

The meaning of $\pi_k^{\mathrm{P}_n} f$ is equally clear:

$$\pi_k^{\mathrm{P}_n} f(x_1, \dots, x_k) := \pi_{k,m}^{\mathrm{P}_n} f(x_1, \dots, x_k) := (\delta_{x_1} - \mathrm{P}_n) \times \cdots \times (\delta_{x_k} - \mathrm{P}_n) \times \mathrm{P}_n^{m-k} f$$

for $x_i \in S$ and $0 \le k \le m$. From the clt for U-statistics, we know that the k-th term in the Hoeffding decomposition (4.4) is asymptotically $O_P(n^{-k/2})$ so that the whole statistic is exactly of the order of the first term. Hence, at least up to first order approximation, only the first term in the Hoeffding expansion needs bootsrapping and we can ignore the rest. Since $\pi_k^{\mathrm{P}} f$ is completely P-degenerate we must replace it by $\pi_k^{\mathrm{P}_n} f$ before bootstrapping, and this is the content of one of the next two theorems. The first is from Arcones and Giné, loc. cit.

4.2. THEOREM. *Let $f(x_1, \dots, x_m)$, $x_i \in S$, be a P–square integrable symmetric kernel, P–degenerate of order $r-1$, so that, in particular,*

$$\binom{n}{r}^{1/2} \Big(U_n(f, \mathrm{P}) - \mathrm{P}^m f\Big) \to_d K_{\mathrm{P},f,r} \tag{4.5}$$

where $K_{\mathrm{P},f,r}$ is a Gaussian chaos variable of order r. Assume also that, for $d := \#\{i_1, \dots, i_m\}$,

$$\mathbb{E}\big|f(X_{i_1}, \dots, X_{i_m})\big|^{2d/m} < \infty. \tag{4.6}$$

Then,

$$\mathcal{L}^*\left[\binom{n}{r}^{1/2}\binom{m}{r}U_n^{(r)}(\pi_r^{\mathrm{P}_n}f,\mathrm{P}_n)\right] \to_w \mathcal{L}(K_{\mathrm{P},f,r}) \quad \text{a.s.} \tag{4.7}$$

We note that $\binom{m}{r}\pi_r^{\mathrm{P}_n}f$ can be replaced in (4.6) by the P_n–orthogonal projection of f onto the space of P_n–degenerate kernels of order $r-1$, and $U_n^{(r)}$ by $U_n^{(m)}$, but this new kernel is more complicated and may not lead to a better approximation.

We need a simple but useful proposition of Sen (1974) on the Marzinkiewicz law of large numbers for U–statistics (the proof here is from Giné and Zinn, 1992, which also has more information on the subject).

4.3. PROPOSTION. *If* $\mathbb{E}|f|^p < \infty$ *with* $0 < p < 1$, *then*

$$\frac{1}{n^{m/p}} \sum_{1 \le i_1 < \ldots < i_m \le n} \left|f(X_{i_1}, \ldots, X_{i_m})\right| \to 0 \quad \text{a.s.}$$

PROOF. We assume $f \ge 0$ and set $s := m/p > m$. By Kronecker's lemma, it suffices to show that

$$\sum_{j=m}^{\infty} \frac{1}{j^s} \sum_{1 \le i_1 < \ldots < i_{m-1} < j} f(X_{i_1}, \ldots, X_{i_{m-1}}, X_j) < \infty \quad \text{a.s.}$$

Since

$$\sum_{j=m}^{\infty} (j-1)^{m-1} \Pr\{f > j^s\} \le \int_{m-1}^{\infty} x^{m-1} \Pr\{f > x^s\} dx \le m^{-1} \mathbb{E} f^{m/s} < \infty,$$

we can truncate at the level j^s and therefore it suffices to show

$$\sum_{j=m}^{\infty} \frac{1}{j^s} \sum_{1 \le i_1 < \ldots < i_{m-1} < j} (f I_{f \le j^s})(X_{i_1}, \ldots, X_{i_{m-1}}, X_j) < \infty \quad \text{a.s.}$$

But this follows by B. Levi's lemma and the following estimate of the series of expected values:

$$\begin{aligned}
\sum_{j=m}^{\infty} \frac{1}{(j-1)^{s-m+1}} \mathbb{E} f I_{f \le j^s} &\le \int_{m-1}^{\infty} x^{m-s-1} \mathbb{E} f I_{f \le x^s} dx \\
&\le \int_{m-1}^{\infty} \int_0^{x^s} x^{m-s-1} \Pr\{f > t\} dt dx \\
&= \frac{(m-1)^{s-m}}{s-m} \int_0^{(m-1)^s} \Pr\{f > t\} dt \\
&\quad + \int_{(m-1)^s}^{\infty} \Big(\int_{t^{1/s}}^{\infty} x^{m-s-1} dx\Big) \Pr\{f > t\} dt \\
&\le \frac{(m-1)^{2s-m}}{s-m} + \frac{1}{s-m} \int_0^{\infty} t^{\frac{m}{s}-1} \Pr\{f > t\} dt < \infty.
\end{aligned}$$

□

The following corollary follows immediately upon decomposing the V–statistic into a sum of U–statistics.

4.4. COROLLARY. *If f satisfies the integrability condition (4.6) then*

$$\frac{1}{n^m} \sum_{1 \le i_1, \dots, i_m \le n} f(X_{i_1} \dots, X_{i_m}) \to \mathrm{P}^m f \quad \text{a.s.} \tag{4.8}$$

This in turn has the following consequence, that will allow us to restrict our attention to very simple functions in the proof of Theorem 4.2.

4.5. COROLLARY. *Let $f_\ell \to 0$ in $L_2(\mathrm{P}^m)$ and suppose the functions f_ℓ are symmetric and satisfy the integrability condition (4.6). Then,*

$$\lim_{\ell\to\infty} \lim_{n\to\infty} \mathbb{E}^* \left[\binom{n}{r}^{1/2} U_n^{(r)*}\left(\pi_r^{\mathrm{P}_n} f_\ell\right) \right]^2 = 0. \tag{4.9}$$

PROOF. Note that the operators π_k^Q are centering operators and therefore they are contractions in $L_2(Q)$, for any probability measures Q, in particular for the random mesures P_n, and recall that $\mathbb{E}^*$ is nothing but integration with respect to the measure P_n. Then, observing that the summands in a U–statistic whose kernel is Q–canonical are Q–orthogonal (assuming they are square integrable), we obtain

$$\begin{aligned}
\mathbb{E}^* \left[\binom{n}{r}^{1/2} U_n^{(r)*}\left(\pi_r^{\mathrm{P}_n} f_\ell\right) \right]^2 &= \mathbb{E}^* \left[\pi_r^{\mathrm{P}_n} f_\ell\right]^2 \\
&\le \mathbb{E}^* f_\ell^2 (X_{n,1}^*, \dots, X_{n,n}^*) \\
&= \frac{1}{n^m} \sum_{1 \le i_1, \dots, i_m \le n} f_\ell^2 (X_{i_1}, \dots, X_{i_m}) \\
&\to \mathrm{P}^m f_\ell^2 \quad \text{a.s. as } n \to \infty \\
&\to 0 \quad \text{as } \ell \to \infty.
\end{aligned}$$

□

In order to implement the reduction to simple kernels it is useful to describe two simple identities, namely, the polarization identities and Newton's identities. The *polarization identities* are as follows: If $\varepsilon_1, \dots, \varepsilon_k$ are i.i.d. Rademacher variables (that is, $\Pr\{\varepsilon_i = 1\} = 1 - \Pr\{\varepsilon_i = -1\} = 1/2$), then, for any k real functions $\phi_1, \dots, \phi_k$ of one variable, not necessarily different, and for any $x_1, \dots, x_k$ in S, we have

$$\sum_\sigma \phi_{\sigma(1)}(x_1) \cdots \phi_{\sigma(k)}(x_k) = \mathbb{E}\Big[\varepsilon_1 \cdots \varepsilon_k \big(\sum_{i=1}^k \varepsilon_i \phi_i(x_1)\big) \cdots \big(\sum_{i=1}^k \varepsilon_i \phi_i(x_k)\big)\Big], \tag{2.4}$$

where σ runs over all the permutations of $\{1, \ldots, k\}$. By developing the expected value in the second term of this identity we obtain a linear combination of at most 2^k functions of the form

$$h_k^{\psi}(x_1, \ldots, x_k) = \psi(x_1) \cdot \ldots \cdot \psi(x_k)$$

with $\psi(x) = \pm\phi_{i_1}(x) \pm \cdots \pm \phi_{i_k}(x)$. Newton's identities, which are useful to handle U–statistics with kernels of the form h_k^{ψ}, are as follows: given $t_1, \ldots, t_n$ in $\mathbb{R}$, if, for $1 \leq r \leq n$, we let $p_r = \sum_{1 \leq i_1 < \ldots < i_r \leq n} t_{i_1} \cdots t_{i_r}$ and $s_r = \sum_{i=1}^{n} t_i^r$, then the identity

$$s_k - p_1 s_{k-1} + p_2 s_{k-2} - \ldots + (-1)^{k-1} s_1 p_{k-1} + (-1)^k k p_k = 0$$

holds for all $k < n$. These identities, which can be checked by induction, give, also by induction, that for every $k \in \mathbb{N}$ there is a real polynomial R_k of degreee k, in k variables, such that for all $n > k$,

$$\sum_{1 \leq i_1 < \ldots < i_k \leq n} t_{i_1} \cdots t_{i_k} = R_k\Big(\sum_{i=1}^{n} t_i, \sum_{i=1}^{n} t_i^2, \ldots, \sum_{i=1}^{n} t_i^k\Big).$$

(Moreover, $R_k(u_1, \ldots, u_k)$ is a sum of monomials of the form $c\prod_{i=1}^{k} u_i^{k_i}$ with $\sum_{i=1}^{k} ik_i = k$ and the coefficient of u_1^k is $1/k!$.)

4.6. Proof of Theorem 4.2. Let $\{\phi_i\}_{i \in I}$ be a complete orthonormal system of $L_2(\mathrm{P})$ consisting of *bounded* functions. Since $f \in L_2(\mathrm{P}^m)$, we have

$$f(x_1, \ldots, x_m) = \sum_{i_1, \ldots, i_m \in I} c_{i_1, \ldots, i_m} \phi_{i_1}(x_1) \ldots \phi_{i_m}(x_m)$$

in the sense of $L_2(\mathrm{P}^m)$ and, f being symmetric in its entries, the coefficients $c_{i_1, \ldots, i_m}$ are invariant under permutations of the indices $i_1, \ldots, i_m$. Then, by polarization,

$$f = \lim \sum_{finite} t_i h^{\psi_i} := \lim_{\ell \to \infty} h_\ell \tag{4.10}$$

also in $L_2(\mathrm{P}^m)$, where the functions ψ_i are bounded and $h^{\psi}(x_1, \ldots, x_m)$ is as defined above, so that h_ℓ is a finite linear combination of h^{ψ_i}'s.

We will not prove here the central limit theorem for U–statistics, but it is required in this proof. See, for example, Arcones and Giné, loc. cit., or the forthcoming book of de la Peña and Giné (1997) –preliminary versions of it were distributed when these lectures were delivered. So, we recall that there is a chaos process K_{P} indexed by the canonical functions of all orders on S^m such that the random variable $K_{\mathrm{P},f,r}$ in the limit (4.5) is precisely $K_{\mathrm{P},f,r} = \binom{m}{r} K_{\mathrm{P}}\big(\pi_r^{\mathrm{P}} f\big)$ for all $f \in L^2(\mathrm{P}^m)$, that $\mathbb{E}\big(K_{\mathrm{P}}(g_1) - K_{\mathrm{P}}(g_2)\big)^2 = \int (g_1 - g_2)^2 d\mathrm{P}^r$ for all square integrable P–canonical functions of r variables, and that the limit (4.5) holds jointly for any finite number of kernels, with limit the corresponding join distributions of the variables $\binom{m}{r} K_{\mathrm{P}}\big(\pi_r^{\mathrm{P}} f\big)$. Let d be any distance metrizing weak convergence in $\mathbb{R}$.

Then, using the just mentioned isometry, (4.10) and Corollary 4.5 on the functions $f_\ell = f - h_\ell$, we have

$$\begin{aligned}
&\lim_{n\to\infty} d\Big[\mathcal{L}^*\Big(\binom{n}{r}^{1/2} U_n^*(\pi_r^{P_n} f)\Big), \mathcal{L}(K_{P,f,r})\Big] \\
&\quad\le \lim_{\ell\to\infty}\lim_{n\to\infty} d\Big[\mathcal{L}^*\Big(\binom{n}{r}^{1/2} U_n^*(\pi_r^{P_n} f)\Big), \mathcal{L}^*\Big(\binom{n}{r}^{1/2} U_n^*(\pi_r^{P_n} h_\ell)\Big)\Big] \\
&\qquad + \lim_{\ell\to\infty}\lim_{n\to\infty} d\Big[\mathcal{L}^*\Big(\binom{n}{r}^{1/2} U_n^*(\pi_r^{P_n} h_\ell)\Big), \mathcal{L}(K_{P,h_\ell,r})\Big] \\
&\qquad\quad + \lim_{\ell\to\infty} d\Big[\mathcal{L}(K_{P,h_\ell,r}), \mathcal{L}(K_{P,f,r})\Big] \\
&\quad= \lim_{\ell\to\infty}\lim_{n\to\infty} d\Big[\mathcal{L}^*\Big(\binom{n}{r}^{1/2} U_n^*(\pi_r^{P_n} h_\ell)\Big), \mathcal{L}(K_{P,h_\ell,r})\Big] \quad \text{a.s.}
\end{aligned}$$

Hence, we have reduced proving Theorem 4.2 to showing that

$$w - \lim_{n\to\infty} \mathcal{L}^*\Big[\binom{n}{r}^{1/2} U_n^{(r)}(\pi_r^{P_n} h^\psi, P_n)\Big] = w - \lim_{n\to\infty} \mathcal{L}\Big[\binom{n}{r}^{1/2} U_n^{(r)}(\pi_r^{P} h^\psi, P)\Big] \quad \text{a.s.} \tag{4.11}$$

for ψ bounded, *jointly* in any finite number of ψ's.

Note that

$$\pi_r^P h^\psi(x_1,\ldots,x_r) = (P\psi)^{m-r}(\psi(x_1) - P\psi)\cdots(\psi(x_n) - P\psi)$$

for all $x_i \in S$, and likewise for P_n. Therefore, if R_r is the polynomial of degree r prescribed by Newton's identities, we have

$$\begin{aligned}
&\frac{1}{n^{r/2}} \sum_{1\le i_1<\ldots<i_r\le n} \pi_r^P h^\psi(X_{i_1},\ldots,X_{i_r}) \qquad\qquad (4.12)\\
&= (P\psi)^{m-r} R_r\Big(\sum_{i=1}^n(\psi(X_i)-P\psi)/n^{1/2}, \sum_{i=1}^n(\psi(X_i)-P\psi)^2/n, \ldots, \sum_{i=1}^n(\psi(X_i)-P\psi)^r/n^{r/2}\Big)
\end{aligned}$$

and, likewise,

$$\begin{aligned}
&\frac{1}{n^{r/2}} \sum_{1\le i_1<\ldots<i_r\le n} \pi_r^{P_n} h^\psi(X_{n,i_1}^*,\ldots,X_{n,i_r}^*) \\
&\quad= (P_n\psi)^{m-r} R_r\Big(\sum_{i=1}^n(\psi(X_{n,i}^*)-P_n\psi)/n^{1/2}, \sum_{i=1}^n(\psi(X_{n,i}^*)-P_n\psi)^2/n, \\
&\qquad\qquad \ldots, \sum_{i=1}^n(\psi(X_{n,i}^*)-P_n\psi)^r/n^{r/2}\Big) \qquad (4.13)
\end{aligned}$$

Now, by the law of large numbers,

$$P_n\psi \to P\psi \quad \text{a.s.};$$

by the bootstrap clt,

$$w-\lim_{n\to\infty}\mathcal{L}^*\left[\frac{1}{n^{1/2}}\sum_{i=1}^n\big(\psi(X^*_{n,i})-\mathrm{P}_n\psi\big)\right]=w-\lim_{n\to\infty}\mathcal{L}\left[\frac{1}{n^{1/2}}\sum_{i=1}^n\big(\psi(X_i)-\mathrm{P}\psi\big)\right]\ \text{a.s.};$$

by the bootstrap lln (see the beginning of this section),

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\big(\psi(X^*_{n,i})-\mathrm{P}_n\psi\big)^2=\mathrm{Var}_{\mathrm{P}}(\psi)$$

in conditional probability, a.s.; and, by Marcinkiewicz's law of large numbers,

$$\lim_{n\to\infty}\frac{1}{n^{k/2}}\sum_{i=1}^n\mathbb{E}^*\big(\psi(X^*_{n,i})-\mathrm{P}_n\psi\big)^k=\lim_{n\to\infty}\frac{1}{n^{k/2}}\sum_{i=1}^n\big(\psi(X_i)-\mathrm{P}\psi\big)^k=0\ \text{ a.s.}$$

for all $k>2$. Note that the last four limits hold *jointly* in any finite number of ψ's (Theorem 1.1 holds in $\mathbb{R}^d$). So, (4.11) follows from (4.12) and (4.13) and these last four limits because polynomials commute with weak limits.

□

4.6. Remark. The previous proof shows, in fact, that if f satisfies the integrability hypothesis (4.6) then, without any degeneracy hypotheses, we have

$$w-\lim_{n\to\infty}\mathcal{L}^*\left[\binom{n}{r}^{1/2}U_n^{(r)}(\pi_r^{\mathrm{P}_n}f,\mathrm{P}_n)\right]=w-\lim_{n\to\infty}\mathcal{L}\left[\binom{n}{r}^{1/2}U_n^{(r)}(\pi_r^{\mathrm{P}}f,\mathrm{P})\right]\ \text{a.s.}\tag{4.11'}$$

(and this last limit exists by the clt for canonical U–statistics).

A similar proof that requires, however, a little extra work for the analogue of Corollary 4.5 (and which we omit), gives the following refinement of Bretagnolle's bootstrap limit theorem for U–statistics.

4.7. Theorem. *Let $f(x_1,\ldots,x_m)$, $x_i\in S$, be a P–square integrable symmetric kernel, P–degenerate of order $r-1$, so that, in particular,*

$$\binom{n}{r}^{1/2}\Big(U_n(f,\mathrm{P})-\mathrm{P}^m f\Big)\to_d K_{\mathrm{P},f,r}\tag{4.5}$$

where $K_{\mathrm{P},f,r}$ is a Gaussian chaos variable of order r. Assume also that, for $d:=\#\{i_1,\ldots,i_m\}$,

$$\mathbb{E}\big|f(X_{i_1},\ldots,X_{i_m})\big|^{2d/m}<\infty.\tag{4.6}$$

Let m_n be a sequence of natural numbers tending to infinity. Then,

$$\mathcal{L}^*\left[\binom{m_n}{r}^{1/2}\Big(U^*_{n_n}(f,\mathrm{P}_n)-\mathrm{P}_n^m f\Big)\to_d K_{\mathrm{P},f,r}\right]\to_w\mathcal{L}(K_{\mathrm{P},f,r})\tag{4.14}$$

in probability if $m_n/n\to 0$ and a.s. if $(m_n\log\log n)/n\to 0$.

It seems clear that the extensions of the bootstrap considered in the previous two sections can also be carried out for U–statistics. For instance, for $m=2$ and for exchangeable weights such as those in Section 2, Hušková and Janssen (1993) modify the bootstrap of Arcones and Giné as follows: for h degenerate, the bootstrap statistic they consider is

$$\sum_{I_n^2} w_{ni}w_{nj}h(X_i,X_j) - \sum_{i=1}^{n} w_{ni}\frac{1}{n}\sum_{j\le n, j\ne i} h(X_i,X_j) \\ - \sum_{j=1}^{n} w_{nj}\frac{1}{n}\sum_{i\le n, i\ne j} h(X_i,X_j) + \frac{1}{n^2}\sum_{I_n^2} h(X_i,X_j),$$

which, for multinomial weights, coincides with $\binom{n}{2}^{1/2}U_n^{(2)*}\big(\pi_2^{\mathrm{P}_n}h\big)$ except for the $h(X_i,X_i)$ terms. We will not pursue this matter.

We will now indicate how to apply the bootstrap theory just developped to a test of independence proposed by Hoeffding (1948). This test follows a general pattern on how to use model based bootstraps to test hypotheses, formalized by Romano (1989) (concrete examples of application of the same scheme already existed –see e.g. Arcones and Giné, 1991).

4.8. EXAMPLE. Let $F(x,y)$ be a bivariate distribution function and let

$$\Delta(F) = \int\int \big[F(x,y) - F(x,\infty)F(\infty,y)\big]^2 dF(x,y).$$

Let Ω'' be the set of bivariate continuous distrubution functions whose marginals are also continuous. If the joint density F of (X,Y) is in Ω'', then X and Y are independent iff $\Delta(F)=0$. It is easy to check that $\Delta(F)$ can be written as

$$\Delta(F) = \frac{1}{4}\int\cdots\int \psi(x_1,x_2,x_3)\psi(x_1,x_4,x_5)\psi(y_1,y_2,y_3)\psi(y_1,y_4,y_5) \\ \times dF(x_1,y_1)\cdots dF(x_5,y_5),$$

where $\psi(t_1,t_2,t_3) = I_{t_1\ge t_2} - I_{t_1\ge t_3}$. A good unbiased estimator of $\Delta(F)$ is then

$$D_n = \frac{1}{\binom{n}{5}}\sum_{1\le i_1,\ldots,i_5\le n} \phi\big((X_{i_1},Y_{i_1}),\ldots,(X_{i_5},Y_{i_5})\big),$$

where we write $\phi\big((x_1,y_1),\ldots,(x_5,y_5)\big)$ for the symmetrization of the function

$$\frac{1}{4}\psi(x_1,x_2,x_3)\psi(x_1,x_4,x_5)\psi(y_1,y_2,y_3)\psi(y_1,y_4,y_5).$$

Now, $\mathbb{E}D_n = \Delta$ and therefore, by the law of large numbers for U–statistics, $D_n = O_{\mathrm{P}}(1)$ if X and Y are not independent. Hoeffding also shows that D_n is degenerate of order 1 if and only if X and Y are independent. In particular, under the independence hypothesis, since moreover $\Delta = 0$, the clt for U–statistics gives that $D_n = O_{\mathrm{P}}\big(n^{-1}\big)$. In this last case, in fact, Hoeffding shows that the limit of $\mathcal{L}\big(nD_n\big)$ has a continuous distribution which is independent of the joint distribution of X and Y as long as it is in Ω''. So, one can always, in principle, tabulate the limiting distribution and construct a test. An alternative is to find, by simulation, $c_n^*(\alpha)$ such that

$$\Pr{}^*\left\{\left|n\binom{5}{2}U_n^{(2)*}(\pi_2^{F_n}\phi)\right| > c_n^*(\alpha)\right\} \simeq \alpha$$

and reject independence if $|nD_n| > c_n^*(\alpha)$. Note that: 1) Under the independence null hypothesis, $\Pr\{|nD_n| > c_n^*(\alpha)\} \to \alpha$, that is, the test has asymptotically type I error α (by Theorem 4.2, since the limiting distribution is continuous). 2) For any cdf F in $\mathbb{R}^2$, whether it is a product cdf or not, $\pi_2^F\phi$ is canonical and therefore the statistics $n\binom{5}{2}U_n^{(2)}(\pi_2^{\mathrm{P}}\phi)$ converge in law (Remark 4.6) and the bootstrap statistics $n\binom{5}{2}U_n^{(2)*}(\pi_2^{\mathrm{P}_n}\phi)$ converge in conditional law to the same limit; this implies that the numbers $c_n^*(\alpha)$ stabilize (a.s.) at some finite number as $n \to \infty$ no matter what the distribution F of the data is (although their limit may depend on F). 3) As mentioned above, $|nD_n| = O_{\mathrm{P}}(n)$ under any alternatives. Conclusion: If we compute the level $c_n^*(\alpha)$ as indicated, and reject independence if $|nD_n| > c_n^*(\alpha)$, then the test has asymptotically level α and the probability of rejecting the null hypothesis under any alternative tends to 1 as $n \to \infty$. Note that there is no need to resort to a sample satisfying the null hypothesis in order to compute the levels $c_n^*(\alpha)$.

We now briefly turn to V–statistics. With $S = \mathbb{R}$, letting F be the cdf of X and F^{-1} its right continuous inverse, and letting β denote the Brownian bridge, Filippova (1961) proved that if f satisfies the integrability conditions (4.6), then

$$n^{m/2}\int_{\mathbb{R}^m} fd(\mathrm{P}_n - \mathrm{P})^m \to_d \int_{\mathbb{R}} \overset{m)}{\cdots} \int_{\mathbb{R}} f(F^{-1}(u_1)),\dots,F^{-1}(u_m))d\beta(u_1)\cdots d\beta(u_m). \tag{4.15}$$

Recall that P_n is the empirical measure corresponding to the sample X_i, $\mathrm{P}_n = \sum_{i=1}^n \delta_{X_i}/n$. Filippova's theorem is important because these statistics arise as the second degree components in the von Mises development of smooth statistics, and therefore, it has some interest to have its bootstrap version.

The left side of (4.15) is a normalized canonical V–statistic, and S needs not be $\mathbb{R}$ for these statistics to have a limit (but then it cannot be expressed as a multiple integral of the classical Brownian bridge, although it is an element of a Gaussian chaos of order m). In general, V statistics of order m and (symmetrtic) kernel $f(x_1,\dots,x_m)$, based on X_i, i.i.d.(P), are defined as

$$V_n^{(m)}(f,\mathrm{P}) := \frac{1}{n^m}\sum_{1\le i_1,\dots,i_m\le n} f(X_{i_1},\dots,X_{i_m}) = \int_{S^m} fd\mathrm{P}_n^m.$$

It follows from this definition that

$$V_n^{(m)}(\pi_m f,\mathrm{P}) := \frac{1}{n^m}\sum_{1\le i_1,\dots,i_m\le n} \pi_m^{\mathrm{P}} f(X_{i_1},\dots,X_{i_m}) = \int_{S^m} fd(\mathrm{P}_n-\mathrm{P})^m,$$

that is, projecting f onto the space of canonical kernels has the effect of centering the empirical measures. V–statistics decompose into the sum of U–statistics, and their theory then reduces to that of U–statistics. But for the bootstrap it seems more appropriate to work by analogy with U–statistics, rather than by reduction. We will only state the bootstrap clt for canonical V-statistics, and then, the proof will only be sketched. (See Arcones and Giné, 1992, for more details.) In the following theorem, P_n^* denotes the *bootstrap empirical measure*, that is,

$$\mathrm{P}_n^* := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_{n,i}^*}.$$

4.9. THEOREM. *Let f be a kernel on S^m such that $f(X_{i_1},\dots,X_{i_m})$ is square integrable for all choices $1 \le i_1,\dots,i_m \le m$. Then,*

$$w-\lim_{n\to\infty}\mathcal{L}^*\Big[n^{m/2}\int_{S^m} fd(\mathrm{P}_n^*-\mathrm{P}_n)^m\Big] = w-\lim_{n\to\infty}\mathcal{L}\Big[n^{m/2}\int_{S^m} fd(\mathrm{P}_n-\mathrm{P})^m\Big] \quad \text{a.s.} \tag{4.16}$$

PROOF. (Sketch) For a partition $-\infty = t_0 < t_1 < \dots < t_{m-1} < t_k = \infty$ of the completed real line, and for numbers $g_{i_1,\dots,i_m}$, let

$$g(x_1,\dots,x_m) = \sum_{1\le j_1,\dots,j_m\le k} g_{i_1,\dots,i_m} I_{A_{j_1}}(x_1)\cdots I_{A_{j_m}}(x_m), \tag{4.17}$$

where $A_j = (t_{j-1},t_j]$, $j<k$, and $A_k = (t_{k-1},\infty)$. Since

$$n^{m/2}(\mathrm{P}_n-\mathrm{P})^m g = \sum_{1\le j_1,\dots,j_m\le k} g_{i_1,\dots,i_m}\Big[n^{1/2}(\mathrm{P}_n-P)(A_{j_1})\Big]\cdots\Big[n^{1/2}(\mathrm{P}_n-P)(A_{j_k})\Big]$$

the bootstrap central limit theorem and the continuous mapping theorem show that the bootstrap clt (4.16) holds true for g. Let us denote this class of functions by $\mathcal{G}$. Next, we define $\tilde{L}_2(\mathrm{P},m)$ as the set of functions f that satisfy the integrability hypotheses of the theorem, with the norm $\|f\|_{\tilde{L}_2(\mathrm{P},m)}$ that we now describe. $\|f\|_{\tilde{L}_2(\mathrm{P},m)}$ is the maximum over all partitions Q of $\{1,\dots,m\}$ of the $L_2(\mathrm{P}^{\#Q})$ norms of the functions f_Q defined as follows: if $Q=\{A_1,\dots,A_r\}$ then $f_Q(x_1,\dots,x_r) = f(x_{Q(1)},\dots,x_{Q(m)})$, where $Q(j) = x_k$ if $j\in A_k$. $\mathcal{G}$ is dense in $\tilde{L}_2(\mathrm{P},m)$. It is easy to show that

$$\big\|n^{m/2}(\mathrm{P}_n-\mathrm{P})^m f\big\|_{L_2(\mathrm{P}^m)} \le c_m\|f\|_{\tilde{L}_2(\mathrm{P},m)},$$

where $c_m<\infty$ is a constant that depends only on m. These comments also apply with P replaced by P_n. Then, if $h\in\tilde{L}_2(\mathrm{P},m)$ we have

$$\mathbb{E}^*\big|n^{m/2}(\mathrm{P}_n^*-\mathrm{P}_n)^m h\big| \le \|h\|_{\tilde{L}_2(\mathrm{P}_n,m)} = \max_Q\|h_Q\|_{L_2(\mathrm{P}_n^{\#Q})} \to \max_Q\|h_Q\|_{L_2(\mathrm{P}^{\#Q})} \quad \text{a.s.} \tag{4.18}$$

by the law of large numbers for U-statistics (as the square of $\|h_Q\|_{L_2(\mathrm{P}_n^{\#Q})}$ is a U-statistic with kernel h_Q^2.) Now, the density of $\mathcal{G}$ in $\tilde{L}_2(\mathrm{P},m)$, the fact that Theorem 4.9 holds for all kernels in $\mathcal{G}$, and the inequality and limit in (4.18) imply (4.16) for all $f\in\tilde{L}_2(\mathrm{P},m)$ by an easy triangle inequality.

□

Although we have sketched an independent proof of Theorem 4.9, this theorem is in fact a corollary of Theorem 4.2.

1.5. A general m out of n bootstrap. In two situations in the previous sections, undersampling has been seen to work when the regular Efron bootstrap does not. Politis and Romano (1994) (preprint from 1992) and Götze (1993) showed that the bootstrap works in great generality if the bootstrap sample corresponding to a sample of size n is obtained by resampling m_n data from $X_1, \ldots, X_n$ *without replacement*, and with $m_n/n \to \infty$. This scheme can be made to fit the exchangeable bootstrap of Section 2 by taking $(w_n(1), \ldots, w_n(n))$ to be a row of m_n times the number $1/m_n$ and $n - m_n$ times the number zero, at random (that is, the subset of subindices $j_1, \ldots, j_{m_n}$ for which $w_n(j_k) \neq 0$ is uniformly distributed over all the subsets of size m_n of $1, \ldots, n$). However we will not exploit this fact. Rather, we will present the surprisingly simple proof of Politis and Romano based on an old exponential inequality of Hoeffding (1946) for U-statistics.

Here is Hoeffding's inequality, where we replace one of the original computations by a more efficient one taken from Ledoux and Talagrand (1991), Lemma 1.5.

5.1. LEMMA. *Let X, X_i, $i \in \mathbb{N}$, be i.i.d. S-valued random variables, with law P on the measurable space $(S, \mathcal{S})$. Let $f : S^m \to \mathbb{R}$ be a symmetric kernel such that $\int f dP^m = 0$ and $\|f\|_\infty = c < \infty$. Then, for all $t > 0$,*

$$\Pr\Big\{\frac{1}{\binom{n}{m}}\Big|\sum_{1\le i_1<\ldots<i_m\le n} f(X_{i_1}, \ldots, X_{i_m})\Big| > t\Big\} \le 2\exp\Big\{-\frac{[\frac{n}{m}]t^2}{2c^2}\Big\}. \tag{5.1}$$

PROOF. First we see that if ξ is a centered random variable whose absolute value is bounded by 1, then

$$\mathbb{E}e^{\lambda\xi} \le e^{\lambda^2/2} \tag{5.2}$$

for all $\lambda \in \mathbb{R}$. For this, we just observe that, since $\lambda\xi = \frac{1+\xi}{2}\lambda + \frac{1-\xi}{2}(-\lambda)$ and $|\xi| \le 1$, convexity of the exponential function gives $e^{\lambda\xi} \le \frac{1+\xi}{2}e^{\lambda} + \frac{1-\xi}{2}e^{-\lambda}$ so that the expected value satisfies $\mathbb{E}e^{\lambda\xi} \le \cosh\lambda \le e^{\lambda^2/2}$, proving (5.2).

We set $k := [n/m]$ and define

$$W(x_1, \ldots, x_n) = \frac{1}{k}\sum_{j=0}^{k-1} f\big(x_{jm+1}, \ldots, x_{(j+1)m}\big),$$

the average of f over disjoint m-blocks of x's. Then,

$$\frac{1}{\binom{n}{m}}\sum_{1\le i_1<\ldots<i_m\le n} f(X_{i_1}, \ldots, X_{i_m}) = \frac{1}{n!}\sum_{\mathbf{i}\in I_n^n} W\big(X_{i_1}, \ldots, X_{i_n}\big),$$

and therefore, by convexity of $e^{\lambda|x|}$,

$$\exp\Big\{\lambda\frac{1}{\binom{n}{m}}\Big|\sum_{1\le i_1<\ldots<i_m\le n} f(X_{i_1}, \ldots, X_{i_m})\Big|\Big\} \le \frac{1}{n!}\sum_{\mathbf{i}\in I_n^n} \exp\Big\{\lambda\big|W\big(X_{i_1}, \ldots, X_{i_n}\big)\big|\Big\} \tag{5.3}$$

The fact that the summands defining $W(X_{i_1},\ldots,X_{i_n})$ are all independent (as they relate to different m–blocks of X_i's) facilitates application of the bound (5.2) to W, hence to the U–statistic, as follows: By (5.2) and (5.3),

$$\begin{aligned}\mathbb{E}\exp\Big\{\lambda\frac{1}{\binom{n}{m}}\Big|\sum_{1\le i_1<\ldots<i_m\le n} f(X_{i_1},\ldots,X_{i_m})\Big|\Big\} &\\ \le \mathbb{E}\exp\Big\{\lambda\big|W(X_1,\ldots,X_n)\big|\Big\} &\\ \le \mathbb{E}\exp\Big\{\lambda W(X_1,\ldots,X_n)\Big\} + \mathbb{E}\exp\Big\{-\lambda W(X_1,\ldots,X_n)\Big\} &\\ = \Big[\mathbb{E}\exp\Big\{\frac{\lambda c}{k}\frac{f(X_1,\ldots,X_m)}{c}\Big\}\Big]^k + \Big[\mathbb{E}\exp\Big\{-\frac{\lambda c}{k}\frac{f(X_1,\ldots,X_m)}{c}\Big\}\Big]^k &\\ \le 2\exp\Big\{\frac{\lambda^2c^2}{2k}\Big\}. &\end{aligned}$$

Then, by Chebyshev's inequality,

$$\Pr\Big\{\frac{1}{\binom{n}{m}}\Big|\sum_{1\le i_1<\ldots<i_m\le n} f(X_{i_1},\ldots,X_{i_m})\Big| > t\Big\} \le 2\exp\Big\{-\lambda t + \frac{\lambda^2c^2}{2k}\Big\},$$

which gives (5.1) upon taking $\lambda = tk/c^2$.

□

As a transition to the second chapter, we will state and prove the main result in this section for Banach space valued statistics (instead of real or multivariate statistics). For this we need a lemma on weak convergence (van der Vaart and Wellner, 1996, page 72, more general and with a different proof). Let B be a Banach space. We recall that

$$BL_1(B) = \big\{f : B \to \mathbb{R}\ :\ \|f\|_\infty \le 1,\ \|f\|_{\mathrm{Lip}} \le 1\big\},$$

where $\|f\|_{\mathrm{Lip}} := \sup_{x\neq y\in B}|f(x)-f(y)|/\|x-y\|$.

5.2. LEMMA. *Let μ be a tight Borel probability mesure on a Banach space B. Then, there exists a countable subset $\mathcal{D}(\mu)$ of $BL_1(B)$ such that, for any sequence μ_n, $n\in\mathbb{N}$, of tight Borel probability measures on B, the following are equivalent:*

i) $\mu_n \to_w \mu$,

ii) $\sup_{f\in\mathcal{D}(\mu)}\Big|\int f d(\mu_n-\mu)\Big| \to 0$,

iii) $\int f d\mu_n \to \int f d\mu$ *for all* $f\in\mathcal{D}(\mu)$.

PROOF. Let K_r, $r\in\mathbb{N}$, be a sequence of compact sets in B such that $\mu(K_n)\to 1$, and let $f_r(x) := d(x,K_r)\wedge 1$, $x\in B$, where d denotes the usual distance from points to sets. Since $BL_1(K_r)$ is separable for the sup norm (by Arzelà–Ascoli) and every function in $BL_1(K_r)$ extends to a function in $BL_1(B)$ (Kirszbraun–McShane's theorem: see e.g. Dudley, 1989, or Araujo and Giné, 1980), there exists a countable set $\tilde{\mathcal{D}}_r \subset BL_1(B)$ which is dense in $BL_1(B)$ for the pseudonorm

$$\|f\|_{K_r} = \sup_{x\in K_r}|f(x)|.$$

We may and do assume that, for each r, $\tilde{\mathcal{D}}_r$ contains the constant function 1_B. We set $\mathcal{D}_r := \{f(1-f_r)/2 : f \in \tilde{\mathcal{D}}_r\}$ and $\mathcal{D}(\mu) := \cup_{r=1}^{\infty}\mathcal{D}_r$. Then, $\mathcal{D}(\mu)$ is a countable subset of $BL_1(B)$.

It is well known (e.g., Dudley, 1989, page 310, or Araujo and Giné, 1980, page 10) that i) implies ii) and therefore iii), and that in order to prove i) it suffices to show that

$$\int h d\mu_n \to \int h d\mu \tag{5.4}$$

for all $h \in BL_1(B)$ (e.g. Dudley, 1989, page 310, or Araujo and Giné, 1980, pages 10–11).

Let us assume that condition iii) holds for $\mathcal{D}(\mu)$ and let $h \in BL_1(B)$ and $\varepsilon \in (0,1)$. Let r be such that $\mu(K_r) > 1 - \dfrac{\varepsilon^2}{12\cdot 64}$. Since by hypothesis, $\int(1-f_r)d(\mu_n - \mu) \to 0$ as $n \to \infty$, there is $N < \infty$ such that

$$\Big|\int f_r d(\mu_n - \mu)\Big| = \Big|\int (1-f_r)d(\mu_n-\mu)\Big| \leq \frac{\varepsilon^2}{12\cdot 64}$$

for all $n > N$, and therefore, for these values of n,

$$\int f_r d\mu_n \leq \mu(K_r^c) + \frac{\varepsilon^2}{12\cdot 64} < \frac{\varepsilon^2}{6\cdot 64}.$$

By construction, there is $f \in \mathcal{D}_r$ such that

$$\|h - f\|_{K_r} \leq \frac{\varepsilon}{16}.$$

Then,

$$\begin{aligned}\Big|\int hd(\mu_n-\mu)\Big| &\leq \Big|\int (f(1-f_r)-h)d(\mu_n-\mu)\Big| + \Big|\int f(1-f_r)d(\mu_n-\mu)\Big| \\ &\leq \Big|\int_{K_r}(f(1-f_r)-h)d(\mu_n-\mu)\Big| + \Big|\int_{0<f_r(x)\leq\varepsilon/16}(f(1-f_r)-h)d(\mu_n-\mu)\Big| \\ &\qquad + \Big|\int_{f_r(x)>\varepsilon/16}(f(1-f_r)-h)d(\mu_n-\mu)\Big| + \frac{\varepsilon^2}{12\cdot 64}.\end{aligned}$$

Now,

$$\Big|\int_{K_r}(f(1-f_r)-h)d(\mu_n-\mu)\Big| = \Big|\int_{K_r}(f-h)d(\mu_n-\mu)\Big| \leq \int \frac{\varepsilon}{16}d(\mu_n+\mu) = \frac{\varepsilon}{8}.$$

If $f_r(x) = d(x, K_r) \leq \varepsilon/16$ then, since $f, h \in BL_1(B)$ and $\|f-h\|_{K_r} \leq \varepsilon/16$, we have

$$|f(x)(1-f_r(x)) - h(x)| \leq |f(x)(1-f_r(x)) - f(x)| + \frac{3\varepsilon}{16} \leq \frac{\varepsilon}{4},$$

and therefore,

$$\Big|\int_{0<f_r(x)\leq\varepsilon/16}(f(1-f_r)-h)d(\mu_n-\mu)\Big| \leq \frac{\varepsilon}{2}.$$

Finally, if $f_r(x) > \varepsilon/16$ we have

$$|f(x)(1-f_r(x)) - h(x)| \leq 2 \leq \frac{32}{\varepsilon} f_r(x),$$

so that

$$\Big|\int_{f_r(x)>\varepsilon/16} (f(1-f_r)-h)\,d(\mu_n-\mu)\Big| \leq \frac{32}{\varepsilon}\int f_r d(\mu_n+\mu) \leq \frac{32}{\varepsilon}\Big(\frac{\varepsilon^2}{6\cdot 64}+\frac{\varepsilon^2}{12\cdot 64}\Big) = \frac{\varepsilon}{8}.$$

Collectiong all these estimates we obtain that $\left|\int h d(\mu_n - \mu)\right| < \varepsilon$ for all $n > N$, proving that (5.4) holds. So, iii) implies i) and the lemma is proved.

□

As is well known, if $B = \mathbb{R}^d$ then $\mathcal{D}(\mu)$ can be taken independent of μ and such that the sup in *ii*) equals the d_{BL} distance (e.g., by a similar proof, taking $K_r = \{|x| \leq r\}$).

Let X, X_i and P be as above. Let $\theta(\mathrm{P})$ be a B-valued function of P and, for each $n \in \mathbb{N}$, let $T_n(x_1, \ldots, x_n)$ be a B-valued measurable function defined on S^n, symmetric in its entries such that the probability measures $P^n \circ T_n^{-1}$ are tight in B. We assume there is a sequence of constants $\tau_n \to \infty$ and a B-valued random variable Z whose law is tight such that

$$\mathcal{L}\Big\{\tau_n\big[T_n(X_1,\ldots,X_n) - \theta(\mathrm{P})\big]\Big\} \to_w \mathcal{L}(Z). \tag{5.5}$$

We further assume that there exists a sequence m_n of positive integers satisfying

$$m_n \to \infty, \quad \frac{m_n}{n} \to 0 \;\text{ and }\; \frac{\tau_{m_n}}{\tau_n} \to 0 \tag{5.6}$$

as $n \to \infty$. The m out of n bootstrap without replacement is defined as follows. For every n, $Y^*_{n,1}, \ldots, Y^*_{n,m_n}$ are m_n samples drawn without replacement from $(X_1, \ldots, X_n)$, that is, if Pr^* denotes, as usual, conditional probability given the sample, then, for any subset $\{i_1, \ldots, i_{m_n}\}$ of $\{1, \ldots, n\}$,

$$\mathrm{Pr}^*\Big[\big\{Y^*_{n,1}, \ldots, Y^*_{n,m_n}\big\} = \big\{X_{i_1}, \ldots, X_{i_{m_n}}\big\}\Big] = \frac{1}{\binom{n}{m_n}}.$$

(Here, the points X_{i_ℓ}, whether they are different or not, are treated as different elements of the set $\{X_{i_1}, \ldots, X_{i_{m_n}}\}$, and likewise for the points $Y^*_{n,\ell}$.) With these definitions and under these assumptions, we have the following theorem.

5.3. THEOREM. *Suppose the limit (5.5) holds and that the sequences of constants* $\{m_n\}$ *and* $\{\tau_n\}$ *satisfy (5.6). Then, the* m_n *out of* n *bootstrap without replacement of the limit theorem (5.5) works in probability, that is,*

$$\mathcal{L}^*\Big(\tau_{m_n}\Big[T_{m_n}(Y^*_{n,1}, \ldots, Y^*_{n,m_n}) - T_n(X_1, \ldots, X_n)\Big]\Big) \to_w \mathcal{L}(Z) \;\text{ in pr.} \tag{5.7}$$

PROOF. To ease notation, let us set $m := m_n$, $T_m^* := T_{m_n}\big(Y_{n,1}^*, \ldots, Y_{n,m_n}^*\big)$ and $T_n := T_n\big(X_1, \ldots, X_n\big)$. The proof of the theorem basically consists of showing that

$$\mathbb{E}^* f\big(\tau_m(T_m^* - T_n)\big) \to \mathbb{E}f(Z) \ \text{ in pr.} \tag{5.8}$$

uniformly in $f \in \mathcal{D} := \mathcal{D}\big(\mathcal{L}(Z)\big)$, the countable set of bounded Lipschitz functions corresponding to the probaility law of Z by Lemma 5.2. Set

$$d_{\mathcal{D}}(\mu, \nu) := \sup_{f\in\mathcal{D}}\Big|\int f d(\mu - \nu)\Big|$$

and recall that, by Lemma 5.2, this distance metrizes weak convergence to $\mu = \mathcal{L}(Z)$. For $f \in \mathcal{D} \subset BL_1(B)$ we have

$$\begin{aligned}
&\Big|\mathbb{E}^* f\big(\tau_m(T_m^* - T_n)\big) - \mathbb{E}f(Z)\Big| \\
&= \Big|\frac{1}{\binom{n}{m}} \sum_{1\le i_1<\ldots<i_m\le n} f\big(\tau_m\big[T_m(X_{i_1}, \ldots, X_{i_m}) - T_n(X_1, \ldots, X_n)\big]\big) - \mathbb{E}f(Z)\Big| \\
&\le \Big|\frac{1}{\binom{n}{m}} \sum_{1\le i_1<\ldots<i_m\le n} f\big(\tau_m(T_m(X_{i_1}, \ldots, X_{i_m}) - \theta(\mathrm{P}))\big) - \mathbb{E}f(Z)\Big| + \tau_m\big|T_n - \theta(P)\big| \\
&\le \Big|\frac{1}{\binom{n}{m}} \sum_{1\le i_1<\ldots<i_m\le n} f\big(\tau_m(T_m(X_{i_1}, \ldots, X_{i_m}) - \theta(\mathrm{P}))\big) - \mathbb{E}f\big(\tau_m(T_m - \theta(\mathrm{P}))\big)\Big| \\
&\qquad + \Big|\mathbb{E}f\big(\tau_m(T_m - \theta(\mathrm{P}))\big) - \mathbb{E}f(Z)\Big| + \frac{\tau_m}{\tau_n}\tau_n\big|T_n - \theta(\mathrm{P})\big| \\
&:= I_n(f) + II_n(f) + III_n.
\end{aligned} \tag{5.9}$$

Now, by (5.5) and the third limit in (5.6), III_n, which does not depend on f, tends to zero in pr. By (5.5), the fact that $m_n \to \infty$, and Lemma 5.2, $i) \Rightarrow ii)$, we have

$$\sup_{f\in\mathcal{D}} II_n(f) = d_{\mathcal{D}}\Big(\mathcal{L}\big(\tau_m(T_m - \theta(\mathrm{P}))\big), \mathcal{L}(Z)\Big) \to 0.$$

I_n is handled via Hoeffding's inequality (5.1): since, for each n, I_n is a U–statistic whose kernel is centered and is bounded by 2, inequality (5.1) gives

$$\sup_{f\in\mathcal{D}} \Pr\big\{I_n(f) > \varepsilon\big\} \le 2\exp\Big\{-\frac{\big[\frac{n}{m}\big]\varepsilon^2}{8}\Big\} \tag{5.10}$$

for all $\varepsilon > 0$. We have thus proved the limit (5.8) uniformly over $f \in \mathcal{D}$, that is, for all $\varepsilon > 0$,

$$\sup_{f\in\mathcal{D}} \Pr\Big\{\big|\mathbb{E}^* f\big(\tau_m(T_m^* - T_n)\big) - \mathbb{E}f(Z)\big| > \varepsilon\Big\} \to 0.$$

By Borel–Cantelli and the fact that $\mathcal{D}$ is countable, this implies that every subsequence n' has a further subsequence n'' such that for all ω in a set $\Omega_{\{n''\}}$ of probability 1,

$$\mathbb{E}^* f\big(\tau_{m_{n''}}(T_{m_{n''}}^*(\omega) - T_{n''}(\omega))\big) \to \mathbb{E}f(Z) \ \text{ for all } f \in \mathcal{D}.$$

Then, by Lemma 5.2, $iii) \Rightarrow i)$, it follows that

$$\mathcal{L}^*\left[\tau_{m_{n''}}(T^*_{m_{n''}} - T_{n''})\right] \to_w \mathcal{L}(Z) \text{ a.s.,} \tag{5.13}$$

which, by Lemma 5.2, $i) \Rightarrow ii)$, can be rewritten as

$$L_{n''} := d_{\mathcal{D}}\Big(\mathcal{L}^*\left[\tau_{m_{n''}}(T^*_{m_{n''}} - T_{n''})\right], \mathcal{L}(Z)\Big) \to 0 \text{ a.s.} \tag{5.14}$$

Since, $\mathcal{D}$ is countable and $\mathbb{E}^* f\big(\tau_{m_n}(T^*_{m_n}(\omega) - T_n(\omega))\big)$ is a random variable for all f measurable, it follows that L_n is measurable for all n. Then, the usual subsequence argument applies to (5.14) and gives that $L_n \to 0$ in probability, proving the theorem. (Recall that the definition of the bootstrap in probability does not depend on the metric, as long as it metrizes weak convergence to the limit.)

□

If m_n is such that $\sum \exp\{-\varepsilon n/m_n\} < \infty$ and $\tau_m(T_n - \theta(\mathrm{P})) \to 0$ a.s., then the limit (5.6) holds a.s.

5.3. REMARK. (*The bootstrap with replacement.*) If one tries this proof on the undersampled bootstrap without replacement, one finds a problem with the main piece of (5.9), namely, I_n, which is now a V–statistic instead of a U–statistic. This V–statistic consists of the diagonal parts, which are typically biased, and the non-diagonal part, which is the same U–statistic as above, but with norming $1/n^m$. The conclusion is that such a simple proof works for the bootstrap with replacement whenever $\frac{n!/(n-m)!}{n^m} \to 1$, that is, when $m_n^2/n \to 0$. So, results like the undersampled bootstrap for the mean in the stable convergence case, or for degenerate U–statistics, where the bootstrap in probability is proved for $m_n/n \to 0$, still require separate proofs.

Politis and Romano, loc. cit., also extend the above procedure to mixing data: In this case one samples one of the data points $X_1, \ldots, X_{n-m+1}$ at random and evaluates the statistic T_m at the m successive observations starting from the sampled point.

Finally we should mention that there is work on the choice of bootstrap sample size m_n and the accuracy and fine–tuning of the procedure (e.g., the loss incurred in simple cases and how to correct the procedure to make it more accurate). Of the three terms in (5.9), I_n is quite small (see (5.10)), so that m_n could be estimated to balance the sizes of II_n and III_n. In the case of the mean, with third moment finite, this would give $m_n = n^{1/2}$ and an error of the order of $n^{-1/4}$. One can do a better calibration using Edgeworth expasions and get, for the same statistic, $m_n = n^{2/3}$ and an error of the order of $n^{-1/3}$ (Politis and Romano, loc. cit.). In the case of the studentized mean, under quite stringent regularity conditions, examination of Edgeworth expansions show that the cdf of

$$\left(\frac{m}{n}\right)^{1/2} \mathcal{L}^*\big(\sqrt{m}(\bar{X}^*_m - \bar{X}_n)/\sigma^*_m\big) + \left(1 - \left(\frac{m}{n}\right)^{1/2}\right) N(0,1)$$

is $o_{\mathrm{P}}\big(n^{-1/2}\big)$ close to the cdf of $\sqrt{n}(\bar{X}_n - \mathbb{E}X)/\sigma_n$. This principle extrapolates to other not too irregular cases, but requires, for its application, relatively precise information on the Edgeworth expansion of the statistic of interest (Bertail, 1994; see also Bickel, Götze and van Zwet, 1994).

In another direction, we mention that Theorem 5.3 has a version for processes, in the setup of Chapter 2. We will omit its proof because the stament is not completely satisfactory: Suppose that B is now $\ell^\infty(T)$, the space of bounded functions on T, that $T_n : S^n \mapsto \ell^\infty(T)$ and that $\theta(\mathrm{P}) \in \ell^\infty(\mathrm{P})$. Suppose that, in the notation of Section 2.1 in the next chapter,

$$\tau_n \big[T_n(X_1, \ldots, X_n) - \theta(\mathrm{P})\big] \to_{\mathcal{L}} Z \quad \text{in} \quad \ell^\infty(T), \tag{5.5$'$}$$

and that τ_n and m_n satisfy conditions (5.6). Then, for every subsequence n' there is another subsequence n'' such that

$$d_{BL}\Big[\mathcal{L}^*\Big(\tau_{m_{n''}}\big(T_{m_{n''}}(Y^*_{n'',1}, \ldots, Y^*_{n'',m_{n''}}) - T_{n''}(X_1, \ldots, X_{n''})\big)\Big), \mathcal{L}(Z)\Big] \to 0 \quad \text{a.s.}$$

The problem is that we do not know how to prove measurablity of this d_{BL} or even of $d_{\mathcal{D}(\mathcal{L}(Z))}$ in order to deduce from this a bootstrap in probability. The proof goes along the same lines as the proof above, using the extension of Lemma 5.2 to this type of convergence (van der Vaart and Wellner, 1996, page 72, or a proof along the lines of Corollary 1.5 below) and simple non-measurable calculus. This version of Theorem 5.3 applies for example to U-processes, with $\tau_n \neq \sqrt{n}$ in the degenerate case (e.g. Arcones and Giné, 1993).

Chapter 2: On the bootstrap for empirical processes

In this chapter we will prove that the central limit theorem for empirical processes in the Vapnik–Červonenkis–Dudley general setting, can always be bootstrapped. This is important because it automatically gives the bootstrap of many limit theorems. We will also show that for a more restricted but still quite general set of classes of functions, basically any sensible model based bootstrap works as well. A small number of applications (to e.g. M–estimators) will be considered. However, the bootstrap for empirical processes based on stationary observations and the bootstrap for U–processes will be ignored because current results in the literature about these two sobjects do not seem to be in final form (see, e.g., Arcones and Giné, 1994, and Radulović, 1996b).

In accordance to standard practice in Statistics, in Chapter 1 we denoted $X^*_{n,i}$, $\Pr^*$, $\mathbb{E}^*$ and $\mathcal{L}^*$ respectively the bootstrap variables, and conditional probability, expectation and law given the sample. In this chapter, the same objects will be denoted as $X^b_{n,i}$, $\Pr^b$, $\mathbb{E}^b$ and $\mathcal{L}^b$ since we must reserve the superscript * for outer probability and expected value (the need for this did not arise in Chapter 1).

2.1. Background from empirical process theory. In this section we present the minimum amount of empirical process theory required for the bootstrap in probability of empirical processes, Efron's version, which is the simplest case: We want to give a clear idea about the amount of technique needed, with the hope of showing it is substantial but not excessive. We assume, however, some knowledge about weak convergence and about sums of independent random vectors (such as the clt in $\mathbb{R}$ and Lévy's maximal inequalities).

2.1.1. Convergence in law of sample bounded processes. Let T be a set and let $X_n(t)$, $t \in T$, be stochastic processes indexed by the set T. Assume all the sample paths of these processes are bounded functions on T. Then $X_n(\cdot) \in \ell^\infty(T)$, the space of all bounded real functions on T. $\ell^\infty(T)$, equipped with the sup norm, $\|\cdot\|_T$, is a Banach space, in particular a metric space, but we do not assume that the finite dimensional distributions of the processes $X_n(t)$, $\mathcal{L}\Big((X_n(t_1), \ldots, X_n(t_k))\Big)$, $k \in \mathbb{N}$, $t_i \in T$, correspond to the finite dimensional projections $\mu_{n,t_1,\ldots,t_k}$ of individually tight Borel probability measures μ_n on $\ell^\infty(T)$. Let now $X(t)$, $t \in T$, be a process whose finite dimensional laws do correspond to the finite dimensional laws of a tight Borel probability measure (= a Radon probability measure) on $\ell^\infty(T)$. X is sample bounded, and we continue denoting as X its version with bounded sample paths. Then, we say that X_n *converges in law to* X *uniformly in* $t \in T$, or that

$$X_n \to_{\mathcal{L}} X \quad \text{in} \quad \ell^\infty(T) \tag{1.2}$$

if

$$\mathbb{E}^* H(X_n) \to \mathbb{E} H(X)$$

for all functions $H : \ell^\infty(T) \to \mathbb{R}$ bounded and continuous. $\mathbb{E}^*$ here denotes outer expectation, as indicated above: the outer expectation of a not necessarily measurable function $H(X_n)$ is the infimum of the expected values of all the measurable functions a.s. larger than or equal to $H(X_n)$. As with regular convergence in law, if F is a continuous function on $\ell^\infty(T)$ with values in another metric space and if $F(X_n)$ is measurable, then (1.2) implies that $F(X_n) \to_{\mathcal{L}} F(X)$ in the usual way, and this is what makes the concept of convergence in law in $\ell^\infty(T)$ useful.

The outer expectation in the definition of convergence in law is necessary because even the most simple bounded processes may fail to induce Borel probability laws on $\ell^\infty(T)$. For instance, let $T = [0,1]$, let U be uniform on $[0,1]$ and let $X(t) = I_{(U,1]}(t)$, $t \in T$. Let A be a non-measurable subset of $[0,1]$ (which exists if we assume the axiom of choice) and let $F_A = \{I_{(s,1]} : s \in A\} \subset \ell^\infty(T)$. Then, F_A is a discrete set for the sup norm, hence a closed subset of $\ell^\infty(T)$. But $\{X \in F_A\} = \{U \in A\}$ is not measurable and therefore the law of X does not extend to a Borel probability measure on $\ell^\infty(T)$. A good reference for all the non-measurable calculus we use here is Dudley and Philipp (1983), Section 2. Other more extensive references for it are Andersen (1985) and Ziegler (1994). van der Vaart and Wellner (1996) contain also an excellent account of the non-measurable calculus required by empirical process theory.

The following lemma clarifies the notion of processes whose laws are tight Borel measures on $\ell^\infty(T)$ (tight Borel finite measures are also called Radon measures and we use both terms interchangeably). It is attributed to Hoffmann–Jørgensen in Andersen and Dobrić (1987). It was reproduced in Giné and Zinn (1986), with minor alterations.

1.1. LEMMA. *Let $X(t)$, $t \in T$, be a sample bounded stochastic process. Then the finite dimensional distributions of X are those of a tight Borel probability measure on $\ell^\infty(T)$ if and only if there exists on T a pseudo-distance d for which (T,d) is totally bounded and such that X has a version with almost all its sample paths uniformly continuous for d.*

PROOF. Let μ, a tight probability measure on $\ell^\infty(T)$, be the law of X, let K_n, $n \in \mathbb{N}$, be an increasing sequence of compact sets in $\ell^\infty(T)$ such that $\mu(\cup_{n=1}^\infty K_n) = 1$, and let $K = \cup_{n=1}^\infty K_n$. Then, it is easy to see that the pseudometric d defined on T by

$$d(s,t) = \sum_{n=1}^{\infty} 2^{-n}\big(1 \wedge d_n(s,t)\big)$$

with

$$d_n(s,t) = \sup\big\{|f(t) - f(s) : f \in K_n\big\},$$

makes (T,d) totally bounded (use that $\cup_{n=1}^m K_n$ is totally bounded and that for any finite number r of functions $f_i \in \cup_{n=1}^m K_n$, the set $\{(f_1(t),\ldots,f_r(t)) : t \in T\}$ is a totally bounded subset of $\mathbb{R}^r$; we skip the details). Moreover, the functions $f \in K$ are uniformly d–continuous since, if $f \in K_n$, then $|f(s) - f(t)| \le d_n(s,t) \le 2^n d(s,t)$. Since $\mu(K) = 1$, the identity map on $(\ell^\infty(T), \mathcal{B}, \mu)$ is a version of X with almost all its trajectories in K, hence in $C_u(T,d)$, the space of bounded uniformly d–continuous functions on T.

Conversely, let $X(t)$, $t \in T$, be a process with a version whose sample paths are almost all in $C_u(T,d)$ for a distance or pseudodistance d on T for which (T,d) is totally bounded. Then, since $C_u(T,d)$ is complete and separable, the law of X is a tight Borel probability measure on $C_u(T,d)$: by separability and the fact that the vectors $(X(t_1),\dots,X(t_k))$ are measurable, it follows that any uniformly continuous version of X is a measurable function from the basic probability space Ω into C_u and, C_u being Polish, its law is tight in C_u. But a tight Borel probability measure on $C_u(T)$ is a tight Borel measure on $\ell^\infty(T)$ since the inclusion of C_u into ℓ^∞ is continuous.

□

1.2. REMARK. Suppose that X has the property that for all $t_n \in T$, $\{X(t_n)\}$ is Cauchy in L_2 whenever it is Cauchy in probability, which is the case if X is Gaussian or if it is a Gaussian chaos process. Then, if X induces a tight Borel probability law on $\ell^\infty(T)$, it has a version with all its sample paths in $C_u(T, d_X)$, where $d_X(s,t) = \left[\mathbb{E}\big(X(t) - X(s)\big)^2\right]^{1/2}$, and the pseudo–metric space (T, d_X) is totally bounded. Here is a sketch of the proof: If $(\bar{T}, \bar{d})$ is the completion of the pseudo–metric space (T, d) prescribed by Lemma 1.1, then d_X extends to $\bar{T}$, and so does any version of X with sample paths in $C_u(T, d)$, by uniform continuity. The metric space $(\bar{T}/d_X, \bar{d})$ is compact and the identity map $j : (\bar{T}/d_X, \bar{d}) \to (\bar{T}/d_X, d_X)$ is continuous, hence, $(\bar{T}/d_X, d_X)$ is compact, in particular totally bounded, and the identity map is bicontinuous. Now one can use the separability of $(\bar{T} \times \bar{T}, \bar{d} \times \bar{d})$ to show that, except for a set of measure 0, the sample paths of X satisfy $X(s) = X(t)$ whenever $\rho_X(s,t) = 0$. Hence, almost all the sample paths of any version of X in $C_u(T, d)$ belong to $C_u(\bar{T}/d_X, \bar{d})$ and, by continuity of the map j^{-1}, also to $C_u(\bar{T}/d_X, d_X)$.

1.3. THEOREM. *Let X_n be processes on T all (or almost all) of whose sample paths are bounded. Then the following statements are equivalent:*

i) The finite dimensional distributions of the processes X_n converge in law and there exists a pseudometric d on T such that (T, d) is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \Pr{}^*\Big\{\sup_{d(s,t)\le\delta} \big|X_n(t) - X_n(s)\big| > \varepsilon\Big\} = 0; \tag{1.3}$$

for all $\varepsilon > 0$.

ii) There exists a process X whose finite dimensional distributions are those of a tight Borel probability measure on $\ell^\infty(T)$ and such that

$$X_n \to_{\mathcal{L}} X \ \text{ in } \ \ell^\infty(T).$$

If i) holds, then the process X in ii), can be chosen to have almost all its sample paths bounded and uniformly continous for d. If X in ii) has a version with almost all of its trajectories in $C_u(T, \rho)$ for a pseudodistance ρ for which (T, ρ) is totally bounded, then the distance d in i) can be taken to be $d = \rho$.

PROOF. Let us assume i) holds. Let T_0 be a countable d–dense subset of T, and let T_k, $k \in \mathbb{N}$, be finite sets incresing to T_0. The limit laws of the finite dimensional distributions of the processes X_n are compatible and thus define a stochastic process

X on T, and moreover,

$$\begin{aligned}\Pr\Big\{\max_{d(s,t)\le\delta,\ s,t\in T_k}|X(t)-X(s)|>\varepsilon\Big\}&\\ &\le\liminf_{n\to\infty}\Pr\Big\{\max_{d(s,t)\le\delta,\ s,t\in T_k}|X_n(t)-X_n(s)|>\varepsilon\Big\}\\ &\le\liminf_{n\to\infty}\Pr\Big\{\max_{d(s,t)\le\delta,\ s,t\in T_0}|X_n(t)-X_n(s)|>\varepsilon\Big\}.\end{aligned}$$

Hence, taking limits as $k\to\infty$ and using condition (1.3), we obtain that there exists a sequence $\delta_r\searrow 0$, $\delta_r>0$, such that

$$\Pr\Big\{\sup_{d(s,t)\le\delta_r,\ s,t\in T_0}|X(t)-X(s)|>2^{-r}\Big\}\le 2^{-r}.$$

Then, by Borel–Cantelli, there exists $r(\omega)<\infty$ a.s. such that

$$\sup_{d(s,t)\le\delta_r,\ s,t\in T_0}|X(t,\omega)-X(s,\omega)|\le 2^{-r}$$

for all $r>r(\omega)$. Hence, the restriction of $X(t,\omega)$ to T_0 is a d–uniformly continuous function of $t\in T_0$ for almost every ω; T being totally bounded, $X(t,\omega)$, $t\in T_0$, is also bounded. X being continuous in probability, the extension to T by uniform continuity of the restriction of X to T_0 (only the ω set where X is uniformly continuous needs be considered) produces a version of X whose trajectories are all in $C_u(T,d)$ and, in particular, the law of X admits a tight extension to the Borel σ–algebra of $\ell^\infty(T)$ (Lemma 1.1).

Before proving convergence, we recall a useful fact (whose simple proof we omit): if $f:\ell^\infty(T)\to\mathbb{R}$ is bounded and continuous, and if $K\subset\ell^\infty(T)$ is compact, then for every $\varepsilon>0$ there exists $\delta>0$ such that

$$\|u-v\|_T<\delta,\quad u\in K,\quad v\in\ell^\infty(T)\quad\Longrightarrow\quad |f(u)-f(v)|<\varepsilon. \tag{1.4}$$

Since (T,d) is totally bounded, for every $\tau>0$ there exists a finite set of points $t_1,\dots,t_{N(\tau)}$ which is τ–dense in (T,d) in the sense that $T\subseteq\cup_{i=1}^{N(\tau)}B(t_i,\tau)$, where $B(t,\tau)$ denotes the open ball of center t and radius τ. Then, for each $t\in T$ we can choose $\pi_\tau(t)\in\{t_1,\dots,t_{N(\tau)}\}$ so that $d(\pi_\tau(t),t)<\tau$. We then define processes $X_{n,\tau}$, $n\in\mathbb{N}$, and X_τ as

$$\begin{aligned}X_{n,\tau}(t)&=X_n(\pi_\tau(t)),\\ X_\tau(t)&=X(\pi_\tau(t)),\quad t\in T.\end{aligned}$$

These are approximations of X_n and X taking only a finite number $N(\tau)$ of values. Convergence of the finite dimensional distributions of X_n to those of X implies that

$$X_{n,\tau}\to_{\mathcal{L}}X_\tau\ \text{ in }\ \ell^\infty(T). \tag{1.5}$$

Moreover, the uniform continuity of the sample paths of X implies

$$\lim_{\tau\to 0}\|X-X_\tau\|_T=0\ \text{ a.s.} \tag{1.6}$$

Let now $f : \ell^\infty(T) \to \mathbb{R}$ be a bounded continuous function. We have

$$\begin{aligned}\left|\mathbb{E}^* f(X_n) - \mathbb{E}f(X)\right| &\le \left|\mathbb{E}^* f(X_n) - \mathbb{E}f(X_{n,\tau})\right| \\ &\quad + \left|\mathbb{E}f(X_{n,\tau}) - \mathbb{E}f(X_\tau)\right| + \left|\mathbb{E}f(X_\tau) - \mathbb{E}f(X)\right| \\ &:= I_{n,\tau} + II_{n,\tau} + III_\tau.\end{aligned}$$

In order to prove that ii) holds we must show that the iterated limit $\lim_{\tau\to 0}\limsup_n$ of each of these three quantities is 0. This is true for $II_{n,\tau}$ by (1.5). Next we show it for III_τ. Given $\varepsilon > 0$ let $K \subset \ell^\infty(T)$ be a compact set such that $\Pr\{X \in K^c\} < \varepsilon/(6\|f\|_\infty)$, let $\delta > 0$ be such that (1.4) works for K and $\varepsilon/6$, and let $\tau_1 > 0$ be such that $\Pr\{\|X_\tau - X\|_T \ge \delta\} < \varepsilon/(6\|f\|_\infty)$ for all $\tau > \tau_1$ (possible by (1.6)). Then,

$$\begin{aligned}\left|\mathbb{E}f(X_\tau) - \mathbb{E}f(X)\right| &\le 2\|f\|_\infty \Pr\{X \in K^c \text{ or } \|X_\tau - X\|_T \ge \delta\} \\ &\quad + \sup\{|f(u) - f(v)| : u \in K, \|u - v\|_T < \delta\} \\ &\le 2\|f\|_\infty\Big(\frac{\varepsilon}{6\|f\|_\infty} + \frac{\varepsilon}{6\|f\|_\infty}\Big) + \frac{\varepsilon}{6} < \varepsilon,\end{aligned}$$

proving $\lim_{\tau\to 0} III_\tau = 0$. For the same ε, δ and K, we have

$$\begin{aligned}\left|\mathbb{E}^* f(X_n) - \mathbb{E}f(X_{n,\tau})\right| &\le 2\|f\|_\infty\Big[\Pr{}^*\{\|X_n - X_{n,\tau}\|_T \ge \frac{\delta}{2}\} + \Pr\{X_{n,\tau} \in \left(K_{\delta/2}\right)^c\}\Big] \\ &\quad + 2\sup\{|f(u) - f(v)| : u \in K, \|u - v\|_T < \delta\}, \qquad (1.7)\end{aligned}$$

where $K_{\delta/2}$ is the $\delta/2$ open neighborhood of the set K for the sup norm. (To verify inequality (1.7) note that if $X_{n,\tau} \in K_{\delta/2}$ and $\|X_n - X_{n,\tau}\|_T < \delta/2$ then there exists $u \in K$ such that $\|u - X_{n,\tau}\| < \delta/2$ and $\|u - X_n\|_T < \delta$.) Since the hypothesis (1.3) implies that there is $\tau_2 > 0$ such that

$$\limsup_{n\to\infty} \Pr{}^*\Big\{\|X_{n,\tau} - X_n\|_T \ge \frac{\delta}{2}\Big\} < \frac{\varepsilon}{6\|f\|_\infty}$$

for all $\tau < \tau_2$, and finite dimensional convergence gives

$$\limsup_{n\to\infty} \Pr\Big\{X_{n,\tau} \in \left(K_{\delta/2}\right)^c\Big\} \le \Pr\Big\{X_\tau \in \left(K_{\delta/2}\right)^c\Big\} \le \frac{\varepsilon}{6\|f\|_\infty},$$

we obtain from (1.7), as above, that for $\tau < \tau_1 \wedge \tau_2$, $\limsup_{n\to\infty}\left|\mathbb{E}^* f(X_n) - \mathbb{E}f(X_{n,\tau})\right| < \varepsilon$, showing that $\lim_{\tau\to 0}\limsup_{n\to\infty} I_{n,\tau} = 0$. Hence, i) implies ii).

For the converse, we first observe that if $X_n \to_{\mathcal{L}} X$ in $\ell^\infty(T)$ then, as for regular convergence in law, $\limsup_{n\to\infty}\Pr^*\{X_n \in F\} \le \Pr\{X \in F\}$ for every closed set of $\ell^\infty(T)$. The proof is the same as for regular convergence in law and is omitted (it can be found in many texts under the heading 'portmanteau's lemma'). Suppose now that ii) holds. Then, by Lemma 1.1, there exists a pseudodistance d on T for which (T, d) is totally bounded and such that X has a version (that we still denote by X) with all its sample paths in $C_u(T, d)$. Take $F_{\delta,\varepsilon} = \{u \in \ell^\infty(T) : \sup_{d(s,t)\le\delta}|u(s) - u(t)| \ge \varepsilon\}$. Then, applying the previous observation to $F_{\delta,\varepsilon}$, we obtain, by the convergence hypothesis, that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr{}^*\Big\{\sup_{d(s,t)\le\delta}\big|X_n(t)-X_n(s)\big|\ge\varepsilon\Big\}$$
$$\le\lim_{\delta\to 0}\Pr{}^*\Big\{\sup_{d(s,t)\le\delta}\big|X(t)-X(s)\big|\ge\varepsilon\Big\}=0$$

for all $\varepsilon>0$.

□

Theorem 1.3 is contained, in one form or other, in Hoffmann–Jørgensen (1984), is attributed to Hoffmann–Jørgensen in Andersen and Dobrić (1987), and the proof given here is just the proof of Theorem 1.3 in Giné and Zinn (1986) with only formal changes. See also Dudley (1984, Theorem 4.1.1).

1.4. REMARK. It is not necessary to check that all the finite dimensional marginals of the processes X_n converge in distribution, when (T,d) is totally bounded and the asymptotic equicontinuity condition (1.3) holds, in order to conclude that X_n converges in law in $\ell^\infty(T)$: It suffices to check convergence in distribution of the marginals of $X_n(t)$ for points t in a *dense* subset D of (T,d). This is obvious because one can choose $t_1,\dots,t_{N(\tau)}$ in D in the definition of π_τ, just below (1.4) in the previous proof (since only the marginals $(X_n(t_1),\dots,X_n(t_{N(\tau)}))$ are involved in the proof of the theorem. This remark, for D countable, simplifies some proofs below.

Let $BL_1\big(\ell^\infty(T)\big)$ be the set of all real functionals f on $\ell^\infty(T)$ such that $\sup_{x\in\ell^\infty(T)}|f(x)|\le 1$ and $\sup_{x\ne y,x,y\in\ell^\infty(T)}|f(y)-f(x)|/\|y-x\|_T\le 1$ (BL_1 stands for the unit ball of the space of bounded Lipschitz functions). It is well known that if B is a separable metric space then the distance between probability measures on B $d_{BL}(\mu,\nu):=\sup\big\{|\int fd(\mu-\nu)|:f\in BL_1(B)\big\}$ metrizes weak convergence (e.g. Dudley, 1989, page 310, or Araujo and Giné, 1980, pages 10–11). If Y is a process on T with almost all its trajectories bounded and X a process whose law is a tight Borel measure on $\ell^\infty(T)$, and we also denote by X one of its versions almost all of whose sample paths are in $\ell^\infty(T)$, we define

$$d_{BL}\big(Y,X\big):=\sup\Big\{\big|\mathbb{E}^*f(Y)-\mathbb{E}f(X)\big|:f\in BL_1(\ell^\infty(T))\Big\}. \tag{1.8}$$

With this definition we have the following corollary (it does not add much to Theorem 1.3 but is sometimes useful).

1.5. COROLLARY. *If the law of X is defined by a tight Borel mesure on $\ell^\infty(T)$ and almost all the trajectories of the processes $X_n(t)$, $t\in T$, are bounded, then*

$$X_n\to_{\mathcal{L}}X\ \text{ in }\ \ell^\infty(T)$$

if and only if

$$d_{BL}(X_n,X)\to 0.$$

PROOF. Let $X_n\to_{\mathcal{L}}X$. We use the same letter X to denote a version of X whose paths are all in $C_u(T,d)$ for some d for which (T,d) is totally bounded, and keep the definitions and notation from the proof of Theorem 1.3. Consider the decomposition of the previous proof,

$$\big|\mathbb{E}^* f(X_n) - \mathbb{E} f(X)\big| \le I_{n,\tau} + II_{n,\tau} + III_\tau.$$

Then, since, for τ fixed, $X_{n,\tau} \to_{\mathcal{L}} X_\tau$ as random vectors in $\mathbb{R}^{N(\tau)}$, a convergence metrized by d_{BL}, it follows that

$$\lim_{n\to\infty} \sup_{f\in BL_1(\ell^\infty(T))} |II_{n,\tau}| = 0$$

for all $\tau > 0$. Since $\|X_\tau - X\|_T \to 0$ a.s.,

$$\lim_{\tau\to 0} \sup_{f\in BL_1(\ell^\infty(T))} |III_\tau| \le \lim_{\tau\to 0} \mathbb{E}\big(\|X_\tau - X\|_T \wedge 2\big) = 0$$

by bounded convergence. Finally,

$$\begin{aligned}\lim_{\tau\to 0} \limsup_{n\to\infty} \sup_{f\in BL_1(\ell^\infty(T))} |I_\tau| &\le \mathbb{E}^*\Big[\sup_{d(s,t)\le\tau} \big|X_n(t) - X_n(s)\big| \wedge 2\Big] \\ &\le 2 \lim_{\tau\to 0} \limsup_{n\to\infty} \Pr{}^*\Big\{\sup_{d(s,t)\le\tau} \big|X_n(t) - X_n(s)\big| > \varepsilon\Big\} + \varepsilon = \varepsilon\end{aligned}$$

for all $\varepsilon > 0$, by Theorem 1.3 ((1.3)). Thus, $d_{BL}(X_n, X) \to 0$.

Conversely, let us assume $d_{BL}(X_n, X) \to 0$. For $\delta > 0$ fixed, and all $\varepsilon, \delta > 0$, let

$$A_{\delta,\varepsilon} = \Big\{x \in \ell^\infty(T) : \sup_{d(s,t)\le\delta} |x(t) - x(s)| \ge \varepsilon\Big\}.$$

If $x \in A_{\delta,\varepsilon}$ and $y \in A^c_{\delta,\varepsilon/2}$ then $\|x - y\|_T \ge \varepsilon/5$. Therefore, the restriction to the set $A_{\delta,\varepsilon} \cup A^c_{\delta,\varepsilon/2}$ of the function $I_{A_{\delta,\varepsilon}}$ is Lipschitz with constant bounded by $5/\varepsilon$. Hence, by the Kirzbraun–McShane extension theorem (e.g., Dudley, 1989, page 141 or Araujo and Giné, 1980, pages 2–3), there exists a bounded Lipschitz function f on $\ell^\infty(T)$ non–negative, bounded by 1 and with Lipschitz constant bounded by $5/\varepsilon$ which is 0 on $A^c_{\delta,\varepsilon/2}$ and 1 on $A_{\delta,\varepsilon}$. Then, the assumption implies

$$\begin{aligned}\limsup_{n\to\infty} \Pr{}^*\{X_n \in A_{\delta,\varepsilon}\} &\le \limsup_{n\to\infty} \mathbb{E}^* f(X_n) = Ef(X) \\ &\le \Pr\Big\{\sup_{d(s,t)\le\delta} \big|X(t) - X(s)\big| > \frac{\varepsilon}{2}\Big\}.\end{aligned}$$

Now, the asymptotic equicontinuity condition (1.3) follows from the uniform continuity with respect to d of the sample paths of X. Then, Theorem 1.3 gives $X_n \to_{\mathcal{L}} X$.

□

Note that in the converse part of this proof, if X is fixed, it is sufficient to assume that $\sup |\mathbb{E}^* f(X_n) - \mathbb{E} f(X)| \to 0$ wirth the sup taken only over a countable subset of

$BL_1(\ell^\infty(T))$ (instead of all of BL_1), the subset consisting of one $f_{\delta,\varepsilon}$ for each $A_{\delta,\varepsilon}$, defined as being Lipschitz with norm 1, zero on $A^c_{\delta,\varepsilon/2}$ and taking the constant value $(\varepsilon/5) \wedge 1$ on $A_{\delta,\varepsilon}$ (such functions exist by the Kirzbraun–McShane theorem).

This comment extends Lemma 5.2 in Chapter 1 to the present setting (as in van der Vaart and Wellner, 1996, Theorem 1.12.2).

Corollary 1.5 has been independently observed by several authors, e.g., Giné and Zinn (1990), to define bootstrap in probability and, more formally Dudley (1990), also in connection with the bootstrap, and van der Vaart and Wellner in a 1989 preprint.

2.1.2. Symmetrization, Lévy-type and Hoffmann–Jørgensen inequalities. The inequalities in this section are basic for most further developments. We will present them in the language of empirical processes, that is, for random processes of the form $\sum f(X_i)$ where X_i are coordinates in a product probability space and f runs over a class of functions $\mathcal{F}$ (instead of the more usual form of sums of independently formed random elements with values in a general Banach space). Some of the proofs will be omitted, particularly if they can be found in Ledoux and Talagrand's 1991 book or in Giné and Zinn's 1986 lecture notes. Here is the general setting. Given $(S, \mathcal{S}, \mathrm{P})$, a probability space, we set $(\Omega, \Sigma, \Pr) = \left(S^{\mathbb{N}} \times \Omega', \mathcal{S}^{\mathbb{N}} \times \Sigma', \mathrm{P}^{\mathbb{N}} \times \Pr'\right)$ and let X_i, $i \in \mathbb{N}$, be the coordinate functions $X_i(s, \omega') = X_i(s) = s_i$, $s \in S^{\mathbb{N}}$. Then, the variables X_i are i.i.d.(P). Sometimes, we will allow P to vary from coodinate to coordinate, that is, we will take $\prod_{i=1}^{\infty} P_i$ instead of $P^{\mathbb{N}}$. We will let $\mathcal{F}$ ba a collection of measurable functions on S, and we will impose, often without explicitly mentioning it, that either

$$F_c(s) := \sup\left\{|f(s) - \mathrm{P}f| : f \in \mathcal{F}\right\} < \infty \text{ for all } s \in S, \tag{1.9}$$

or that

$$F(s) := \sup\left\{|f(s)| : f \in \mathcal{F}\right\} < \infty \text{ for all } s \in S, \tag{1.10}$$

according as to whether we are considering the processes $f(X_i) - P$ or $f(X_i)$, $f \in \mathcal{F}$; in this way these processes are random elements of the space $\ell^\infty(\mathcal{F})$ of all bounded functions on $\mathcal{F}$. $\ell^\infty(\mathcal{F})$, equipped with the sup norm, that we denote $\|\cdot\|_{\mathcal{F}}$, is a Banach space.

We are interested in comparing the tail probabilities and the p–th moments of the sup norms of processes of the form $\sum_{i=1}^{n} f(X_i)$ (or $\sum_{i=1}^{n}\left(f(X_i) - Pf\right)$) with their symmetrized or randomized counterparts. We are also intersted on Lévy type maximal inequalities and an inequality of Hoffmann–Jørgensen (extending the converse Kolmogorov maximal inequality) for these processes. The newest material in this section is a Lévy type maximal inequality for sums of i.i.d. but not necessarily symmetric random elements taking values in a Banach space, due to Montgomery–Smith (1994).

We begin with randomization inequalities for expected values, which are easiest. They are standard under measurablity, but require some comments when measurablity is not assumed. We summarize in the next lemma most of the facts we need on the calculus of outer expectations, including the symmetrization inequalities. See e.g. Dudley and Philipp (1983), Andersen (1985) or van der Vaart and Wellner (1996) for the definition of outer integral, $\mathbb{E}^*$ and for the fact that $\mathbb{E}^* f = \mathbb{E} f^*$, where f^* is the measurable envelope or cover of f. This lemma can also be found in van der Vaart and Wellner's 1996 book.

1.6. LEMMA. *a) If X and Y are the coordinate functions of $(S^2, \mathcal{S}, \mathrm{P}\times\mathrm{Q})$ and if $\mathbb{E}f(Y) = 0$ for all $f \in \mathcal{F}$, then*

$$\mathbb{E}^*\|f(X)\|_{\mathcal{F}} \le \mathbb{E}^*\|f(X) + f(Y)\|_{\mathcal{F}}. \tag{1.11}$$

b) If X and Y are coordinate functions as in a) and if $g(x,y) \ge 0$ is a (not necessarily measurable) real function on S^2, then, letting $\mathbb{E}_X$ (resp. $\mathbb{E}_Y$) denote integration with respect to P (resp. Q), we have

$$\mathbb{E}_X^*\mathbb{E}_Y^* g(X,Y) \le \mathbb{E}g^*(X,Y) = \mathbb{E}^*g(X,Y), \tag{1.12}$$

and if, moreover, Y is discrete,

$$\mathbb{E}_X^*\mathbb{E}_Y g(X,Y) \le \mathbb{E}^*g(X,Y) = \mathbb{E}_Y\mathbb{E}_X^* g(X,Y). \tag{1.13}$$

c) If ε_i, $i \in \mathbb{N}$, is a Rademacher sequence defined on the Ω' part of the basic probability space, and X_i, $i \in BbbN$, are the coordinate functions defined on the $S^{\mathbb{N}}$ part, then

$$\frac{1}{2}\mathbb{E}^*\Big\|\sum_{i=1}^n \varepsilon_i(f(X_i) - \mathrm{P}f)\Big\|_{\mathcal{F}} \le \mathbb{E}^*\Big\|\sum_{i=1}^n (f(X_i) - \mathrm{P}f)\Big\|_{\mathcal{F}} \le 2\mathbb{E}^*\Big\|\sum_{i=1}^n \varepsilon_i(f(X_i) - \mathrm{P}f)\Big\|_{\mathcal{F}}, \tag{1.14}$$

where $\mathrm{P}f$ can be deleted from the expression at the right.

PROOF. To prove a) we just note that, by Jensen, monotonicity of $\mathbb{E}^*$, and Tonelli,

$$\begin{aligned}\mathbb{E}^*\|f(X)\|_{\mathcal{F}} &= \mathbb{E}^*\Big(\sup_{f\in\mathcal{F}} |f(X) + \mathbb{E}f(Y)|\Big) \le \mathbb{E}^*\Big(\sup_{f\in\mathcal{F}} \mathbb{E}_Y|f(X) + f(Y)|\Big)\\ &\le \mathbb{E}_X\mathbb{E}_Y\|f(X) + f(Y)\|_{\mathcal{F}}^* = \mathbb{E}\|f(X) + f(Y)\|_{\mathcal{F}}^*\\ &= \mathbb{E}^*\|f(X) + f(Y)\|_{\mathcal{F}},\end{aligned}$$

proving (1.11).

For b) we first note that

$$\mathbb{E}_X^*\mathbb{E}_Y^* g(X,Y) \le \mathbb{E}_X\mathbb{E}_Y g^*(X,Y) = \mathbb{E}g^*(X,Y) = \mathbb{E}^*g(X,Y)$$

since g^* is jointly measurable, and inequality (1.12) follows (the left side of (1.13) follows as well because $g(X,Y)$ is a measurable function of Y if Y is discrete). Suppose now Y is discrete and denote $g^{*,X}(X,Y)$ the measurable envelope of $g(\cdot,Y)$, Y fixed. Note $g^{*,X}(X,Y) = \sum I_{Y=r}g^{*,X}(X,r)$ by definition. We have $g^*(X,Y) \le \sum I_{Y=r}g^{*,X}(X,r)$ a.s. because the latter function is jointly measurable; so, $g^*(X,Y) \le g^{*,X}(X,Y)$ a.s. The reverse inequality is obvious because, by Fubini, $g^*(X,Y)$ is measurable in X for each value of Y. Hence,

$$\mathbb{E}_Y\mathbb{E}_X^* g(X,Y) = \mathbb{E}_Y\mathbb{E}_X g^{*,X}(X,Y) = \mathbb{E}_Y\mathbb{E}_X g^*(X,Y) = \mathbb{E}^*g(X,Y),$$

giving the identity at the right of (1.13).

To prove (1.14), we can assume $\mathrm{P}f(X_i) = 0$. Now,

$$\begin{aligned}
\mathbb{E}^*\Big\|\sum_{i=1}^n \varepsilon_i f(X_i)\Big\|_{\mathcal{F}} &= \mathbb{E}_\varepsilon \mathbb{E}_X^*\Big\| \sum_{i:\varepsilon_i=1, i\le n} \varepsilon_i f(X_i) + \sum_{i:\varepsilon_i=-1, i\le n} \varepsilon_i f(X_i)\Big\|_{\mathcal{F}} \\
&\le \mathbb{E}_\varepsilon \mathbb{E}_X^*\Big\| \sum_{i:\varepsilon_i=1, i\le n} \varepsilon_i f(X_i)\Big\|_{\mathcal{F}} + \mathbb{E}_\varepsilon \mathbb{E}_X^*\Big\| \sum_{i:\varepsilon_i=-1, i\le n} \varepsilon_i f(X_i)\Big\|_{\mathcal{F}} \\
&\le 2\mathbb{E}_\varepsilon \mathbb{E}_X^*\Big\|\sum_{i=1}^n f(X_i)\Big\|_{\mathcal{F}} = 2\mathbb{E}^*\Big\|\sum_{i=1}^n f(X_i)\Big\|_{\mathcal{F}},
\end{aligned}$$

where the identity follows from (1.13), the first inequality is obvious and the second inequality is justified by (1.11). This gives the left side of (1.14). For the right side, increase the basic probability space with another copy of $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, \mathrm{P}^{\mathbb{N}})$, and denote the new coordinate functions by X_i'. Then, by invariance of $\mathrm{P}^{\mathbb{N}} \times \mathrm{P}^{\mathbb{N}}$ under permutations of the coordinates, we have that if τ_i, $i = 1, \ldots, n$ is any sequence of ± 1's,

$$\mathbb{E}^*\Big\|\sum_{i=1}^n f(X_i)\Big\|_{\mathcal{F}} \le \mathbb{E}^*\Big\|\sum_{i=1}^n (f(X_i) + f(X_i'))\Big\|_{\mathcal{F}} = \mathbb{E}^*\Big\|\sum_{i=1}^n \tau_i (f(X_i) + f(X_i'))\Big\|_{\mathcal{F}}.$$

Therefore, using (1.12) once more,

$$\begin{aligned}
\mathbb{E}^*\Big\|\sum_{i=1}^n f(X_i)\Big\|_{\mathcal{F}} &= \mathbb{E}_\varepsilon \mathbb{E}_X^*\Big\|\sum_{i=1}^n \varepsilon_i (f(X_i) + f(X_i'))\Big\|_{\mathcal{F}} \\
&\le 2\mathbb{E}_\varepsilon \mathbb{E}_X^*\Big\|\sum_{i=1}^n \varepsilon_i f(X_i)\Big\|_{\mathcal{F}} = \mathbb{E}^*\Big\|\sum_{i=1}^n \varepsilon_i f(X_i)\Big\|_{\mathcal{F}}. \qquad \square
\end{aligned}$$

The inequalities between tail probabilities of the centered empirical process and its randomized version are not as neat as (1.14). They are as follows:

1.7. LEMMA. *Let ε_i, $i \in \mathbb{N}$, be a Rademacher sequence defined on the Ω' part of the basic probability space, in particular independent of the sequence $\{X_i\}$, which are the coordinates of $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}})$, and let $\mathcal{F}$ be a class of measurable functions on $(S, \mathcal{S}, \mathrm{P})$. Then: a) for all $t > 0$ and $n \in \mathbb{N}$, we have*

$$\begin{aligned}
\Pr{}^*\Big\{\Big\|\sum_{i=1}^n \varepsilon_i f(X_i)\Big\|_{\mathcal{F}} > t\Big\} &\le 2\max_{k\le n} \Pr{}^*\Big\{\Big\|\sum_{i=1}^k f(X_i)\Big\|_{\mathcal{F}} > \frac{t}{2}\Big\} \\
&\le 20\Pr{}^*\Big\{\Big\|\sum_{i=1}^n f(X_i)\Big\|_{\mathcal{F}} > \frac{t}{128}\Big\}, \qquad (1.15)
\end{aligned}$$

and b) if

$$\alpha^2 := \sup_{f\in\mathcal{F}} \mathrm{Var}_{\mathrm{P}}(f) < \infty,$$

then

$$\Pr^*\Big\{\Big\|\sum_{i=1}^n (f(X_i)-\mathrm{P}f)\Big\|_{\mathcal{F}} > t\Big\} \le 4\Pr^*\Big\{\Big\|\sum_{i=1}^n \varepsilon_i(f(X_i)-\mathrm{P}f)\Big\|_{\mathcal{F}} > \frac{t-2^{1/2}\alpha n^{1/2})}{2}\Big\}, \tag{1.16}$$

where $\mathrm{P}f$ *can be deleted from the term at the right.*

PROOF. The easy first inequality in (1.15) is obtained by developing the integral with respect to the Rademacher variables after applying (1.13), and comes from Giné and Zinn (1984), Lemma 2.3 (a) (= (1986), Lemma 2.3 (a)). The deeper second inequality in (1.15) has been recently obtained by Montgomery–Smith (1994) and will be proved immediately below (Theorem 1.10). Inequality (1.16) is Lemma 2.7 (b) in Giné and Zinn (1984) (= Lemma 2.3 (b) in Giné and Zinn, 1986). It is based on a more basic symmetrization inequality (Lemma 2.5, loc. cit. '84 =Lemma 2.1, loc. cit. '86).

□

Now we turn to Lévy type inequalities. Kahane's proof of the classical Lévy inequalities for independent symmetric random variables (e.g., Araujo and Giné, 1980, or Ledoux and Talagrand, 1991) extends to the present setting with only formal changes that we omit.

1.8. THEOREM. (Lévy's inequalities). *For* $n \in \mathbb{N}$, *let* X_i, $i \le n$, *be the coordinate functions on the product probability space* $(S^n, \mathcal{S}^n, \prod_{i=1}^n \mathrm{P}_i)$, *and let* $\mathcal{F}$ *be a class of measurable functions on* S *such that* $f(X_i)$, $i \le n$, *is a symmetric random variable. Then,*

$$\Pr^*\Big\{\max_{k\le n}\|\sum_{i=1}^k f(X_i)\|_{\mathcal{F}} > t\Big\} \le 2\Pr^*\Big\{\|\sum_{i=1}^n f(X_i)\|_{\mathcal{F}} > t\Big\} \tag{1.17}$$

and

$$\Pr^*\Big\{\max_{k\le n}\|f(X_k)\|_{\mathcal{F}} > t\Big\} \le 2\Pr^*\Big\{\|\sum_{i=1}^n f(X_i)\|_{\mathcal{F}} > t\Big\} \tag{1.18}$$

for all $t > 0$.

Ottaviani's inequality, with a complement, is contained in the following lemma (e.g. Kwapień and Woyczynski (1992)):

1.9. LEMMA. *For* $n \in \mathbb{N}$, *let* X_i, $i \le n$, *be the coordinate functions on the product probability space* $(S^n, \mathcal{S}^n, \prod_{i=1}^n \mathrm{P}_i)$, *and let* $\mathcal{F}$ *be a class of measurable functions. Then,*

$$\Pr\Big\{\max_{k\le n}\|\sum_{i=1}^k f(X_i)\|_{\mathcal{F}}^* > t\Big\} \le 3\max_{k\le n}\Pr^*\Big\{\|\sum_{i=1}^k f(X_i)\|_{\mathcal{F}} > \frac{t}{3}\Big\} \tag{1.19}$$

PROOF. Let us set $S_k(f) := \sum_{i=1}^k f(X_i)$ for all k and drop the subindex $\mathcal{F}$ from the norm signs. Define, for all $u, v \ge 0$ and $1 \le k \le n$, $A_k = \{\|S_i\|^* \le u + v \text{ for } i < k, \text{ and } \|S_k\|^* > u + v\}$. The sets A_k are disjoint and their union is

$\{\max_{k\leq n}\|S_k\|^* > u+v\}$. $\|S_n - S_k\|^*$ and $\|S_k\|^*$ are independent random variables by Dudley and Philipp (1983, Lemma 2.3). Therefore,

$$\begin{aligned}
\Pr\{\|S_n\|^* > u\} &\geq \Pr\{\|S_n\|^* > u,\ \max_{k\leq n}\|S_k\|^* > u+v\} \\
&\geq \sum_{k=1}^{n}\Pr\{A_k \cap \{\|S_n - S_k\|^* \leq v\}\} \\
&= \sum_{k=1}^{n}\Pr(A_k)\Pr\{\|S_n - S_k\|^* \leq v\} \\
&\geq \left[1 - \max_{k\leq n}\Pr\{\|S_n - S_k\|^* > v\}\right]\Pr\{\max_{k\leq n}\|S_k\|^* > u+v\} \qquad (1.20)
\end{aligned}$$

This is the Lévy–Ottaviani inequality. Taking $u = t/3$ and $v = 2t/3$ in this inequality gives

$$\begin{aligned}
\Pr\{\max_{k\leq n}\|S_k\|^* > t\} &\leq \frac{\Pr\{\|S_n\|^* > t/3\}}{1 - \max_{k\leq n}\Pr\{\|S_n - S_k\|^* > 2t/3\}} \\
&\leq \frac{\max_{k\leq n}\Pr\{\|S_k\|^* > t/3\}}{1 - 2\max_{k\leq n}\Pr\{\|S_k\|^* > t/3\}}.
\end{aligned}$$

This proves (1.19) if $\max_{k\leq n}\Pr^*\{\|S_k\| > t/3\} < 1/3$. But (1.19) is trivially satisfied otherwise. (We should recall here that for any B–valued random element U, $\Pr^*\{\|U\| > t\} = \Pr\{\|U\|^* > t\}$: see e.g. Andersen, 1985, page I.14 or van der Vaart and Wellner, 1996, page 7.)

□

Next we will give the main argument to prove that Lévy's maximal inequalities, with different constants, are also true if the symmetry assumption is replaced by the assumption of identical distribution of the X_i's. This remarkable result is due to Montgomery–Smith (1994) (he proved it for measurable random elements, but his proof goes thru without measurability because the measurable outer envelope of a norm basically works like a norm).

1.10. THEOREM. *For $n \in \mathbb{N}$, let X_i, $i \leq n$, be the coordinate functions on the product probability space $(S^n, \mathcal{S}^n, P^n)$, and let $\mathcal{F}$ be a class of measurable functions. Then, for $1 \leq k \leq n$,*

$$\Pr^*\Big\{\|\sum_{i=1}^{k} f(X_i)\|_{\mathcal{F}} > t\Big\} \leq 10\Pr^*\Big\{\|\sum_{i=1}^{n} f(X_i)\| > \frac{t}{64}\Big\}. \qquad (1.21)$$

For $n = 2$ this theorem has a surprisingly simple proof: Let X, Y, Z, be i.i.d. random vectors. Then,

$$\begin{aligned}
\Pr\{\|X\| > t\} &\leq \Pr\{\|(X+Y) + (X+Z) - (Y+Z)\| > 2t\} \\
&\leq 3\Pr\{\|X+Y\| > \frac{2t}{3}\},
\end{aligned}$$

and the same argument works with non–measurable vectors and outer probabilities.

The general case is more delicate and its proof rests on the lemma that follows. First, an auxiliary definition: we say that $x \in \ell^\infty(\mathcal{F})$ is a *t-concentration point* for the process $f(X)$, $f \in \mathcal{F}$ if $\Pr^*\{\|f(X) - x(f)\|_{\mathcal{F}} > t\} \le 1/10$.

1.11. LEMMA. *Let X_i, $i \le n$, and $\mathcal{F}$ be as in Theorem 1.10. If $S_j(f) = \sum_{i=1}^{j} f(X_i)$, $f \in \mathcal{F}$, has a t–concentration point s_j for $1 \le j \le k \le n$, then*

$$\|ks_j - js_k\|_{\mathcal{F}} \le 3(k+j)t. \tag{1.22}$$

PROOF. First we observe that for $f(X)$ and $f(Y)$, $f \in \mathcal{F}$, arbitrary, if x is a t–concentration point for $f(X)$, y is a t–concentration point for $f(Y)$ and z is a t–concentration point for $f(X) + f(Y)$ then

$$\|x + y - z\| \le 3t. \tag{1.22'}$$

To see this just note

$$\begin{aligned}
\Pr\{\|x + y - z\| > 3t\} & \\
&= \Pr\{\|f(X) - x(f) + f(y)Y - y(f) - (f(X) + f(Y) - z(f))\| > 3t\} \\
&\le \Pr^*\{\|f(X) - x(f)\| > t\} + \Pr^*\{\|f(Y) - y(f)\| > t\} \\
&\qquad\qquad\qquad\qquad + \Pr^*\{\|f(X) + f(Y) - z(f)\| > t\} \\
&\le 3/10,
\end{aligned}$$

so that $\Pr\{\|x+y-z\| \le 3t\} > 0$ and therefore (1.22') holds since x, y and z are non–random. To prove the lemma we now proceed by induction. The lemma obviously holds for $j = k$, and (1.22') gives it for $k = 2$. Hence, it suffices to show that if the lemma holds for $1 \le j < r$ for all $r < k$, then it also holds for $1 \le j < k$. Now,

$$js_k - ks_j = js_k - (k-j)s_j - js_j = (js_{k-j} - (k-j)s_j) + j(s_k - s_j - s_{k-j}).$$

Hence, applying (1.22') and the induction hypothesis, we obtain

$$\begin{aligned}
\|js_k - ks_j\| &\le \|js_{k-j} - (k-j)s_j\| + j\|s_k - s_{k-j} - s_j\| \\
&\le 3(k - j + j)t + 3jt = 3(k+j)t.
\end{aligned}$$

□

PROOF OF THEOREM 1.10. We distinguish three cases. Suppose first $\Pr\{\|S_j\|^* > t/4\} \le 1/2$. Enlarging the probability space $(S^n, \mathcal{S}^n, P^n)$ to a product with itself and letting X'_j be the second set of coordinates and $S'_j(f) = \sum_{k=1}^{j} f(X'_k)$, we have

$$\begin{aligned}
\Pr\{\|S_j\|^* > t\} &= \Pr\{\|S_k - S_{k-j}\|^* > t\} \\
&\le \Pr\{\|S_k\|^* > \frac{t}{2}\} + \Pr\{\|S_{k-j}\|^* > \frac{t}{2}\} \\
&\le \Pr\{\|S_k\|^* > \frac{t}{2}\} + 2\Pr\{\|S_{k-j}\|^* > \frac{t}{2},\ \|S'_j\|^* \le \frac{t}{4}\} \\
&\le \Pr\{\|S_k\|^* > \frac{t}{2}\} + 2\Pr\{\|S_{k-j} + S'_j\|^* > \frac{t}{4}\} \\
&\le 3\Pr\{\|S_k\|^* > \frac{t}{4}\},
\end{aligned}$$

where we used independence of $\|S_{k-j}\|^*$ and $\|\tilde{S}_j\|^*$, as in the proof of Lemma 1.9.

Next we asume that there exists some $1 \le i \le k$ such that S_i does not have any $(t/64)$-concentration points. Denote Pr_i the product of the first i P's. Then

$$\mathrm{Pr}_i^*\big\{\|S_i(f) + f(X_{i+1}) + \ldots + f(X_k)\| > \frac{t}{64}\big\} \ge \frac{1}{10}$$

for all values of $X_{i+1}, \ldots, X_k$. Then, applying (1.12) for the random elements S_i and $S_k - S_i$, we obtain

$$\mathrm{Pr}^*\big\{\|S_k\| > \frac{t}{64}\big\} \ge \frac{1}{10} \ge \frac{1}{10}\mathrm{Pr}\{\|S_j\| > t\}$$

for all $1 \le j \le k$.

Suppose, finally, that $\mathrm{Pr}\big\{\|S_j\|^* > t/4\big\} \ge 1/2$ and that S_i has a $(t/64)$-concentration point s_i for all $1 \le i \le k$. Then $\{\|S_j\|^* > t/4\} \cap \{\|S_j - s_j\|^* \le t/64\} \ne \emptyset$ and therefore, $\|s_j\| \ge 15t/64$. Hence, by Lemma 1.11,

$$\|s_k\| \ge \frac{k}{j}\|s_j\| - 3\frac{k+j}{j}\frac{t}{64} \ge \frac{15kt}{64j} - \frac{6kt}{64j} \ge \frac{9t}{64}.$$

This gives

$$\mathrm{Pr}\big\{\|S_k\|^* \ge \frac{8t}{64}\big\} \ge \mathrm{Pr}\big\{\|S_k - s_k\|^* \le \frac{t}{64}\big\} \ge \frac{9}{10} \ge \frac{9}{10}\mathrm{Pr}\{\|S_j\|^* > t\}. \qquad \square$$

Theorem 1.10, combined with Lemma 1.9, yields *Lévy's maximal inequality for i.i.d. random vectors*:

1.12. COROLLARY. *For $n \in \mathbb{N}$, let X_i, $i \le n$, be the coordinate functions on the product probability space $(S^n, \mathcal{S}^n, P^n)$, and let $\mathcal{F}$ be a class of measurable functions. Then, for $1 \le k \le n$,*

$$\mathrm{Pr}\Big\{\max_{k\le n}\Big\|\sum_{i=1}^{k} f(X_i)\Big\|_{\mathcal{F}}^* > t\Big\} \le 30\mathrm{Pr}^*\Big\{\Big\|\sum_{i=1}^{n} f(X_i)\Big\| > \frac{t}{192}\Big\}. \tag{1.23}$$

Finally, we turn to Hoffmann–Jørgensen's inequality. We state it here without proof. References: Hoffmann-Jørgensen (1974); see also Araujo and Giné (1980) and, particularly, Ledoux and Talagrand (1991), in the measurable case; the proof for measurable envelopes in the non measurable case is analogous (just as the proof of Lemma 1.9 above is 'the same' as the proof of the same lemma in the measurable case).

1.13. THEOREM. *Let $0 < p < \infty$, let $n \in \mathbb{N}$, and let X_i be the coordinate functions in the product probability space $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, \prod_{i=1}^{\infty} \mathrm{P}_i)$, and let $\mathcal{F}$ be a class of measurable functions such that $\mathbb{E}\|f(X_i\|_{\mathcal{F}}^p < \infty$, $i = 1, \ldots, n$. Letting*

$$t_0 = \inf\Big[t > 0 : \mathrm{Pr}\big\{\max_{k\le n}\Big\|\sum_{i=1}^{k} f(X_i)\Big\|_{\mathcal{F}}^* > t\Big\} \le 1/(2\cdot 4^p)\Big],$$

we have,

$$\mathbb{E}\max_{k\le n}\Big(\big\|\sum_{i=1}^{k} f(X_i)\big\|_{\mathcal{F}}^*\Big)^p \le 2\cdot 4^p \mathbb{E}\max_{k\le n}\big(\|f(X_i)\|_{\mathcal{F}}^*\big)^p + 2(4t_0)^p. \tag{1.24}$$

This inequality is the best tool available to derive uniform integrability for uniformly tight sequences of sums of independent random vectors *when*

$$\Pr\{\max_{k\le n}\big\|\sum_{i=1}^{k} f(X_i)\big\|_{\mathcal{F}}^* > t\} \ge c_1 \overset{*}{\Pr}\{\big\|\sum_{i=1}^{n} f(X_i)\big\|_{\mathcal{F}} > c_2 t\}$$

for all $t>0$ *and some fixed* $c_i>0$. By Lévy's inequality this happens if the variables $f(X_i)$ are symmetric, or if $\mathcal{F}=\{f\}$ and $f\ge 0$, and it happens *always* when the X_i's are i.i.d. because of Corollary 1.12.

2.1.3. Donsker classes. As in the last subsection, $(S,\mathcal{S},\mathrm{P})$ is a probability space, $(\Omega,\Sigma,\Pr)=\big(S^{\mathbb{N}}\times\Omega',\mathcal{S}^{\mathbb{N}}\times\Sigma',\mathrm{P}^{\mathbb{N}}\times\Pr'\big)$, X_i, $i\in\mathbb{N}$, are the coordinate functions, which in particular are i.i.d.(P), and $\mathcal{F}$ is a class of measurable functions. From this point on we assume that $\mathcal{F}$ is a subset of $\mathcal{L}_2(\mathrm{P})$. The *centered empirical measure* corresponding to the data X_i, is defined as

$$\mathrm{P}_n-\mathrm{P} := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} - P \tag{1.25}$$

and it is a random element with values in the space $\ell^\infty(\mathcal{F})$ provided we assume $F_c(s)<\infty$ for all $s\in S$, that is, condition (1.9). The *empirical process* is defined as

$$\nu_n^{\mathrm{P}} := \nu_n := \sqrt{n}(\mathrm{P}_n-\mathrm{P}), \tag{1.26}$$

and it also takes its values in $\ell^\infty(\mathcal{F})$ if condition (1.9) holds. The P–Brownian bridge G_P is the centered Gaussian process on $\mathcal{L}_2(\mathrm{P})$ with covariance

$$\mathrm{Cov}\big(G_{\mathrm{P}}(f),G_{\mathrm{P}}(g)\big)=\mathrm{Cov}_P\big(f,g\big).$$

We say that the class of functions $\mathcal{F}$ is P–*pregaussian* if the restriction of the process G_{P} to $\mathcal{F}$ induces a tight Borel probability law on $\ell^\infty(\mathcal{F})$. By Remark 1.2, if $\mathcal{F}$ is P–pregaussian then $\{G_P(f): f\in\mathcal{F}\}$ admits a version with almost all its sample paths bounded and uniformly continuous for its intrinsic pseudometric

$$\rho_{\mathrm{P}}^2(f,g)=\mathbb{E}\big(G_{\mathrm{P}}(f)-G_{\mathrm{P}}(g)\big)^2=\mathrm{P}(f-g)^2-\big(\mathrm{P}(f-g)\big)^2, \tag{1.27}$$

and $(\mathcal{F},\rho_{\mathrm{P}})$ is totally bounded. By Sudakov's minorization (e.g., Ledoux and Talalgrand, 1991, pages 79–84), if the process $\{G_P(f): f\in\mathcal{F}\}$ admits such a version then automatically $(\mathcal{F},\rho_{\mathrm{P}})$ is totally bounded and therefore, $\mathcal{F}$ is P–pregaussian by Lemma 1.1 (although we will not use this fact in what follows). When $\mathcal{F}$ is P–pregaussian, we will keep the same notation G_{P} to denote its version(s) with sample paths in $C_u(\mathcal{F},\rho_{\mathrm{P}})$. We now have the ingredients for the following definition of Dudley (1978, 1985):

1.14. DEFINITION. *We say that a class $\mathcal{F}$ satisfying condition (1.9) is* P*–Donsker, or $\mathcal{F} \in CLT(\mathrm{P})$, if $\mathcal{F}$ is* P*–pregaussian and*

$$\nu_n^{\mathrm{P}} \to_{\mathcal{L}} G_{\mathrm{P}} \quad \text{in } \ell^\infty(\mathcal{F}).$$

Theorem 1.3 then has the following corollary:

1.15. THEOREM. *a) A class $\mathcal{F}$ satisfying (1.9) is P–Donsker if and only if there exists a pseudo–distance d on $\mathcal{F}$ such that $(\mathcal{F}, d)$ is totally bounded and*

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \Pr^*\Big\{\sup_{f,g\in\mathcal{F},d(f,g)\le\delta} \big|\nu_n(f-g)\big| > \varepsilon\Big\} = 0 \tag{1.28}$$

for all $\varepsilon > 0$, and then G_{P} has a version with all its sample paths in $C_u(\mathcal{F}, d)$.
b) A class $\mathcal{F}$ satisfying (1.9) is P–Donsker if and only if $(\mathcal{F}, \rho_{\mathrm{P}})$ is totally bounded and the asymptotic equicontinuity condition (1.28) holds with d replaced by ρ_{P}.

This theorem is due to Dudley (1978, 1984) for ρ_{P} and to Andersen and Dobrić (1987) for general d. In a sense, it goes back to Prohorov's tightness criterion in $C(S)$.

In $\mathbb{R}$, if X satisfies the clt then $\mathbb{E}X^2 < \infty$, and in infinite dimensions, if X satisfies the clt then $t^2 \Pr\{\|X\| > t\} \to 0$ as $t \to \infty$. The analogue statement is true in $\ell^\infty(\mathcal{F})$:

1.16. THEOREM. *If $\mathcal{F}$ is P-Donsker (this presuposes that $\mathcal{F}$ satisfies (1.9)) and F_c is its centered envelope then*

$$\lim_{t\to\infty} t^2 \Pr^*\big\{F_c > t\big\} = 0. \tag{1.29}$$

PROOF. Postponed to Proposition 1.20 below.

□

We will require several other conditions equivalent to $\mathcal{F}$ being P–Donsker, besides the asymptotic equicontinuity condition of Theorem 1.15, and randomization and symmetrization, as developed in the previous section, are invaluable tools for this. The following theorem was obtained by Giné and Zinn (1984; see also 1986). To simplify notation, here and in what follows, for every $\delta > 0$, we write

$$\mathcal{F}'_{\delta,\rho_P} := \mathcal{F}_\delta := \big\{f - g : f, g \in \mathcal{F}, \rho_P(f,g) \le \delta\big\}$$

and adhere to the notation $\|\cdot\|_{\mathcal{F}'_\delta}$ for the sup norm over the class $\mathcal{F}'_\delta$.

1.17. THEOREM. *Let $\mathcal{F}$ be a class satisfying condition (1.9). Then the following are equivalent:*

a) $\mathcal{F}$ is P*–Donsker;*

b) $(\mathcal{F}, \rho_{\mathrm{P}})$ is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \Pr^*\Big\{\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i(f(X_i) - \mathrm{P}f)\Big\|_{\mathcal{F}'_\delta} > \varepsilon\Big\} = 0$$

for all $\varepsilon > 0$;

b') $\mathcal{F}$ is P–pregaussian and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i(\delta_{X_i}-\mathrm{P})\to_{\mathcal{L}} G_{\mathrm{P}} \quad \text{in } \ell^\infty(\mathcal{F});$$

c) $(\mathcal{F},\rho_{\mathrm{P}})$ is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i(f(X_i)-\mathrm{P}f)\Big\|_{\mathcal{F}'_\delta}=0;$$

d) $(\mathcal{F},\rho_{\mathrm{P}})$ is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}^*\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'_\delta}=0.$$

PROOF. The symmetrization inequalities (Lemma 1.7), together with Theorem 1.15 show that *a*) implies *b*); Hoffmann-Jørgensen (together with Lévy) and Theorem 1.16, that *b*) implies *c*); *c*) implies *d*) by inequality (1.14), and *d*) implies a) by Theorem 1.15 and Chebyshev. *b*) and *b'*) are equivalent by Theorem 1.3. See Giné and Zinn (1986), Theorem 2.8, for details.

□

One of the reasons for symmetrizing in the previous theorem is that one can use Lévy's inequality for more than one purpose, but in particular to obtain an efficient version of Hoffmann-Jørgensen's inequality. However, in the case of i.i.d. summands, which is the present situation, as mentioned above, Corollary 1.12 (Lévy for i.i.d.) allows us to pass from *a*) to *d*) in Theorem 1.17 without symmetrizing. Although Corollary 1.12 makes randomization/symmetrization less necessary than it used to be, it is not clear that it has become completely superfluous, as we will see along the way.

It will be convenient to further extend Theorem 1.17 since multipliers other than Rademacher will be needed in conditions *b*), *b'*) and *c*) in connection with the bootstrap. The following inequality will do just this. It requires the following definition: for real random variables ξ, we let

$$\Lambda_{2,1}(\xi):=\int_0^\infty\sqrt{\Pr\{|\xi|>t\}}\,dt. \tag{1.30}$$

Note that $\Lambda_{2,1}(\xi)<\infty$ implies $\mathbb{E}\xi^2<\infty$, and that $\mathbb{E}|\xi|^{2+\delta}<\infty$ for some $\delta>0$ implies $\Lambda_{2,1}(\xi)<\infty$.

1.18. THEOREM. *Let $\mathcal{F}$ be a class of P–integrable functions, and let ε_i, ξ_i, $i\in\mathbb{N}$, be respectively a Rademacher sequence and a sequence of symmetric i.i.d. real random variables, independent of each other and defined on the Ω' factor of the basic probability space hence, in particular, independent of the sequence $\{X_i\}$. Then, for every $0\le n_0<\infty$ and $n_0<n\in\mathbb{N}$, we have*

$$(\mathbb{E}|\xi_1|)\mathbb{E}^*\Big\|\frac{\sum_{i=1}^n\varepsilon_i f(X_i)}{\sqrt{n}}\Big\|_{\mathcal{F}}\le\mathbb{E}^*\Big\|\frac{\sum_{i=1}^n\xi_i f(X_i)}{\sqrt{n}}\Big\|_{\mathcal{F}}$$

$$\leq n_0 \big(\mathbb{E}^* \|f(X_1)\|_{\mathcal{F}}\big) \mathbb{E}\left[\max_{i\leq n} \frac{|\xi_i|}{\sqrt{n}}\right] + \Lambda_{2,1}(\xi_1) \max_{n_0<k\leq n} \mathbb{E}^* \left\| \frac{\sum_{i=n_0+1}^k \varepsilon_i f(X_i)}{\sqrt{k}} \right\|_{\mathcal{F}}. \tag{1.31}$$

If the variables ξ_i are centered but not necessarily symmetric, inequality (1.31) holds with the following modifications: $\mathbb{E}|\xi_1|$ at the left is replaced by $\mathbb{E}|\xi_1 - \xi_2|/2$, the first summand at the right is multiplied by 2 and the second by $2\sqrt{2}$.

PROOF. The left side inequality in (1.31) follows from the observation that, by symmetry, the joint distribution of the variables ξ_i coincides with the joint distribution of the variables $\varepsilon_i|\xi_i|$, so that

$$\mathbb{E}^* \left\| \frac{\sum_{i=1}^n \xi_i f(X_i)}{\sqrt{n}} \right\|_{\mathcal{F}} = \mathbb{E}^* \left\| \frac{\sum_{i=1}^n \varepsilon_i |\xi_i| f(X_i)}{\sqrt{n}} \right\|_{\mathcal{F}} \geq \mathbb{E}^* \left\| \frac{\sum_{i=1}^n \varepsilon_i (\mathbb{E}|\xi_i|) f(X_i)}{\sqrt{n}} \right\|_{\mathcal{F}}.$$

The following chain of inequalities, which are self-explanatory, gives the proof of the main part of theorem (the subindex $\mathcal{F}$ is suppressed from the norm signs):

$$\begin{aligned}
\mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i f(X_i) \right\| &= \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i |\xi_i| f(X_i) \right\| \\
&= \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\int_0^\infty I_{t\leq|\xi_i|} dt \Big) \varepsilon_i f(X_i) \right\| \\
&= \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \int_0^\infty \Big(\sum_{i=1}^n I_{t\leq|\xi_i|} \varepsilon_i f(X_i) \Big) dt \right\| \\
&\leq \int_0^\infty \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{t\leq|\xi_i|} \varepsilon_i f(X_i) \right\| dt \\
&= \int_0^\infty \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{\#\{i\leq n:\ |\xi_i|\geq t\}} \varepsilon_i f(X_i) \right\| dt \\
&\leq \int_0^\infty \left(\sum_{k=1}^n \Pr\Big\{ \sum_{i=1}^n I_{|\xi_i|\geq t} = k \Big\} \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^k \varepsilon_i f(X_i) \right\| \right) dt
\end{aligned}$$

$$\leq \left(\int_0^\infty \Pr\Big\{ \sum_{i=1}^n I_{|\xi_i|\geq t} > 0 \Big\} dt \right) \max_{k\leq n_0} \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^k \varepsilon_i f(X_i) \right\|$$

$$+ \left(\frac{1}{\sqrt{n}} \int_0^\infty \sum_{k=n_0+1}^\infty \sqrt{k} \Pr\Big\{ \sum_{i=1}^n I_{|\xi_i|\geq t} = k \Big\} dt \right) \max_{n_0<k\leq n} \mathbb{E}^* \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0+1}^k \varepsilon_i f(X_i) \right\|$$

$$\leq \Big(\int_0^\infty \Pr\{\max_{i\leq n}|\xi_i|\geq t\}dt\Big)\frac{n_0}{\sqrt{n}}\mathbb{E}^*\|f(X_i)\|$$
$$+\Lambda_{2,1}(\xi)\max_{n_0<k\leq n}\mathbb{E}^*\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^{k}\varepsilon_i f(X_i)\Big\|$$
$$= n_0\mathbb{E}^*\|f(X_1)\|\mathbb{E}\Big(\max_{k\leq n}\frac{|\xi_i|}{\sqrt{n}}\Big)+\Lambda_{2,1}(\xi)\max_{n_0<k\leq n}\mathbb{E}^*\Big\|\frac{1}{\sqrt{k}}\sum_{n_0<i\leq k}\varepsilon_i f(X_i)\Big\|.$$

When ξ is not symmetric, but still centered, the theorem follows from the previous estimates applied to $\xi_i-\xi_i'$, where $\{\xi_i'\}$ is an independent copy of $\{\xi_i\}$, and inequality (1.12).

□

Inequalities (1.31) are due to Pisier (private communication) and possibly also, independently, to Fernique (both of them told it to me in 1977 and none of the three of us has exact recollections); they were published, with a different proof, in Giné and Zinn (1984, 1986), and with the present proof in Giné (1996). Ledoux and Talagrand (1985) prove that the integrability condition on ξ cannot in general be relaxed.

The previous inequalities obviously imply the following extension of Theorem 1.17 (note that if $\Lambda_{2,1}(\xi)<\infty$ we have $\mathbb{E}\max_{i\leq n}|\xi_i|/\sqrt{n}\to 0$ and we can take $n_0\to\infty$ in (1.31)).

1.19. THEOREM. *Let $\mathcal{F}$ be a class satifying condition (1.9) and let ξ_i be i.i.d. centered real random variables defined on Ω' and such that $\Lambda_{2,1}(\xi_1)<\infty$. Then the following are equivalent:*

a) $\mathcal{F}$ is P*-Donsker;*

b) $(\mathcal{F},\rho_{\mathrm{P}})$ is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr{}^*\Big\{\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(f(X_i)-\mathrm{P}f)\Big\|_{\mathcal{F}'_\delta}>\varepsilon\Big\}=0$$

for all $\varepsilon>0$;

c) $(\mathcal{F},\rho_{\mathrm{P}})$ is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(f(X_i)-\mathrm{P}f)\Big\|_{\mathcal{F}'_\delta}=0.$$

We complete this subsection with a useful remark on the necessary integrability associated with the Donsker property. We need a statement slightly more general than Theorem 1.16 in order to include the bootstrap clt. We will use implicitly in the proof below the fact that $I_{\|X\|_T^*>t}=\big(I_{\|X_T\|>t}\big)^*$ already used in the previous subsection (e.g. Andersen, 1985, I, Proposition 2.3 or van der Vaart and Wellner, 1996, Lemma 1.2.2).

1.20. THEOREM. *Let (T,d) be a separable pseudo-metric space and, for each $n \in \mathbb{N}$, let $X_{n,1}(t),\dots,X_{n,n}(t)$, $t \in T$, be i.i.d. processes with bounded sample paths, defined on different factors of a product probability space. Assume*
i) for all t in a d-dense subset D of T and for all $\lambda > 0$,

$$\lim_{n\to\infty} n\Pr^*\{|X_{n,1}(t)| > \lambda\sqrt{n}\} = 0, \tag{1.32}$$

and ii)

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \Pr^*\Big\{\frac{1}{\sqrt{n}}\sup_{d(s,t)\le\delta}\Big|\sum_{i=1}^{n}(X_{n,i}(t)-\mathbb{E}X_{n,i}(t)-X_{n,i}(s)+\mathbb{E}X_{n,i}(s))\Big| > \varepsilon\Big\} = 0 \tag{1.33}$$

for all $\varepsilon > 0$.
Then,

$$\lim_{n\to\infty} n\Pr^*\Big\{\big\|X_{n,1}\big\|_T > \lambda\sqrt{n}\Big\} = 0 \tag{1.34}$$

for all $\lambda > 0$. If hypotheses 1) and 2) hold only along a subsequence n_k, then the same is true for the conclusion.

PROOF. Let $\{X'_{n,i}\}$ be an independent copy of $\{X_{n,i}\}$, defined on a different factor of the general product probability space. Then, since

$$\Pr^*\Big\{\Big\|\sum_i(X_{n,i}-X'_{n,i})\Big\| > t\Big\} \le 2\Pr^*\Big\{\Big\|\sum_i X_{n,i}\Big\| > \frac{t}{2}\Big\} \tag{1.35}$$

for any pseudo-norm $\|\cdot\|$ and any $t > 0$, condition ii) implies

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \Pr^*\Big\{\frac{1}{\sqrt{n}}\sup_{d(s,t)\le\delta}\Big|\sum_{i=1}^{n}(X_{n,i}(t)-X'_{n,i}(t)-X_{n,i}(s)+X'_{n,i}(s))\Big| > \varepsilon\Big\} = 0 \tag{1.36}$$

Set $Y_{n,i} := X_{n,i}-X'_{n,i}$. For each $\tau > 0$ let $A_{1,\tau},\dots,A_{N(\tau),\tau}$ be a finite number $N(\tau)$ of subsets of T whose union covers T, whose diameter is smaller than T and such that $A_{i,\tau}\cap D \ne \emptyset$, $i = 1,\dots,N(\tau)$. Choose $t_i \in A_{i,\tau}\cap D$ and define $Y_{n,j,\tau}(t) = Y_{n,j}(t_i)$ if $t \in A_{i,\tau}$, $j = 1,\dots,N(\tau)$ (very much as in the proof of Theorem 1.3 as modified by Remark 1.4). Then,

$$\Pr^*\Big\{\big\|Y_{n,1}\big\|_T > 2\lambda\sqrt{n}\Big\} \le \Pr^*\Big\{\big\|Y_{n,1,\tau}\big\|_T > \lambda\sqrt{n}\Big\}+\Pr^*\Big\{\big\|Y_{n,1}-Y_{n,1,\tau}\big\|_T > \lambda\sqrt{n}\Big\} \tag{1.37}$$

Now,

$$n\Pr^*\Big\{\big\|Y_{n,1,\tau}\big\|_T > \lambda\sqrt{n}\Big\} \le \sum_{i=1}^{N(\tau)} n\Pr\Big\{\big|Y_{n,1}(t_i)\big| > \lambda\sqrt{n}\Big\} \to 0 \tag{1.38}$$

by hypothesis i) and (1.35) (with $X_{n,i}$ replaced by $X_{n,i}(t)-\mathbb{E}X_{n,i}(t)$, and likewise for $X'_{n,i}$). Moreover, by Lévy's inequality (1.17) (which works for outer probabilities because $\|\cdot\|^*$ behaves a.s. as a measurable norm),

$$\begin{aligned}1-\exp\Big[-n\Pr{}^*\Big\{\big\|Y_{n,1}-Y_{n,1,\tau}\big\|_T>\lambda\sqrt{n}\Big\}\Big]\\ &\le \Pr{}^*\Big\{\max_{1\le i\le n}\big\|Y_{n,i}-Y_{n,i,\tau}\big\|_T>\lambda\sqrt{n}\Big\}\\ &\le 2\Pr{}^*\Big\{\big\|\sum_{i=1}^{n}(Y_{n,i}-Y_{n,i,\tau})\big\|_T>\lambda\sqrt{n}\Big\}\\ &\le 2\Pr{}^*\Big\{\frac{1}{\sqrt{n}}\sup_{d(s,t)\le\tau}\Big|\sum_{i=1}^{n}(X_{n,i}(t)-X'_{n,i}(t)-X_{n,i}(s)+X'_{n,i}(s))\Big|>\lambda\Big\}.\end{aligned}$$

Hence, (1.36) gives

$$\lim_{\tau\to0}\limsup_{n\to\infty} n\Pr{}^*\Big\{\big\|Y_{n,1}-Y_{n,1,\tau}\big\|_T>\lambda\sqrt{n}\Big\}=0,$$

and this, together with (1.37) and (1.38), yields

$$\lim_{n\to\infty} n\Pr{}^*\Big\{\big\|Y_{n,1}\big\|_T>\lambda\sqrt{n}\Big\}=0. \tag{1.39}$$

To desymmetrize, we observe that, by independence of $\|X_{n,i}\|_T^*$ and $\|X'_{n,i}\|_T^*$ (which follows e.g. from Dudley and Philipp, 1983, Lemma 2.3, already used in the previous subsection),

$$n\Pr{}^*\Big\{\big\|Y_{n,1}\big\|_T>\lambda\sqrt{n}\Big\}\ge n\Pr\Big\{\big\|X_{n,1}\big\|_T^*>2\lambda\sqrt{n}\Big\}\Pr\Big\{\big\|X_{n,1}\big\|_T^*\le\lambda\sqrt{n}\Big\}$$

and that by hypothesis i), for any $t\in D$,

$$\Pr\Big\{\big\|X_{n,1}\big\|_T^*\le\lambda\sqrt{n}\Big\}\le\Pr\Big\{\big|X_{n,1}(t)\big|\le\lambda\sqrt{n}\Big\}\to1$$

as $n\to\infty$, so that (1.39) yields the limit (1.34). The proof for subsequences does not differ from the proof for the whole sequence $\mathbb{N}$.

□

de Acosta, Araujo and Giné (1978, Corollary 2.11) proved a more general version of the above proposition in separable Banach spaces. Pisier and Zinn (1978, Proposition 5.2), independently, obtained it for the usual normal domain of attraction case, independently. The proof presented here is inspired by theirs, which consists basically of reducing to finite dimensions, where the result is classical. On the other hand, de Acosta et al. obtained the result from their study of the general clt, Poisson convergence included, which makes their proof less simple.

2.2. Poissonization inequalities and Efron's bootstrap in probability. Take the factor $(\Omega',\sigma',\Pr')$ of the basic probability space $(\Omega,\sigma,\Pr)$ to be a countable product of copies of $([0,1],\mathcal{B},\lambda)$, λ being Lebesgue measure (we might later on increase Ω' in order to fit it to any new set of variables we may define), and let U_i, defined on $(\Omega', \sigma', \Pr')$ denote the $[0,1]$–coordinates, $U_i(s,s') = s'_i$, where $s' \in [0,1]^{\mathbb{N}}$. For each $n \in \mathbb{N}$ define variables

$$X^b_{n,i} = \sum_{j=1}^{n} X_j I_{U_i \in A(j,n)}, \tag{2.1}$$

with $A(j,n) = \big((j-1)/n, j/n\big]$, as in Chapter 1 (note that this makes sense because only one of the summands in not zero). Recall that in this chapter the bootstrap variables are denoted by a superindex b instead of the usual superindex * and that now $\mathbb{E}^b$, $\Pr^b$ and $\mathcal{L}^b$ denote conditional expectation probability and law given the sample, or what is the same, expectation, probability and law only with repect to the ω' coordinate of the basic probability space, with ω fixed. For each ω, the variables $X^b_{n,i}(\omega)$, $i = 1, \ldots, n$, are i.i.d.$(\mathrm{P}_n(\omega))$ and the bootstrap empirical measure $\mathrm{P}^b_n(\omega)$ is the measure that places mass $1/n$ on each of the points $X^b_{n,i}(\omega)$ of the bootstrap sample. So, for each n and ω, the (n–th) bootstrap empirical process is defined as

$$\nu^b_n(\omega) := \nu^{\mathrm{P}_n}_n(\omega) := \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} \delta_{X^b_{n,i}(\omega)} - \mathrm{P}_n(\omega)\Big). \tag{2.2}$$

Whereas the empirical process ν_n does not have a Radon law in $\ell^\infty(\mathcal{F})$, the conditional law of $\nu^b_n(\omega)$ is Radon for each ω as it only takes a finite number of values.

2.1. DEFINITION. *We say that a class of measurable functions $\mathcal{F} \subset \mathcal{L}_2(\mathrm{P})$ satisfying condition (1.9) is bootstrap* P*–Donsker in probability, or $\mathcal{F} \in B_{pr}CLT(\mathrm{P})$, if $\mathcal{F}$ is P–pregaussian and*

$$\lim_{n\to\infty} d_{BL}\big[\mathcal{L}^b\big(\nu^{\mathrm{P}_n}_n(\omega)\big), \mathcal{L}(G_{\mathrm{P}})\big] = 0 \text{ in outer probability.} \tag{2.3}$$

Our object in this section is to prove the following theorem of Giné and Zinn (1990) (with an improvement on measurability by Strobl, 1994, and van der Vaart and Wellner, 1996). Recall that condition (1.9) is part of the definition of P–Donsker classes. Define the envelope F of $\mathcal{F}$ as $F(s) := \sup\{|f(s)| : f \in \mathcal{F}\}$.

2.2. THEOREM. *If $\mathcal{F}$ is P–Donsker then $\mathcal{F}$ is also bootstrap* P*–Donsker in probability. Conversely, if a class of measurable functions $\mathcal{F}$ with everywhere finite envelope F is image admissible Suslin and there exists a centered Gaussian process G indexed by $\mathcal{F}$ whose law is Radon in $\ell^\infty(\mathcal{F})$ and such that*

$$\lim_{n\to\infty} d_{BL}\big[\mathcal{L}^b\big(\nu^{\mathrm{P}_n}_n(\omega)\big), \mathcal{L}(G)\big] = 0 \text{ in outer probability,} \tag{2.4}$$

then $\mathcal{F}$ is P–Donsker and $G = G_{\mathrm{P}}$.

See Dudley (1984, Section 10.3) for the definition of the image admissible Suslin property and its consequences. For instance, if $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ where Θ is a Polish space, and if the evaluation map $(\theta, s) \mapsto f_\theta(s)$ is jointly measurable, then $\mathcal{F}$ is image admissible Suslin (previously it was required that $(S, \mathcal{S})$ be also Suslin, but this requirement is not necessary: see a forthcoming book of Dudley on empirical processes). The image admissible property implies that $\|P_n - P\|_{\mathcal{F}}$ and its randomized versions are measurable, so that in this case there is no need of non-measurable calculus. Most classes of interest are image admissible Suslin. The converse part of Theorem 2.2 holds under weaker measurability conditions (see e.g. van der Vaart and Wellner, 1996), but the Suslin property is very convenient and, this part being less useful than the direct part (basically, it just says that the direct part cannot be improved), we have chosen to give it under this stronger measurability hypothesis.

Zinn and I (1990) stated Theorem 2.2 under measurability assumptions for both, direct and converse, but our original proof actually gives the direct part of the theorem without any measurability, as noted in the dissertation of F. Strobl (1994) and in the recent book of van der Vaart and Wellner (1996). We present here our proof of the direct part and a new proof of the converse, different, but not too different, from our original proof. This proof of the converse, based on a result on Poissonization of sums of i.i.d. symmetric (or centered) random variables of Araujo and Giné (1980) is inspired in part on work by Klaassen and Wellner (1992). Another difference with our original proof, also inspired by Klaassen and Wellner, is that here we make less extensive use of symmetry.

Besides the general theory sketched in Section 1, plus the bootstrap of the mean in $\mathbb{R}$, this theorem is based on two simple inequalities relating the original and the bootstrap empirical processes. These inequalities are based on Poissonization. So, we start with some basics about this (see Araujo and Giné, 1980, for a more extensive treatment of the subject in connection with the clt).

LeCam (1970) proved that in a general Banach space, if the accompanying Poisson laws of a triangular array of (row-wise) independent symmetric random variables are tight, so are the row sums. We need something slightly different, namely comparison of the expected values of the norms, a result that has a different, simpler proof than LeCam's (Giné and Zinn, 1990). We gave this inequality only for symmetric random variables, which is what we needed but both, statement and proof, work without changes for centered variables and without any measurability assumptions, as first noted by Klaassen and Wellner, loc. cit.

2.3. LEMMA. (First Poissonization inequality). *For any $n \in \mathbb{N}$, let X_i, $i = 1, \ldots, n$, $j \in \mathbb{N}$, be centered independent stochastic processes on an index set T, and, for each $i \le n$, let $X_{i,j}$, $j \in \mathbb{N}$, be independent copies of X_i, all independent. Let $X_{i,0} \equiv 0$. Let N_i be i.i.d. Poisson random variables with unit expectation, independent of the $X_{i,j}$. Assume, in fact, that each of the above variables (the X's and the N's) is defined on a different factor of an infinite product probability space. Then,*

$$\mathbb{E}^*\Big\|\sum_{i=1}^n X_i\Big\|_T \le \frac{e}{e-1}\mathbb{E}^*\Big\|\sum_{i=1}^n\sum_{j=0}^{N_i} X_{i,j}\Big\|_T. \tag{2.5}$$

PROOF. Using inequalities (1.16) and (1.17), we have

$$\begin{aligned}\frac{e-1}{e}\mathbb{E}^*\Big\|\sum_{i=1}^n X_{i,1}\Big\|_T &= \mathbb{E}^*\Big\|\sum_{i=1}^n(\mathbb{E}(N_i\wedge 1))X_{i,1}\Big\|_T \le \mathbb{E}_X^*\mathbb{E}_N\Big\|\sum_{i=1}^n(N_i\wedge 1)X_{i,1}\Big\|_T\\ &\le \mathbb{E}_N\mathbb{E}_X^*\Big\|\sum_{i=1}^n(N_i\wedge 1)X_{i,1}\Big\|_T \le \mathbb{E}_N\mathbb{E}_X^*\Big\|\sum_{i=1}^n\sum_{j=0}^{N_i}X_{i,j}\Big\|_T = \mathbb{E}^*\Big\|\sum_{i=1}^n\sum_{j=0}^{N_i}X_{i,j}\Big\|_T.\end{aligned}$$

□

A reverse inequality is true provided the X_i are i.i.d. This observation comes from Araujo and Giné (1980, Theorem 4.9, Chapter 3, page 122, where the word symmetric is missing, and exercise 2 on the same page), who used it to prove that Le Cam's theorem for accompanying Poisson laws has a converse if the variables in each row of a triangular array are i.i.d. and symmetric. The same principle, basically a natural coupling of the sums and the Poison variables, has surely been used before. Here is an adaptation of the result for expected values (instead of probabilities).

2.4. LEMMA. (Second Poissonization inequality). *Let X_i, $i \in \mathbb{N}$, $j \in \mathbb{N}$, be centered independent identically distributed stochastic processes on an index set T. Let $X_0 \equiv 0$. Let $n \in \mathbb{N}$ and let $N(n)$ be a Poisson random variable with parameter n independent of the X's. Assume, in fact, that each of the above variables (the X's and N) is defined on a different factor of an infinite product probability space. Then,*

$$\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i - \sum_{i=0}^{N(n)} X_i\Big\|_T \leq 2\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i\Big\|_T, \tag{2.6}$$

and, in particular,

$$\mathbb{E}^*\Big\|\sum_{i=0}^{N(n)} X_i\Big\|_T \leq 3\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i\Big\|_T. \tag{2.7}$$

PROOF. The key estimate is the following

$$\begin{aligned}
\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i - \sum_{i=0}^{N(n)} X_i\Big\|_T &= \sum_{r=-n}^{\infty} \Pr\{N_n - n = r\}\mathbb{E}^*\Big\|\sum_{i=0}^{|r|} X_i\Big\|_T \\
&\leq \sum_{r=-n}^{n} \Pr\{N_n - n = r\}\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i\Big\|_T + 2\sum_{r=n+1}^{2n} \Pr\{N_n - n = r\}\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i\Big\|_T \\
&\quad + \ldots + k\sum_{r=(k-1)n+1}^{kn} \Pr\{N_n - n = r\}\mathbb{E}^*\Big\|\sum_{i=1}^{n} X_i\Big\|_T + \ldots \\
&= \Big[\Pr\{N_n \leq n\} + \sum_{k=1}^{\infty} k\Pr\{kn < N_n \leq (k+1)n\}\Big]\mathbb{E}\Big\|\sum_{i=1}^{n} X_i\Big\| \\
&\leq 2\mathbb{E}\Big\|\sum_{i=1}^{n} X_i\Big\|,
\end{aligned}$$

where the first identity uses the proof of Lemma 1.6 b), the first inequality follows from (1.12) (as the summands are centered), and the last inequality is obtained by comparison of the sum to $\mathbb{E}(N_n/n) = 1$. This proves inequality (2.6). Now (2.7) is immediate.

□

If the law of $X_{i,1}$ is a Radon probability measure, say μ_i, then the law of the variable $\sum_{j=0}^{N_i} X_{i,j}$ is

$$\mathrm{Pois}\mu_i := e^{-1}\sum_{k=0}^{\infty}\frac{\mu_i^k}{k!} = e^{\mu_i - 1},$$

where powers and exponential are understood in the sense of convolution. Then, by the properties of the exponential function, the law of the random variable $\sum_{i=1}^n \sum_{j=0}^{N_i} X_{i,j}$ is

$$\mathrm{Pois}\mu_1 * \cdots * \mathrm{Pois}\mu_n = e^{\mu_1 - 1} * \cdots * e^{\mu_n - 1} = \exp\Big\{\sum_{i=1}^n \mu_i - n\Big\} := \mathrm{Pois}\big(\sum_{i=1}^n \mu_i\big),$$

where, $\mathrm{Pois}\mu$ is defined, for any finite measure μ, as

$$\mathrm{Pois}\mu = e^{\mu - |\mu|},$$

$|\mu|$ being the total mass of μ. (This corresponds to the compound Poisson law associated to the probability measure $\mu/|\mu|$, with intensity $|\mu|$.). With these definitions, if the processes X_i are just centered random vectors in a Banach space whose laws are Radon and if all the norms involved in (2.5) are measurable, then (2.5) becomes

$$\mathbb{E}\Big\|\sum_{i=1}^n X_i\Big\| \le \frac{e}{e-1}\int \|x\| d\mathrm{Pois}\Big(\sum_{i=1}^n \mathcal{L}(X_i)\Big). \tag{2.5'}$$

Likewise, since $N(n) =_d \sum_{i=1}^n N_i$, N_i i.i.d. Poisson with parameter 1, inequality (2.7) for X_i i.i.d. and centered, becomes

$$\int \|x\| d\mathrm{Pois}\Big(n\mathcal{L}(X_1)\Big) \le 3\mathbb{E}\Big\|\sum_{i=1}^n X_i\Big\|. \tag{2.7'}$$

Two easy to check properties of the operator Pois, that we use below, are that

$$\mathrm{Pois}\big(\sum \mu_i\big) = \mathrm{Pois}\mu_1 * \cdots * \mathrm{Pois}\mu_n$$

and that

$$\mathrm{Pois}(a\delta_x) = \mathcal{L}(N_a x)$$

where N_a is a Poisson random variable with prameter a. The following are the basic inequalities for the proof of Theorem 2.2.

2.5. PROPOSITION. *Let B be a Banach space, and, for any $n \in \mathbb{N}$, let $v_1, \dots, v_n$ be points in B and let $v = \sum_{i=1}^n v_i/n$ be their average. Let $V_1^b, \dots, V_n^b$ be i.i.d. B-valued random variables with law $\Pr\{V_i^b = v_j\} = 1/n$, $j = 1, \dots, n$. Let N_i, $i = 1, \dots, n$ be i.i.d. Poisson random variables with parameter 1. Assume these random vectors and variables are all independent. Then,*

$$\frac{1}{3}\mathbb{E}\Big\|\sum_{i=1}^n (N_i - 1)(v_i - v)\Big\| \le \mathbb{E}\Big\|\sum_{i=1}^n (V_i^b - v)\Big\| \le \frac{e}{e-1}\mathbb{E}\Big\|\sum_{i=1}^n (N_i - 1)(v_i - v)\Big\|. \tag{2.8}$$

PROOF. Since the variables V_i^b and N_i are discrete, all the variables involved in (2.8) are measurable. We then note

$$\sum_{i=1}^{n} \mathcal{L}(V_i^b - v) = n\mathcal{L}(V_1^b - v) = \sum_{j=1}^{n} \delta_{v_j - v}.$$

Therefore,

$$\begin{aligned}
\mathrm{Pois}\Big(\sum_{i=1}^{n} \mathcal{L}(V_i^b - v)\Big) &= \mathrm{Pois}\Big(\delta_{v_1 - v}\Big) * \ldots * \mathrm{Pois}\Big(\delta_{v_n - v}\Big) \\
&= \mathcal{L}\Big(\sum_{i=1}^{n} N_i(v_i - v)\Big) \\
&= \mathcal{L}\Big(\sum_{i=1}^{n} (N_i - 1)(v_i - v)\Big), \qquad (2.9)
\end{aligned}$$

where we just use the properties of Poissonizastion listed above and the fact that $\sum v_i = nv$. Now we observe that, by (2.9), the inequalities (2.8) are nothing but (2.5') and (2.7') for $X_i = V_i^b - v$.

□

2.6. REMARK. Given the previous Poissonization lemmas, the following symmetrization of (2.8) has an equally simple, analogous proof (alternatively, (2.8) together with Lemma 1.6 c), give the following inequalities, but with worse constants): if M_i, M_i' are i.i.d. Poisson with parameter 1/2, independent from $\{V_i^b\}$, and if ε_i are i.i.d. Rademacher variables independent from all the others, then

$$\frac{1}{3}\mathbb{E}\Big\|\sum_{i=1}^{n}(M_i - M_i')v_i\Big\| \le \mathbb{E}\Big\|\sum_{i=1}^{n}\varepsilon_i V_i\Big\| \le \frac{e}{e-1}\mathbb{E}\Big\|\sum_{i=1}^{n}(M_i - M_i')v_i\Big\|. \qquad (2.8s)$$

The inequalities obtained in Proposition 2.5 are quite natural: given that the weights in Efron's bootstrap are multinomial, replacing them by Poisson weights should not change things much asymptotically; what is interesting here is that we have concrete inequalities. Technically, what facilitates the above proof is the fact that $\mathrm{Pois}(\sum \mu_i) = \mathrm{Pois}(\sum \nu_i)$ whenever $\sum \mu_i = \sum \nu_i$.

2.7. PROOF OF THE DIRECT PART OF THEOREM 2.2. Assume $\mathcal{F}$ is P–Donsker. We can also assume, without loss of generality, that $\mathrm{P}f = 0$ because both $\nu_n(f) = \nu_n(f - \mathrm{P}f)$ and $\nu_n^b(f) = \nu_n^b(f - \mathrm{P}f)$. We must prove (2.3). Since $(\mathcal{F}, \rho_{\mathrm{P}})$ is totally bounded (Theorem 1.15 b), for every $\tau > 0$ there is a map $\pi_\tau : \mathcal{F} \to \mathcal{F}$ which takes no more than $N(\tau) < \infty$ values and such that $\rho_{\mathrm{P}}(\pi_\tau f, f) < \tau$, and we define $\nu_{n,\tau}^b(f) := \nu_n^b(\pi_\tau f)$, $G_{\mathrm{P},\tau}(f) := G_{\mathrm{P}}(\pi_\tau f)$, as in the proof of Theorem 1.3 and Corollary 1.5. Then, as in these proofs, we consider the decomposition

$$\begin{aligned}
\big|\mathbb{E}^b H(\nu_n^b) - \mathbb{E}H(G_{\mathrm{P}})\big| \le& \big|\mathbb{E}^b H(\nu_n^b) - \mathbb{E}^b H(\nu_{n,\tau}^b)\big| \\
&+ \big|\mathbb{E}^b H(\nu_{n,\tau}^b) - \mathbb{E}H(G_{\mathrm{P},\tau})\big| + \big|\mathbb{E}H(G_{\mathrm{P},\tau}) - \mathbb{E}H(G_{\mathrm{P}})\big| \\
:=& I_{n,\tau}(H) + II_{n,\tau}(H) + III_\tau(H).
\end{aligned}$$

Now, $\sup_{H\in BL_1(\ell^\infty(\mathcal{F}))} III_\tau(H) \to 0$ as $\tau \to 0$ since, by sample uniform continuity with respect to ρ_{P}, $G_{\mathrm{P},\tau} \to_{\mathcal{L}} G_P$ in $\ell^\infty(\mathcal{F})$, a convergence metrized by d_{BL}. Note that III_τ is not random. The sup over BL_1 of $II_{n,\tau}(H)$, $d_{BL_1}\big(\mathcal{L}^b(\nu_{n,\tau}^b), \mathcal{L}(G_{\mathrm{P},\tau})\big)$, is measurable, and it tends to zero a.s. by the bootstrap clt in $\mathbb{R}^{N(\tau)}$. Finally, since

$$\sup_{H\in BL_1(\ell^\infty(\mathcal{F}))} I_{n,\tau}(H) \le E^b\big\|\nu_n^b\big\|_{\mathcal{F}'_\tau},$$

the limit (2.3) will be proved if we show

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \Pr{}^*\Big\{\mathbb{E}^b\big\|\nu_n^b\big\|_{\mathcal{F}'_\delta} > \varepsilon\Big\} = 0 \tag{2.10}$$

for all $\varepsilon > 0$. (Recall the definition of $\|\cdot\|_{\mathcal{F}'_\delta}$ from the paragraph immediately above Theorem 1.17.) The right side of inequality (2.8), applied with $B = \ell^\infty(\mathcal{F}'_\delta)$ and $v_i = \delta_{X_i(\omega)}$ for each fixed ω, gives, using (1.13),

$$\begin{aligned}
\mathbb{E}^*\mathbb{E}^b\big\|\nu_n^b\big\|_{\mathcal{F}'_\delta} &= \mathbb{E}^*\mathbb{E}^b\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(\delta_{X_{n,i}^b} - \mathrm{P}_n\Big)\Big\|_{\mathcal{F}'_\delta} \\
&\le \frac{e}{e-1}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (N_i - 1)(\delta_{X_i} - \mathrm{P}_n)\Big\|_{\mathcal{F}'_\delta} \\
&\le \frac{e}{e-1}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (N_i - 1)\delta_{X_i}\Big\|_{\mathcal{F}'_\delta} \\
&\quad + \frac{e}{e-1}\mathbb{E}^*\Big(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n (N_i - 1)\Big|\|\mathrm{P}_n\|_{\mathcal{F}'_\delta}\Big).
\end{aligned}$$

Now, $\mathcal{F}$ being P–Donsker, the $\lim_\delta \limsup_{n\to\infty}$ of the first summand in the last term of this string of inequalities is zero by Theorem 1.19, and so is the $\lim_{\delta\to 0}\limsup_{n\to\infty}$ of the second by Theorem 1.17 d). The direct part of Theorem 2.2 is thus proved.

□

2.8. Proof of the converse part of Theorem 2.2. We assume that $\mathcal{F}$ is image admissible Suslin, and that $\mathcal{F} \in B_{pr}CLT$. Then, the bootstrap clt holds in probability for $f(X_1)$ with norming constants $a_n = \sqrt{n}$, where f is any finite linear combination of functions in $\mathcal{F}$. By Theorem 1.1 in Chapter 1, this implies that $f^2(X_1)$ is integrable and that $\sum_{i=1}^n \big(f(X_i) - \mathbb{E}f(X_1)\big)/\sqrt{n} \to_d G(f)$. In particular, by the definition of G_P, $\mathbb{E}G^2(f) = \mathbb{E}G_{\mathrm{P}}^2(f)$. We thus have

$$\mathcal{F} \subset \mathcal{L}_2(\mathrm{P}) \ \text{ and } \ G = G_{\mathrm{P}}. \tag{2.11}$$

At this point we can assume $\mathbb{E}f(X_1) = 0$ for all $f \in \mathcal{F}$. If (2.4) holds then for every subsequence there is a further subsequence, say $\{n_k\}$, such that

$$\lim_{k\to\infty} d_{BL}\big[\mathcal{L}^b\big(\nu_{n_k}^b(\omega)\big), \mathcal{L}(G_{\mathrm{P}})\big] = 0 \quad \omega - \text{a.s.} \tag{2.12}$$

where G is replaced by G_{P} by virtue of (2.11). (Actually, this limit holds for the measurable cover of $d_{BL}\big[\mathcal{L}^b\big(\nu_{n_k}^b(\omega)\big), \mathcal{L}(G_{\mathrm{P}})\big]$.) Let Ω_1 be the set of probability 1 where (2.12) holds. The argument for the proof of the converse part of Corollary 1.5 then shows that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr{}^b\Big\{\big\|\nu^b_{n_k}(\omega)\big\|_{\mathcal{F}'_\delta}>\varepsilon\Big\}=0,\quad \omega\in\Omega_1,\quad \varepsilon>0.$$

Also, $\mathcal{L}(G)=\mathcal{L}(G_{\mathrm{P}})$ being Radon, $(\mathcal{F},\rho_{\mathrm{P}})$ is separable. If D be a countable ρ_{P}-dense subset of $\mathcal{F}$ then there exists a set of probability one, say Ω_2, such that $\max_{1\le i\le n} f^2(X_i(\omega))/n\to 0$ for all $f\in D$ and all $\omega\in\Omega_2$ (since $\mathbb{E}f^2(X_1)<\infty$), and therefore we have

$$n\Pr{}^b\Big\{\big|f(X^b_{n,1}(\omega))\big|>\lambda\sqrt{n}\Big\}=\sum_{i=1}^n I_{|f(X_i(\omega))|>\lambda\sqrt{n}}\to 0,\quad \omega\in\Omega_2,\quad \lambda>0.$$

Hence, we can apply Theorem 1.20 to $\nu^b_{n_k}(\omega)$ and obtain

$$n_k\Pr{}^b\Big\{\big\|f(X^b_{n,1}(\omega))\big\|_{\mathcal{F}}>\lambda\sqrt{n_k}\Big\}\to 0,\quad \omega\in\Omega_1\cap\Omega_2,\quad \lambda>0,$$

or what is the same,

$$\sum_{i=1}^{n_k} I_{F(X_i)>\lambda\sqrt{n_k}}\to 0\quad \omega\in\Omega_1\cap\Omega_2,\quad \lambda>0, \tag{2.13}$$

where $F=F_c$ is the measurable envelope of the class $\mathcal{F}$ (that we are assuming centered). By Hoffmann-Jørgensen's inequality for sums of independent non-negative random variables (Theorem 1.13 and comments following it), this limit also holds in expectation (in fact for all moments), that is,

$$n_k\Pr\{F(X_1)>\delta\sqrt{n_k}\}\to 0. \tag{2.14}$$

The limit (2.12) can be rewritten as

$$\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\Big(\delta_{X^b_{n_k,i}}(\omega)-\mathrm{P}_{n_k}(\omega)\Big)\to G_{\mathrm{P}}\ \text{ in } \ell^\infty(\mathcal{F}),\quad \omega\in\Omega_1. \tag{2.15}$$

Now we need to improve this limit to include convergence of bootstrap moments. In the next arguments we leave ω implicit, but we assume it to be in $\Omega_1\cap\Omega_2$. In order to apply Hoffmann-Jørgensen's inequality we must control the bootstrap expected value of the maximum of the norms of the individual summands. For any $p>0$ and $a>0$,

$$\begin{aligned}\mathbb{E}^b\max_{i\le n_k}\left(\frac{\big\|\delta_{X^b_{n_k,i}}-\mathrm{P}_{n_k}\big\|_{\mathcal{F}}}{\sqrt{n_k}}\right)^p&\le 2^p\max_{i\le n_k}\left(\frac{F(X_i)}{\sqrt{n_k}}\right)^p\\&\le 2^p\left[a+\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}F(X_i)I_{F(X_i)>a\sqrt{n_k}}\right]^p.\end{aligned}$$

Since the sum in (2.13) takes on only integer values, we have that $\sum_{i=1}^{n_k} I_{F(X_i)>a\sqrt{n_k}}$ is eventually zero. So, the last quantity is eventually $(2a)^p$ and therefore,

$$\sup_k \mathbb{E}^b \max_{i\le n_k}\left(\frac{\|\delta_{X^b_{n_k,i}} - \mathrm{P}_{n_k}\|_{\mathcal{F}}}{\sqrt{n_k}}\right)^p < \infty.$$

Then, using this and (2.15) in Hoffmann–Jørgensen's inequality for sums of i.i.d. random vectors (Theorem 1.13 together with Corollary 1.12), we obtain

$$\sup_k \mathbb{E}^b\|\nu^b_{n_k}\|^p_{\mathcal{F}} = \sup_k \mathbb{E}^b\left\|\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\left(\delta_{X^b_{n_k,i}} - \mathrm{P}_{n_k}\right)\right\|^p_{\mathcal{F}} < \infty$$

for all $p > 0$. This provides enough uniform integrability in (2.15) to conclude, by the continuous mapping theorem, that, for all $\omega \in \Omega_1 \cap \Omega_2$,

$$\lim_{k\to\infty} \mathbb{E}^b\left\|\nu^b_{n_k}(\omega)\right\|_{\mathcal{F}} = \mathbb{E}\left\|G_{\mathrm{P}}\right\|_{\mathcal{F}}, \tag{2.16}$$

$$\lim_{k\to\infty} \mathbb{E}^b\left\|\nu^b_{n_k}(\omega)\right\|_{\mathcal{F}'_\delta} = \mathbb{E}\left\|G_{\mathrm{P}}\right\|_{\mathcal{F}'_\delta}, \tag{2.17}$$

including $\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}} < \infty$ and $\lim_{\delta\to 0}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'_\delta} = 0$ (which we already knew from the theory of Gaussian processes). Now, we apply the left side inequality in Proposition 2.5 together with (2.7), to obtain, using bounded convergence and Fatou,

$$\begin{aligned}
&\lim_{\delta\to 0}\limsup_{n_k\to\infty}\mathbb{E}\left(\left\|\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}(N_i - 1)(\delta_{X_i} - \mathrm{P}_{n_k})\right\|_{\mathcal{F}'_\delta} \wedge M\right) \\
&\quad \le \lim_{\delta\to 0}\limsup_{n_k\to\infty}\mathbb{E}\left[\left(\mathbb{E}_N\left\|\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}(N_i - 1)(\delta_{X_i} - \mathrm{P}_{n_k})\right\|_{\mathcal{F}'_\delta}\right) \wedge M\right] \\
&\quad \le \lim_{\delta\to 0}\limsup_{n_k\to\infty}\mathbb{E}\left[\left(3\mathbb{E}^b\left\|\nu^b_{n_k}\right\|_{\mathcal{F}'_\delta}\right) \wedge M\right] \\
&\quad \le 3\lim_{\delta\to 0}\mathbb{E}\limsup_{n_k\to\infty}\mathbb{E}^b\left\|\nu^b_{n_k}\right\|_{\mathcal{F}'_\delta} \\
&\quad = 3\lim_{\delta\to 0}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'_\delta} = 0
\end{aligned} \tag{2.18}$$

for all $0 < M < \infty$.

Our last technical point will be to see that we can dispense with the centering P_{n_k} in the limit (2.18). Since $(\mathcal{F}, \rho_{\mathrm{P}})$ is totally bounded and $\mathbb{E}f(X_1) = 0$, the symmetrization inequality (1.16) gives that for all large n and for all $\tau > 0$,

$$\begin{aligned}
\Pr\left\{\|\mathrm{P}_n\|_{\mathcal{F}} > \tau\right\} &\le 4\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i\delta_{X_i}\right\|_{\mathcal{F}} > \frac{\tau}{4}\right\} \\
&\le \frac{16}{\tau}\mathbb{E}\left[\left(\mathbb{E}_\varepsilon\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i\delta_{X_i}\right\|_{\mathcal{F}}\right) \wedge \frac{\tau}{4}\right].
\end{aligned}$$

Using the fact that $\varepsilon_i|N_i - N_i'| =_d N_i - N_i'$ and Jensen, the last quantity is bounded from above by

$$\frac{16}{\tau}\mathbb{E}\Big[\Big(e^{-1}\mathbb{E}_{N,N'}\Big\|\frac{1}{n}\sum_{i=1}^{n}(N_i - N_i')\delta_{X_i}\Big\|_{\mathcal{F}}\Big)\wedge\frac{\tau}{4}\Big]$$
$$\leq \frac{16}{\tau}\mathbb{E}\Big[\Big(2e^{-1}\mathbb{E}_N\Big\|\frac{1}{n}\sum_{i=1}^{n}(N_i - 1)\delta_{X_i}\Big\|_{\mathcal{F}}\Big)\wedge\frac{\tau}{4}\Big]$$
$$= \frac{16}{\tau}\mathbb{E}\Big[\Big(2e^{-1}\mathbb{E}_N\Big\|\frac{1}{n}\sum_{i=1}^{n}(N_i - 1)(\delta_{X_i} - \mathrm{P}_n)\Big\|_{\mathcal{F}}\Big)\wedge\frac{\tau}{4}\Big],$$

where we use the fact that $n\mathrm{P}_n = \sum_{i=1}^{n}\delta_{X_i}$ in the last identity. In conclusion, we have that for every $\tau > 0$ there are $0 < \tau' < \infty$ and $c < \infty$ such that

$$\Pr\Big\{\|\mathrm{P}_{n_k}\|_{\mathcal{F}} > \tau\Big\} \leq \frac{c}{\tau'}\mathbb{E}\Big[\Big(\mathbb{E}_N\Big\|\frac{1}{n_k}\sum_{i=1}^{n_k}(N_i - 1)(\delta_{X_i} - \mathrm{P}_{n_k})\Big\|_{\mathcal{F}}\Big)\wedge\tau'\Big]$$

for all k large enough. But, by bounded convergence and Fatou, as in (2.18), the limit (2.14) implies that the limsup as $n_k \to \infty$ of this last sequence is zero (actually, the general term is dominated by a constant times $n_k^{-1/2}$). This implies that

$$\lim_{n_k\to\infty}\mathbb{E}\Big[\Big(\Big|\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}(N_i - 1)\Big|\|P_{n_k}\|_{\mathcal{F}'_\delta}\Big)\wedge M\Big] = 0$$

for all $\delta > 0$ and therefore, the limit (2.18) becomes

$$\lim_{\delta\to 0}\limsup_{n_k\to\infty}\mathbb{E}\Big(\Big\|\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}(N_i - 1)\delta_{X_i}\Big\|_{\mathcal{F}'_\delta}\wedge M\Big) = 0.$$

Hence, by Theorem 1.3 and Corollary 1.5,

$$d_{BL(\ell^\infty(\mathcal{F}))}\Big(\frac{1}{\sqrt{n'_k}}\sum_{i=1}^{n_k}(N_i - 1)\delta_{X_i}, G_{\mathrm{P}}\Big) \to 0.$$

Since this happens for a subsequence of every subsequence, we conclude

$$d_{BL(\ell^\infty(\mathcal{F}))}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(N_i - 1)\delta_{X_i}, G_{\mathrm{P}}\Big) \to 0.$$

But then Theorem 1.19 implies that $\mathcal{F}$ is P–Donsker.

□

As in the case of real random variables, proving that the bootstrap clt implies the clt for the original sample seems to require more work than proving that the clt can be bootstrapped, which is a more useful result.

2.9. REMARK. Imbedded in the proof of the converse part of Theorem 2.2 is the interesting fact that if the bootstrap of the empirical process works in outer probability then all the bootstrap moments of the sup norm of the empirical bootstrap process converge in outer probability to the corresponding moments of the limit, at least under the image admissible Suslin hypothesis.

2.10. REMARK. It can be rightly argued that the object of the bootstrap is not to recover the limit distribution of the statistics of interest, but to approximate the distributions of these statistics (for each n, as $n \to \infty$). In connection with this it is worth mentioning that recently D. Radulović has shown that, contrary to the situation in $\mathbb{R}$, there do exist classes $\mathcal{F}$ which are not P–Donsker and for which $d_{BL(\ell^\infty(\mathcal{F}))}(\nu_n, \nu_n^b) \to 0$ in probability.

2.3. The almost sure Efron's bootstrap. We recall that if f is a not necessaraly measurable random element then f^* denotes its measurable cover. We say that $f_n \to f$ *almost uniformly*, a.u. for short, if $\|f_n - f\|^* \to 0$ a.s. (Dudley, 1985).

3.1. DEFINITION. *We say that a class of measurable functions $\mathcal{F} \subset \mathcal{L}_2(\mathrm{P})$ satisfying condition (1.9) is bootstrap* P*–Donsker a.s., or $\mathcal{F} \in B_{a.s.}CLT(\mathrm{P})$, if*

$$\lim_{n\to\infty} d_{BL}\left[\mathcal{L}^b\left(\nu_n^{\mathrm{P}_n}(\omega)\right), \mathcal{L}(G_{\mathrm{P}})\right] = 0 \quad \text{a.u.} \tag{3.1}$$

In analogy with the previous section, our object here is to prove the following theorem of Giné and Zinn (1990) on the a.s. bootstrap of empirical processes. The comments on measurability in Theorem 2.2 apply also to Theorem 3.2. In particular, the version we present assumes no measurability in the direct part, but it does assume some for the converse. The measurability assumption for the direct part, which required no changes from our proof, was removed by Strobl (1994) and Van der Vaart and Wellner (1996).

3.2. THEOREM. *If $\mathcal{F}$ is P–Donsker and $\mathbb{E}^*\|f(X_1) - \mathrm{P}f\|^2 < \infty$ then $\mathcal{F}$ is also bootstrap* P*–Donsker a.s. Conversely, if a class of measurable functions $\mathcal{F}$ with everywhere finite envelope F is image admissible Suslin and there exists a centered Gaussian process G indexed by $\mathcal{F}$ whose law is Radon in $\ell^\infty(\mathcal{F})$ and such that*

$$\lim_{n\to\infty} d_{BL}\left[\mathcal{L}^b\left(\nu_n^{\mathrm{P}_n}(\omega)\right), \mathcal{L}(G)\right] = 0 \quad \text{a.u.} \tag{3.2}$$

then $\mathcal{F}$ is P–Donsker, $\mathbb{E}F^2(X_1) < \infty$ and $G = G_{\mathrm{P}}$.

3.3. PROOF OF THE CONVERSE PART OF THEOREM 3.2. In view of Theorem 2.2, and since $B_{a.s.}CLT \subset B_{pr}CLT$, it just remains to be proved that $\mathbb{E}F^2(X_1) < \infty$. Since $\mathbb{E}|f(X_1)| < \infty$ by the proof of Theorem 2.2, we can assume $f(X_1)$ centered for all $f \in \mathcal{F}$. Proceeding as in the proof of (2.13) in the previous section, we now obtain, instead,

$$\sum_{i=1}^n I_{F(X_i) > \sqrt{n}} \to 0 \quad \text{a.s.}$$

In fact, since these variables only take (non–negative) integer values, this sum is 0 eventually a.s. (that is, for all $n > n(\omega)$ with $n(\omega) < \infty$ a.s.). In particular, $F(X_n) \le \sqrt{n} \le 1$ eventually a.s. Therefore, Borel–Cantelli gives

$$\sum_{n=1}^{n} \Pr\{F^2(X_n) > n\} < \infty,$$

that is, $\mathbb{E}F^2(X_1) < \infty$.

□

Regarding technique, the novelty here resides in the proof of the direct part of the theorem. It requires a lemma of Ledoux, Talagrand and Zinn (Ledoux and Talagrand, 1988) expressing in infinite dimensions the fact that if $\mathbb{E}X^2 < \infty$ and ξ is symmetric with $\mathbb{E}\xi^2 = 1$, then, letting X_i, ξ_i, be independent copies of X and ξ respectively, all independent,

$$\lim_{n\to\infty} \mathbb{E}_\xi \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i X_i\Big)^2 = \mathbb{E}X^2 \quad \text{a.s.}$$

by the law of large numbers. Surprisingly, the analogous result in infinite dimensions is not easy to prove. We know of two proofs for it, one using 'Yurinskii's trick' as modified by Talagrand, which was the original proof, and the other using more recent inequalities based on Talagrand's isoperimetric methods (as given in the book of Ledoux and Talagrand, 1991). We give here the original proof because it requires less technique. But before, we give an auxiliary lemma on a small variation about the main argument in Hoffmann-Jørgensen's inequality. That it works without measurability assumptions seems to have been noticed first by Strobl (1994) and van der Vaart and Wellner (1996).

3.4. Lemma. *For any $n \in \mathbb{N}$, let ξ_i and Y_i, $i = 1, \dots, n$, be the coordinates in a product probability space, ξ_i real, symmetric random variables with $\mathbb{E}|\xi_i| \le 1$, and Y_i S–valued with $\|f(Y_i)\|_{\mathcal{F}} \le \varepsilon$ for some $0 < \varepsilon < \infty$, where $\mathcal{F}$ is a class of measurable functions on S. Then,*

$$\Pr\Big\{\mathbb{E}_\xi\Big\|\sum_{i=1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > 2t + \varepsilon\Big\} \le \Big(\Pr\Big\{\mathbb{E}_\xi\Big\|\sum_{i=1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > t\Big\}\Big)^2, \tag{3.3}$$

where h^* denotes the mesurable cover of h with respect to all the variables jointly.

Proof. Set $U_k := \mathbb{E}_\xi\big\|\sum_{i=1}^{k} \xi_i f(Y_i)\big\|_{\mathcal{F}}^*$, $1 \le k \le n$, and $T := \min\{k : U_k \ge t\}$. Then,

$$\begin{aligned} U_n &\le U_{j-1} + \mathbb{E}_\xi\big\|\xi_j f(Y_j)\big\|_{\mathcal{F}}^* + \mathbb{E}_\xi\Big\|\sum_{i=j+1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* \\ &\le U_{j-1} + \varepsilon + \mathbb{E}_\xi\Big\|\sum_{i=j+1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* \text{ a.s.} \end{aligned}$$

and note that $U_{j-1} < t$ on $T = j$. Therefore, up to a set of probability zero,

$$\{T = j, U_n > 2t + \varepsilon\} \subseteq \Big\{T = j, \mathbb{E}_\xi \Big\| \sum_{i=j+1}^{n} \xi_i f(Y_i) \Big\|_{\mathcal{F}}^* > t\Big\}.$$

Hence, since the two variables in the last set are independent as each of them is in fact a function of different coordinates of the product probability space (see Dudley and Philipp, 1983, Section 2), we have

$$\begin{aligned}\Pr\Big\{\mathbb{E}_\xi \Big\|\sum_{i=1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > 2t + \varepsilon\Big\} &\le \sum_{j=1}^{n} \Pr\Big\{T = j, \mathbb{E}_\xi \Big\| \sum_{i=j+1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > t\Big\} \\ &= \sum_{j=1}^{n} \Pr\{T = j\} \Pr\Big\{\mathbb{E}_\xi \Big\| \sum_{i=j+1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > t\Big\}.\end{aligned}$$

Now, a slight modification of the arguments in the proof of (1.11) give

$$\mathbb{E}_\xi \Big\| \sum_{i=j+1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* \le \mathbb{E}_\xi \Big\|\sum_{i=1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^*,$$

and the lemma follows because

$$\sum_{j=1}^{n} \Pr\{T = j\} \le \Pr\Big\{\mathbb{E}_\xi \Big\|\sum_{i=1}^{n} \xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > t\Big\}.$$

□

In order to estimate expected values of norms of truncated sums at an important step in the next proposition, we will also need the following observation. It is an instance of a 'contraction principle' (Dudley, unpublished; also given in van der Vaart and Wellner, 1996). If the three random variables ξ, (τ, X) and Y are coordinate functions on a product probability space, X and Y are S–valued, ξ and τ are real valued, ξ is centered and $0 \le \tau \le 1$, then

$$\mathbb{E}^* \|\xi \tau f(X) + f(Y)\|_{\mathcal{F}} \le \mathbb{E}^* \|\xi f(X) + f(Y)\|_{\mathcal{F}}. \tag{3.4}$$

To see this just note the following estimates, where the last inequality is a consequence of inequality (1.11), and the last identity follows from the fact that if h is finite and measurable, $(hf)^* = hf^*$ a.s.:

$$\begin{aligned}\mathbb{E}^* \|\xi \tau f(X) + f(Y)\|_{\mathcal{F}} &\le \mathbb{E}^* \big\|\tau\big(\xi f(X) + f(Y)\big)\big\|_{\mathcal{F}} + \mathbb{E}^* \|f(Y)(1-\tau)\|_{\mathcal{F}} \\ &= \mathbb{E}\big(\|\xi f(X) + f(Y)\|_{\mathcal{F}}^* \tau\big) + \mathbb{E}^*\big(\|f(Y) + (\mathbb{E}\xi) f(X)\|_{\mathcal{F}} (1-\tau)\big) \\ &\le \mathbb{E}\big(\|\xi f(X) + f(Y)\|_{\mathcal{F}}^* \tau\big) + \mathbb{E}\big(\|f(Y) + \xi f(X)\|_{\mathcal{F}}^* (1-\tau)\big) \\ &= E^* \|\xi f(X) + f(Y)\|_{\mathcal{F}}.\end{aligned}$$

The Ledoux–Talagrand–Zinn lemma is as follows:

3.5. THEOREM. *Let ξ_i and Y_i, $i \in \mathbb{N}$, be coordinate functions on an infinite product probability space such that the variables ξ_i are real, symmetric, i.i.d. and $\mathbb{E}|\xi_i| \le 1$, and the variables Y_i are i.i.d. and take values in S. Let $\mathcal{F}$ be a class of measurable functions on S such that $\|f(Y_1)\|_{\mathcal{F}} < \infty$ and $\mathbb{E}^*\|f(Y_1)\|_{\mathcal{F}}^2 < \infty$. Then,*

$$\limsup_{n\to\infty} \mathbb{E}_\xi \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i f(Y_i) \Big\|_{\mathcal{F}}^* \le 2\sqrt{2} \limsup_{n\to\infty} \mathbb{E}^* \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i f(Y_i) \Big\|_{\mathcal{F}} \quad \text{a.s.,} \tag{3.5}$$

where h^* denotes the measurable cover of h with respect to all the variables jointly.

PROOF. Denote by M the limsup at the right of (3.5), and assume it is finite (otherwise, there is nothing to prove). For ease of notation we drop the subindex $\mathcal{F}$ from the norm signs. By Borel–Cantelli, the proof of the theorem reduces to showing

$$\sum_{n\ge n_0} \Pr\Big\{ \max_{2^{n-1}<k\le 2^n} \frac{1}{\sqrt{k}} \mathbb{E}_\xi \Big\| \sum_{i=1}^{k} \xi_i f(Y_i) \Big\|^* > \sqrt{2}(2M + 5\varepsilon) \Big\} < \infty$$

for all $\varepsilon > 0$ and for some $n_0 < \infty$. Since $\mathbb{E}_\xi \Big\| \sum_{i=1}^{k} \xi_i f(Y_i) \Big\|^*$ increases with k a.s. by (1.11), we only have to prove

$$\sum_{n\ge n_0} \Pr\Big\{ \frac{1}{2^{n/2}} \mathbb{E}_\xi \Big\| \sum_{i=1}^{2^n} \xi_i f(Y_i) \Big\|^* > 2M + 5\varepsilon \Big\} < \infty$$

for all $\varepsilon > 0$. Because $\mathbb{E}^*\|f(Y_1)\|^2 < \infty$ we can truncate $f(Y_i)$ at the level $\varepsilon\sqrt{2^n}$: letting

$$u_i = \big\{ \{ f(Y_i) I_{\|f(Y_i)\|^* \le \varepsilon 2^{n/2}} / 2^{n/2} : f \in \mathcal{F} \big\} \in \ell_\infty(\mathcal{F})$$

we have

$$\sum \Pr^* \Big\{ \text{there exist } i \le 2^n \text{ such that } u_i \ne \frac{\delta_{Y_i}}{2^{n/2}} \Big\} \le \sum 2^n \Pr\big\{ \|f(Y_1)\|^* > \varepsilon 2^{n/2} \big\} < \infty.$$

So, we only need to show

$$\sum_{n\ge n_0} \Pr\Big\{ \mathbb{E}_\xi \Big\| \sum_{i=1}^{2^n} \xi_i u_i \Big\|^* > 2M + 5\varepsilon \Big\} < \infty.$$

Furthermore, by Lemma 3.4 (applied with $t = M + 2\varepsilon$) this will follow if we prove

$$\sum_{n\ge n_0} \bigg(\Pr\Big\{ \mathbb{E}_\xi \Big\| \sum_{i=1}^{2^n} \xi_i u_i \Big\|^* > M + 2\varepsilon \Big\} \bigg)^2 < \infty. \tag{3.6}$$

Inequality (3.4) gives, by iteration,

$$\mathbb{E}\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* \le \mathbb{E}\Big\|\sum_{i=1}^{2^n}\xi_i f(Y_i)/2^{n/2}\Big\|^*$$

(take $\tau = I_{\|\delta_{Y_1}\|_{\mathcal{F}}^* \le \varepsilon 2^{n/2}}$ for the first application of (3.4), and so on). Now, by the definition of M,

$$\mathbb{E}\Big\|\sum_{i=1}^{2^n}\xi_i f(Y_i)/2^{n/2}\Big\|^* < M+\varepsilon, \tag{3.7}$$

for all large n. So, we can center in (3.6), which reduces the problem to showing

$$\sum_{n=1}^{\infty}\Big(\Pr\Big\{\mathbb{E}_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* - \mathbb{E}_Y\mathbb{E}_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* > \varepsilon\Big\}\Big)^2 < \infty. \tag{3.8}$$

We will show the series (3.8) converges by means of a martingale difference decomposition of its terms due to Yurinskii (1974) (with a key observation by Talagrand). Let $\mathbb{E}_{X,i}$ denote expectation with respect to the variables $X_{i+1},\dots,X_n$ only. We then have

$$\begin{aligned} E_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* - E\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* &= \sum_{i=1}^{2^n}(\mathbb{E}_{X,i}-\mathbb{E}_{X,i-1})\Big(E_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^*\Big)\\ &= \sum_{i=1}^{2^n}(\mathbb{E}_{X,i}-\mathbb{E}_{X,i-1})\Big(E_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* - E_\xi\Big\|\sum_{j\le 2^n, j\ne i}\xi_j u_j\Big\|^*\Big) \end{aligned} \tag{3.9}$$

since the subtracted term does not depend on X_i (as indicated above). Set

$$f_i := E_\xi\Big\|\sum_{k=1}^{2^n}\xi_k u_k\Big\|^* - E_\xi\Big\|\sum_{j\le 2^n, j\ne i}\xi_j u_j\Big\|^*.$$

Since $0\le f_i\le \mathbb{E}_\xi\|\xi_i u_i\|^* = \|u_i\|^*\mathbb{E}|\xi_i|$ a.s. and $\mathbb{E}|\xi_i|\le 1$, we have

$$0\le f_i\le \|u_i\|^* \quad \text{a.s.} \tag{3.10}$$

Since

$$\sum_{k=1}^{2^n}\xi_k u_k = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j\le 2^n, j\ne i}\xi_j u_j,$$

it follows that

$$\mathbb{E}\Big\|\sum_{k=1}^{2^n}\xi_k u_k\Big\|^* \le \frac{2^n}{2^n-1}\mathbb{E}\Big\|\sum_{j\le 2^n, j\ne i}\xi_j u_j\Big\|^*$$

and therefore, using (3.6) and (3.7),

$$\mathbb{E}f_i \le \frac{1}{2^n}\mathbb{E}\Big\|\sum_{j\le 2^n, j\ne i}\xi_j u_j\Big\|^* \le \frac{1}{2^n}\mathbb{E}\Big\|\sum_{j\le 2^n, j\ne i}\xi_j u_j\Big\|^* < \frac{M+\varepsilon}{2^n}. \tag{3.11}$$

((3.11) is Talagrand's observation). The summands in the decomposition (3.9) are orthogonal. Therefore, (3.10) and (3.11) give

$$\begin{aligned}
\mathbb{E}\Big(E_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* - E\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^*\Big)^2 &= \sum_{i=1}^{2^n}\mathbb{E}f_i^2 \le \sum_{i=1}^{2^n}\mathbb{E}\Big((\|u_i\|^*)^{3/2}f_i^{1/2}\Big)\\
&\le \sum_{i=1}^{2^n}\Big(\mathbb{E}(\|u_i\|^*)^3\Big)^{1/2}(\mathbb{E}f_i)^{1/2}\\
&\le \sum_{i=1}^{2^n}\Big(\frac{M+\varepsilon}{2^n}\mathbb{E}(\|u_i\|^*)^3\Big)^{1/2}\\
&= \Big(2^n(M+\varepsilon)\mathbb{E}(\|u_1\|^*)^3\Big)^{1/2}.
\end{aligned}$$

Finally, we apply this estimate to bound the series (3.8) using Chebyshev. We obtain

$$\begin{aligned}
&\sum_{n=1}^{\infty}\Big(\Pr\Big\{\mathbb{E}_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* - \mathbb{E}_Y\mathbb{E}_\xi\Big\|\sum_{i=1}^{2^n}\xi_i u_i\Big\|^* > \varepsilon\Big\}\Big)^2\\
&\qquad\le \frac{M+\varepsilon}{\varepsilon^2}\sum_{n=1}^{\infty}2^n\mathbb{E}\Big[\Big(\frac{\|f(Y_1)\|^*}{2^{n/2}}\Big)^3 I_{\|f(Y_1)\|^*\le\varepsilon 2^{n/2}}\Big]\\
&\qquad= \frac{M+\varepsilon}{\varepsilon^2}\mathbb{E}\Big[(\|f(Y_1)\|^*)^3\sum_{n:\|f(Y_1)\|\le\varepsilon 2^{n/2}}\frac{1}{2^{n/2}}\Big]\\
&\qquad\le \frac{M+\varepsilon}{\varepsilon^2}\frac{2^{1/2}}{2^{1/2}-1}\mathbb{E}^*\|f(Y_1)\|^2 < \infty.
\end{aligned}$$

□

3.6. REMARK. If the variables ξ_i in Theorem 3.5 are taken to be centered (instead of symmetric), and the remaining hypotheses in Theorem 3.5 are left unchanged, then inequality (3.5) still holds, but with a different multiplicative constant at the right. This is an immediate consequence of inequality (1.11) and its proof.

3.7. REMARK. Theorem 3.5 can be complemented with moment boundedness, to be precise, basically the same proof above shows that, under the hypotheses of Theorem 3.5, we also have

$$\mathbb{E}\Big(\sup_{n\in\mathbb{N}}\mathbb{E}_\xi\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i f(Y_i)\Big\|_{\mathcal{F}}^*\Big) \le C\Big[\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i f(Y_i)\Big\|_{\mathcal{F}} + \mathbb{E}^*\|f(Y_1)\|_{\mathcal{F}}^2\Big]. \quad (3.12)$$

To see this, proceed as in the previous proof with the following two changes: replace M by $\bar{M} := \sup_{n\in\mathbb{N}}\mathbb{E}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i f(Y_i)\Big\|_{\mathcal{F}}^*$ (this will allow all the series in the proof to start at 1 instead of at an unspecified n_0) and observe that the series that controls the truncation, $\sum 2^n \Pr\{\|f(Y_1)\|^* > \varepsilon 2^{n/2}\}$ is dominated by $\mathbb{E}^*\|f(Y_1)\|^2/\varepsilon^2$. With these two changes, the previous proof gives

$$\Pr\Big\{\sup_{k\in\mathbb{N}} \Big\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\xi_i f(Y_i)\Big\|_{\mathcal{F}}^* > \sqrt{2}(2\bar{M}+5\varepsilon)\Big\} \le C\frac{\mathbb{E}^*\|f(Y_1)\|^2}{\varepsilon^2},$$

and (3.12) follows.

3.8. PROOF OF THE DIRECT PART OF THEOREM 3.2. We assume $\mathcal{F}$ is P–Donsker, $\mathrm{P}f = 0$ for all $f \in \mathcal{F}$ and $\mathbb{E}^*F^2(X_1) < \infty$. In complete analogy with the proof of Theorem 2.2, direct part, proving almost uniform convergence in (3.1) reduces to showing

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{E}^b\|\nu_n^b\|_{\mathcal{F}'_\delta} = 0 \quad \text{a.u.} \tag{3.13}$$

By Lemma 1.6 (c) (right side of inequality (1.14) and the comment on centering following it), and by the right side of inequality (2.8s) we have, for all $\delta > 0$,

$$\mathbb{E}^b\|\nu_n^b\|_{\mathcal{F}'_\delta} \le 2\mathbb{E}^b\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\delta_{X^b_{n,i}}\Big\|_{\mathcal{F}'_\delta} \le \frac{2e}{e-1}\mathbb{E}_{M,M'}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(M_i - M_i')\delta_{X_i}\Big\|_{\mathcal{F}'_\delta}, \tag{3.14}$$

where M_i, M_i',, $i \in \mathbb{N}$, are i.i.d. Poisson variables with parameter $1/2$, defined on the Ω' part of the basic probability space, and $\mathbb{E}_{N,N'}$ denotes expectation with respect to these variables only (as usual). Theorem 3.5 gives that for all $\delta > 0$,

$$\limsup_{n\to\infty}\mathbb{E}_{M,M'}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(M_i-M_i')\delta_{X_i}\Big\|_{\mathcal{F}'_\delta} \le 2\sqrt{2}\limsup_{n\to\infty}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(M_i-M_i')\delta_{X_i}\Big\|_{\mathcal{F}'_\delta}, \tag{3.15}$$

since $\mathbb{E}^*F^2 < \infty$. Now, since $\Lambda_{2,1}(M_1 - M_1') < \infty$ and $\mathcal{F}$ is P–Donsker, Theorem 1.19 implies

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(M_i - M_i')\delta_{X_i}\Big\|_{\mathcal{F}'_\delta} = 0. \tag{3.16}$$

Combining (3.14), (3.15) and (3.16) yields the limit (3.13).

□

2.4. The exchangeable bootstrap. Præstgaard and Wellner (1993) extended the exchangeable bootstrap clt (Section 1.2) to empirical processes. This section is devoted to their work. We change the notation about the weights from Section 1.2 by setting $W_n(j) := nw_n(j)$. We assume that in what follows the vector of weights, $\mathbf{W}_n = (W_n(1), \dots, W_n(n))$, $n \in \mathbb{N}$, is exchangeable and satisfies the conditions

E1. $W_n(j) \ge 0$ for all n and j, and $\sum_{j=1}^n W_n(j) = n$;

E2'. $\sup_n \Lambda_{2,1}(W_n(1)) := M(\mathbf{W}) < \infty$.

E3. $\max_{1\le j\le n} \frac{1}{\sqrt{n}}|W_n(j) - 1| \to_P 0$.

E4. $\frac{1}{n}\sum_{j=1}^n (W_n(j) - 1)^2 \to_P 1$.

Note that E.2' is a different condition from E.2 in Section 1.2 and that E.2' implies E.2. We will also assume throughout that the weights W are defined on the Ω' component of the basic probability space (in particular they are independent from the data $\mathbf{X}$). The weighted empirical measure, $\mathrm{P}_n^w(\omega)$ is defined as

$$\mathrm{P}_n^w(\omega) = \frac{1}{n}\sum_{j=1}^{n} W_n(j)\delta_{X_j(\omega)}, \quad \omega \in \Omega, \quad n \in \mathbb{N}, \tag{4.1}$$

and the weighted empirical process, ν_n^w, as

$$\begin{aligned}\nu_n^w(\omega) &:= \sqrt{n}\big(\mathrm{P}_n^w(\omega) - \mathrm{P}_n(\omega)\big) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big(W_n(j) - 1\big)\delta_{X_j(\omega)} \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} W_n(j)\big(\delta_{X_j(\omega)} - \mathrm{P}_n(\omega)\big), \quad \omega \in \Omega, \quad n \in \mathbb{N}. \end{aligned}\tag{4.2}$$

We skip reference to ω whenever no confusion may arise, and denote conditional probability and expectation with respect to the data $\mathbf{X}$ as Pr^w, $\mathbb{E}^w$, although sometimes we will be more specific and write e.g. $\mathbb{E}_N$ for integration with respect to only the variables N_i.

Properties E.2' and E.3 imply:

4.1. LEMMA. *If the sequence of non-negative weights $\mathbf{W}_n$ satisfies condition E.2' then the sequence $\{W_n(1)\}_{n=1}^{\infty}$ is uniformly square integrable, and if moreover it satisfies E.3, then*

$$\frac{1}{\sqrt{n}}\mathbb{E}\max_{1\le i\le n} W_n(i) \to 0. \tag{4.3}$$

PROOF. By monotonicity, condition E.2', that is,

$$\Lambda_{2,1}\big(W_n(1)\big) := \int_0^\infty \sqrt{\Pr\{W_n(1) > t\}}\,dt < \infty,$$

implies

$$\Pr\{W_n(1) > t\} = o(t^{-2}) \quad \text{as } t \to \infty,$$

so that

$$\begin{aligned}&\mathbb{E}W_n(1)^2 I_{W_n(1)>t} = t^2 \Pr\{W_n(1) > t\} + 2\int_t^\infty u \Pr\{W_n(1) > u\}du \\ &\le t^2 \Pr\{W_n(1) > t\} + 2\Big[\sup_{u\ge t}\sqrt{u^2 \Pr\{W_n(1) > u\}}\Big]\int_t^\infty \sqrt{\Pr\{W_n(1) > u\}}du \to 0\end{aligned}$$

as $t \to \infty$ and, taking $t = 0$ in these inequalities, $\sup_n \mathbb{E}W_n(1)^2 < \infty$. This and E.3 give (4.3).

□

The main ingredient in the proof of the bootstrap clt in probability for empirical processes is the following modification of the multiplier inequality given in Theorem 1.18.

4.2. THEOREM. *Let $\xi = (\xi_1, \dots, \xi_n)$ be a non-negative, exchangeable vector such that $\Lambda_{2,1}(\xi_1) < \infty$, let $R = (R(1), \dots, R(n))$ be a random vector uniformly distributed over the set π_n of all the permutations of $\{1, \dots, n\}$ and let $\mathbf{Z} = (Z_1, \dots, Z_n)$ be random elements of $\ell^\infty(\mathcal{F})$. Assume ξ, R and $\mathbf{Z}$ are coordiante functions on a product probability space. Then, for all $0 \le n_0 \le n$,*

$$\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Z_i\Big\|_{\mathcal{F}} \le n_0\Big(\mathbb{E}\max_{1\le i\le n}\frac{\xi_i}{\sqrt{n}}\Big)\mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^n \|Z_i\|^*\Big) + \Lambda_{2,1}(\xi_1)\max_{n_0<k\le n}\mathbb{E}^*\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^k Z_{R(i)}\Big\|_{\mathcal{F}}. \tag{4.4}$$

PROOF. Proceeding as in the first steps in the proof of Theorem 1.18, we have

$$\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Z_i\Big\|_{\mathcal{F}} = \mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Z_{R(i)}\Big\|_{\mathcal{F}} \le \int_0^\infty \mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\xi_i\ge t} Z_{R(i)}\Big\|_{\mathcal{F}} dt,$$

where the first identity is a direct consequence of the exchangeability of the vector ξ. If S is any random permutation then $R\circ S$ is again uniformly distributed on π_n and is independent of S, as is easy to check. Hence, if we take any (random) permutation S for which $\xi_{S(1)} \ge \xi_{S(2)} \ge \dots \ge \xi_{S(n)}$, we have, for all t,

$$\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\xi_i\ge t} Z_{R(i)}\Big\|_{\mathcal{F}} = \mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{\#\{i:\xi_i\ge t\}} Z_{R(i)}\Big\|_{\mathcal{F}}.$$

This allows us to continue as in the proof of Theorem 1.18 with only formal changes, except that we use the following obvious estimate at the right place:

$$\max_{1\le k\le n_0}\mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^k Z_{R(i)}\Big\|_{\mathcal{F}} \le \frac{n_0}{\sqrt{n}}\mathbb{E}\|Z_{R(i)}\|_{\mathcal{F}}^* = \frac{n_0}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|Z_i\|_{\mathcal{F}}^*,$$

where the last identity follows by (1.13).

□

4.3. THEOREM. *If $\mathcal{F}$ is P–Donsker and the exchangeable weights satisfy E.1–E.4 and are defined on Ω', then*

$$\lim_{n\to\infty} d_{BL}\Big[\mathcal{L}^w(\nu_n^w), \mathcal{L}(G_{\mathrm{P}})\Big] = 0 \text{ in outer probability.} \tag{4.5}$$

PROOF. Proceeding as in the beginning of the proof of Theorem 2.2, where the proof of that theorem is reduced to (2.7), with only formal changes, but using the exchangeable bootstrap of the mean (Theorem 2.1, Section 1, plus Cramér–Wold) to control $II_{n,r}(H)$ instead of the regular bootstrap of the mean, the proof of Theorem 4.3 reduces to showing

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}^*\Big(\mathbb{E}^w\|\nu_n^w\|_{\mathcal{F}'_\delta}\Big) = 0. \tag{4.6}$$

Now, by Theorem 4.2 applied with $Z_i = \delta_{X_i} - \mathrm{P}_n$ and $\xi = \mathbf{W_n}$, we have

$$\begin{aligned}
\mathbb{E}^*\Big(\mathbb{E}^w \big\|\nu_n^w\big\|_{\mathcal{F}'_\delta}\Big) &\le \mathbb{E}^*\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_n(i)\big(\delta_{X_i} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \\
&\le n_0\Big(\mathbb{E}\max_{1\le i\le n}\frac{W_n(i)}{\sqrt{n}}\Big)\mathbb{E}\Big\|\delta_{X_1} - \mathrm{P}_n\Big\|^*_{\mathcal{F}'_\delta} \\
&\quad + \Lambda_{2,1}\big(W_n(1)\big)\max_{n_0\le k\le n}\mathbb{E}^*\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^k \big(\delta_{X_{R(i)}} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \qquad (4.7)
\end{aligned}$$

where the first inequality follows from (1.13). Since

$$\mathbb{E}^*\Big\|\delta_{X_1} - \mathrm{P}_n\Big\|_{\mathcal{F}'_\delta} \le \mathbb{E}^*\Big\|\delta_{X_1} - \mathrm{P}\Big\|_{\mathcal{F}'_\delta} + \mathbb{E}^*\Big\|\mathrm{P}_n - \mathrm{P}\Big\|_{\mathcal{F}'_\delta} \le 4\mathbb{E}^*\Big\|\delta_{X_1} - \mathrm{P}\Big\|_{\mathcal{F}} < \infty$$

by Theorem 1.16 (the weak L_2 integrability of F_c^*), it follows from Lemma 4.1 that the first summand in the last term of (4.7) converges to 0 for all $n_0 \in \mathbb{N}$. By condition E.2', the second term is dominated by a constant times

$$\max_{n_0\le k\le n}\mathbb{E}^*\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^k \big(\delta_{X_{R(i)}} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta}.$$

Now, using (1.13) and (1.11)

$$\begin{aligned}
&\mathbb{E}^*\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^k \big(\delta_{X_{R(i)}} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \\
&\qquad = \frac{1}{\sqrt{k}}\frac{1}{\binom{n}{k-n_0}}\sum_{1\le j_1<\ldots<j_{k-n_0}\le n}\mathbb{E}^*\Big\|\sum_{i=n_0+1}^k \big(\delta_{X_{j(i)}} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \\
&\qquad = \frac{1}{\sqrt{k}}\mathbb{E}^*\Big\|\sum_{i=1}^{k-n_0}\big(\delta_{X_i} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \\
&\qquad \le \frac{1}{\sqrt{k}}\mathbb{E}^*\Big\|\sum_{i=1}^{k-n_0}\big(\delta_{X_i} - \mathrm{P}\big)\Big\|_{\mathcal{F}'_\delta} + \sqrt{k}\mathbb{E}^*\big\|\mathrm{P}_n - \mathrm{P}\big\|_{\mathcal{F}'_\delta} \\
&\qquad \le \frac{1}{\sqrt{k}}\mathbb{E}^*\Big\|\sum_{i=1}^{k}\big(\delta_{X_i} - \mathrm{P}\big)\Big\|_{\mathcal{F}'_\delta} + \sqrt{k}\mathbb{E}^*\big\|\mathrm{P}_n - \mathrm{P}\big\|_{\mathcal{F}'_\delta}.
\end{aligned}$$

$\mathcal{F}$ being P–Donsker, Theorem 1.17 c) gives

$$\lim_{\delta\to 0}\lim_{n_0\to\infty}\sup_{k\ge n_0}\frac{1}{\sqrt{k}}\mathbb{E}^*\Big\|\sum_{i=1}^k\big(\delta_{X_i} - \mathrm{P}\big)\Big\|_{\mathcal{F}'_\delta} = 0$$

and, of course, $\lim_{\delta\to 0}\limsup_{n\to\infty}\sqrt{n}\mathbb{E}^*\|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{F}'_\delta} = 0$. Now, (4.6) follows from (4.7) and these limits and inequalities.

□

The exchangeable bootstrap also works a.s. if $\mathbb{E}^* F_c^2(X_1) := \mathbb{E}^*\|f(X_1)-\mathrm{P}f\|^2 < \infty$:

4.4. THEOREM. *If $\mathcal{F}$ is P–Donsker and $\mathbb{E}^*\|f(X_1)-\mathrm{P}f\|^2 < \infty$ and the exchangeable weights satisfy E.1–E.4 and are defined on Ω', then*

$$\lim_{n\to\infty} d_{BL}\Big[\mathcal{L}^w(\nu_n^w), \mathcal{L}(G_\mathrm{P})\Big] = 0 \quad \text{a.u.} \tag{4.8}$$

The proof of this theorem is considerably subtler than that of the exchangeable bootstrap in probability. The main extra tool it uses, besides the Ledoux–Talagrand–Zinn inequality, is the following interesting observation of Hoeffding (1963) to the effect that a V–statistic with a symmetric kernel g can be written as a U–statistic with a uniquely determined symmetric kernel $\tilde{g}$ which is a convex combination of all the functions obtained from g by making some coordinates equal and then symmetrizing. The proof will also require the extension of the right side of inequality (2.8s) when the number of variables V_i^b is different from n (that is, for bootstrap sample size different from n) and some standard estimates for Poisson random variables. Here are statements and indications of proofs for these results, beginning with Hoeffding's.

4.5. THEOREM. *Let L be a real vector space, let $f : L \to \mathbb{R}$ be a convex function, and let $v_1, \dots, v_n \in L$ for some $n < \infty$. For $m \le n$, let $X_1, \dots, X_m$ be a random sample without replacement from $\{v_1, \dots, v_n\}$, and let $Y_1, \dots, Y_m$ be a random sample with replacement from the same set (that is, the Y's are independent and uniformly distributed over $v_1, \dots, v_n$). Then,*

$$\mathbb{E}f\Big(\sum_{i=1}^m X_i\Big) \le \mathbb{E}f\Big(\sum_{i=1}^m Y_i\Big). \tag{4.9}$$

PROOF. Let $g : L^m \to H$ be a symmetric function (that is, a function invariant under permutations of its entries), where H is another real vector space. Then,

$$\begin{aligned}\mathbb{E}g(X_1, \dots, X_m) &= \frac{1}{\binom{n}{m}} \sum_{1\le i_1<\dots<i_m\le n} g(v_{i_1} \dots, v_{i_m}),\\ \mathbb{E}g\Big(Y_1, \dots, Y_m\Big) &= \frac{1}{n^m} \sum_{1\le i_1,\dots,i_m\le n} g(v_{i_1} \dots, v_{i_m}),\end{aligned} \tag{4.10}$$

expressions of U and V statistic type respectively. The decomposition of the V–statistic based on g into a sum of U–statistics can be described as follows: There is a unique function $\tilde{g} : L^m \to H$ such that: 1) it is symmetric in its entries; 2) it is of the form

$$\tilde{g}(x_1, \dots, x_m) = \sum p(k; r_1, \dots, r_k; i_1, \dots, i_k) g(x_{i_1}, \overset{r_1)}{\dots}, x_{i_1}, \dots, x_{i_k}, \overset{r_k)}{\dots}, x_{i_k}), \tag{4.11}$$

where the weights p are positive, add up to one and are independent of g, L and H, and where the sum runs over all the positive integers $k, r_1, \dots, r_k, i_1, \dots, i_k$, such that $1 \le k \le m$, $i_1, \ldots, i_k$ are all different and do not exceed m, and $r_1 + \ldots + r_k = m$; and 3)

$$\frac{1}{n^m} \sum_{1 \le i_1, \ldots, i_m \le n} g(x_{i_1} \ldots, x_{i_m}) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \ldots < i_m \le n} \tilde{g}(x_{i_1} \ldots, x_{i_m}) \tag{4.12}$$

for all $x_1, \ldots, x_n \in L$. In particular,

$$\mathbb{E}g(Y_1, \ldots, Y_m) = \mathbb{E}\tilde{g}(X_1, \ldots, X_m). \tag{4.12'}$$

For instance, for $m = 2$ and n arbitrary,

$$\tilde{g}(x_1, x_2) = \frac{n-1}{n} g(x_1, x_2) + \frac{1}{2n} g(x_1, x_1) + \frac{1}{2n} g(x_2, x_2).$$

The existence and uniqueness of $\tilde{g}$ satisfying the specified conditions can be proved by induction on m. If we take $H = L$ and $g(x_1, \ldots, x_m) = x_1 + \ldots + x_m$ then the function $\tilde{g}(x_1, \ldots, x_m)$ is a linear symmetric function of $x_1, \ldots, x_m$, hence it is a constant times $x_1 + \ldots + x_m$; but, since the variables X_i and Y_j all have the same expected value, it follows from (4.12') and (4.12) that this constant is 1, that is, $\tilde{g}(x_1, \ldots, x_m) = x_1 + \ldots + x_m$. (We are abusing notation here: we mean different sets of variables X_i and Y_j, defined for each possible set of n points in L.) Then, (4.11) gives

$$\sum p(k; r_1, \ldots, r_k; i_1, \ldots, i_k)(r_1 x_{i_1} + \ldots + r_k x_{i_k}) = x_1 + \ldots + x_m. \tag{4.13}$$

Let now f be a convex real function on L and let $g(x_1, \ldots, x_m) := f(x_1 + \ldots + x_m)$. Then, (4.11), (4.13) and the convexity of f give

$$\begin{aligned}
g(x_1, \ldots, x_m) &:= f(x_1 + \ldots + x_m) \\
&= f\Big(\sum p(k; r_1, \ldots, r_k; i_1, \ldots, i_k)(r_1 x_{i_1} + \ldots + r_k x_{i_k})\Big) \\
&\le \sum p(k; r_1, \ldots, r_k; i_1, \ldots, i_k) f(r_1 x_{i_1} + \ldots + r_k x_{i_k}) \\
&= \sum p(k; r_1, \ldots, r_k; i_1, \ldots, i_k) g(x_{i_1}, \overset{r_1)}{\ldots}, x_{i_1}, \ldots, x_{i_k}, \overset{r_k)}{\ldots}, x_{i_k}) \\
&= \tilde{g}(x_1, \ldots, x_m).
\end{aligned}$$

Therefore, by (4.12'),

$$\mathbb{E}f\Big(\sum_{i=1}^m X_i\Big) \le \mathbb{E}\tilde{g}(X_1, \ldots, X_m) = \mathbb{E}g(Y_1, \ldots, Y_m) = \mathbb{E}f\Big(\sum_{i=1}^m Y_i\Big).$$

□

The proof of the following lemma differs only formally from that of the right side inequalities in Proposition 2.5 and Remark 2.6, and therefore is omitted.

4.6. LEMMA. *Under the hypotheses for (2.8s) in Remark 2.6, but with M_i and M_i' replaced by $N_i(m/2n)$ and $N_i'(m/2n)$, i.i.d. Poisson variables with parameter $m/2n$, we have, for any $m \in \mathbb{N}$,*

$$\mathbb{E}\Big\|\sum_{i=1}^{m} \varepsilon_i V_i^b\Big\| \leq \frac{e}{e-1}\mathbb{E}\Big\|\sum_{i=1}^{n}\Big(N_i\Big(\frac{m}{2n}\Big) - N_i'\Big(\frac{m}{2n}\Big)\Big)v_i\Big\|. \tag{4.14}$$

If $\tilde{N}(\lambda) := N(\lambda) - N'(\lambda)$ with $N(\lambda)$ and $N'(\lambda)$ independent Poisson variables with parameter λ, then direct computation gives $\mathbb{E}\tilde{N}(\lambda)^4 = 12\lambda^2 + 2\lambda$ for all $\lambda > 0$. By standard computations we can then derive from this the following consequences:

4.7. LEMMA. *$\tilde{N}(\lambda)$ satisfies*

$$\Lambda_{2,1}\big(\tilde{N}(\lambda)/\sqrt{\lambda}\big) \leq 4 \quad \text{for all } \lambda > 0, \tag{4.15}$$

and, if $\tilde{N}_i(\lambda)$ are i.i.d. copies of $\tilde{N}(\lambda)$, then

$$\mathbb{E}\max_{1\leq i\leq n}\frac{\tilde{N}_i(\lambda)}{\sqrt{\lambda}} \leq 2\sqrt{n}\Big(\frac{12}{n} + \frac{2}{\lambda n}\Big)^{1/4} \quad \text{for all } n \in \mathbb{N},\ \lambda > 0. \tag{4.16}$$

4.8. PROOF OF THEOREM 4.4. We assume $\mathcal{F}$ is P–Donsker, $\mathrm{P}f = 0$ for all $f \in \mathcal{F}$ and $\mathbb{E}^*F^2 < \infty$. In analogy with previous proofs (Theorems 2.2, 3.2 and 4.3), the proof of the theorem reduces to showing

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}^w\big\|\nu_n^w\big\|_{\mathcal{F}'_\delta} = 0 \quad \text{a.u.} \tag{4.17}$$

First we apply Theorem 4.2 conditionally on the sample, to the effect that

$$\begin{aligned}
\mathbb{E}^w\big\|\nu_n^w\big\|_{\mathcal{F}'_\delta} &= \mathbb{E}^w\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_n(i)(\delta_{X_i} - \mathrm{P}_n)\Big\|_{\mathcal{F}'_\delta} \\
&\leq n_0\Big(\mathbb{E}\max_{1\leq i\leq n}\frac{W_n(i)}{\sqrt{n}}\Big)\frac{1}{n}\sum_{i=1}^{n}\big\|\delta_{X_i} - \mathrm{P}_n\big\|_{\mathcal{F}'_\delta} \\
&\quad + M(\mathbf{W})\max_{n_0<k\leq n}\mathbb{E}_R\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^{k}\big(\delta_{X_{R_n(i)}} - \mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \\
&:= I_{n,n_0,\delta} + II_{n,n_0,\delta},
\end{aligned} \tag{4.18}$$

where R_n is a random vector unformly distributed over the permutations of $1, \ldots, n$ independent of everything else, and $\mathbb{E}_R$ denotes integration with respect to R_n only. Since $\Big(\sum_{i=1}^{n}\|\delta_{X_i} - \mathrm{P}_n\|_{\mathcal{F}'_\delta}/n\Big)^* \leq 4\mathrm{P}_n F^* \to 4\mathrm{P}F^*$ a.e. by the law of large numbers, Lemma 4.1 then gives

$$\lim_{\delta\to 0}\limsup_{n\to\infty} I_{n,n_0,\delta} = 0 \quad \text{a.u.} \tag{4.19}$$

for all $n_0 < \infty$. Regarding the second term in (4.18), we start by applying Hoeffding's inequality (4.9), to obtain

$$\begin{aligned}\mathbb{E}_R\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^{k}\big(\delta_{X_{R_n(i)}}-\mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} &\le \mathbb{E}^b\Big\|\frac{1}{\sqrt{k}}\sum_{i=n_0+1}^{k}\big(\delta_{X^b_{n,i}}-\mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta}\\ &\le \mathbb{E}^b\Big\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\big(\delta_{X^b_{n,i}}-\mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta},\end{aligned}$$

(where the last inequality follows from (1.11)) and then apply the Poissonization inequality (4.14) to the effect that

$$\mathbb{E}^b\Big\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\big(\delta_{X^b_{n,i}}-\mathrm{P}_n\big)\Big\|_{\mathcal{F}'_\delta} \le \frac{2e}{e-1}\mathbb{E}_{N,\varepsilon}\Big\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\tilde{N}_i(k/2n)}{\sqrt{k/n}}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'_\delta}.$$

(here we have also used the symmetrization inequality (1.14) and introduced Rademacher variables, which is allowed by symmetry). Contrary to the proof of Theorem 3.2, we are not yet prepared to apply the Ledoux–Talagrand–Zinn lemma. What we do before is to apply the multiplier inequality (4.4) once more, in conjunction with Lemma 4.7 and, noting that $n_0 \le k \le n$, obtain

$$\begin{aligned}\mathbb{E}_{N,\varepsilon}\Big\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\tilde{N}_i(k/2n)}{\sqrt{k/n}}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'_\delta} &= \mathbb{E}_{N,\varepsilon}\Big\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{|\tilde{N}_i(k/2n)|}{\sqrt{k/n}}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'_\delta}\\ &\le cn_1n_0^{-1/4}\mathrm{P}_nF^* + c\max_{n_1\le\ell\le n}\mathbb{E}_{R,\varepsilon}\Big\|\frac{1}{\sqrt{\ell}}\sum_{j=n_1+1}^{\ell}\varepsilon_i\delta_{X_{R_n(i)}}\Big\|_{\mathcal{F}'_\delta}\\ &:= I'_{n_1,n_0,n} + II'_{n_1,n,\delta},\end{aligned}$$

for all $n_1 \in \mathbb{N}$ and for some universal $c < \infty$ (which can be easily specified but whose value is irrelevant for this proof). Now,

$$\lim_{n_0\to\infty}\limsup_{n\to\infty} I'_{n_1,n_0,n} = 0 \quad \text{a.u.}$$

for all $n_1 \in \mathbb{N}$, by the law of large numbers for F^*. The last three sets of inequalities and the last two limits show that there is $c < \infty$ such that

$$\lim_{\delta\to 0}\limsup_{n_0\to\infty}\limsup_{n\to\infty}\big(II_{n,n_0,\delta}\big)^* \le \lim_{\delta\to 0}\limsup_{n_1\to\infty}\limsup_{n\to\infty} c\big(II'_{n_1,n,\delta}\big)^* \quad \text{a.s.}$$

Hence, by (4.18), the theorem will be proved if we show

$$\lim_{\delta\to 0}\limsup_{n_1\to\infty}\limsup_{n\to\infty}\big(II'_{n_1,n,\delta}\big)^* = 0 \quad \text{a.s.} \tag{4.20}$$

Let now

$$U_\ell := \mathbb{E}_\varepsilon\Big\|\frac{1}{\sqrt{\ell}}\sum_{j=1}^{\ell}\varepsilon_i\delta_{X_i}\Big\|^*_{\mathcal{F}'_\delta}$$

and let $\mathcal{S}_n$ denote the σ-algebra generated by all the measurable functions $f : S^{\mathbb{N}} \to \mathbb{R}$ symmetric in the first n coordinates. Then,

$$\mathbb{E}_{R,\varepsilon}\Big\|\frac{1}{\sqrt{\ell}}\sum_{j=n_1+1}^{\ell}\varepsilon_i\delta_{X_{R_n(i)}}\Big\|_{\mathcal{F}'_\delta} \leq \mathbb{E}_{R,\varepsilon}\Big\|\frac{1}{\sqrt{\ell}}\sum_{j=1}^{\ell}\varepsilon_i\delta_{X_{R_n(i)}}\Big\|^*_{\mathcal{F}'_\delta} = \mathbb{E}\Big(U_\ell\big|\mathcal{S}_n\Big) \text{ a.s.} \quad (4.21)$$

It follows from Remark 3.7 on boundedness of moments in the Ledoux–Talagrand–Zinn lemma and the hypotheses that

$$\mathbb{E}\sup_{\ell\in\mathbb{N}} U_\ell < \infty. \quad (4.22)$$

Therefore, the sequence $\mathbb{E}\Big(\sup_{\ell>n_1} U_\ell\big|\mathcal{S}_n\Big)$ is a reverse martingale, hence it converges a.s. and in L_1. But, its limit being measurable for the symmetric σ-algebra, it is a.s. a constant by the Hewitt–Savage zero–one law. Thus, we have

$$\mathbb{E}\Big(\sup_{\ell>n_1} U_\ell\big|\mathcal{S}_n\Big) \to \mathbb{E}\sup_{\ell>n_1} U_\ell \text{ a.s. as } n\to\infty. \quad (4.23)$$

By the Ledoux–Talagrand–Zinn lemma (Theorem 3.5)

$$\lim_{n_1\to\infty}\sup_{\ell>n_1} U_\ell \leq 2\sqrt{2}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'_\delta} \ \ a.s.$$

so that by (4.22) and dominated convergence,

$$\lim_{n_1\to\infty}\mathbb{E}\sup_{\ell>n_1} U_\ell \leq 2\sqrt{2}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'_\delta}, \quad (4.24)$$

which tends to zero as δ tends to zero (Remark 1.2). Therefore, by (4.21)–(4.24) and the definition of II', we have

$$\begin{aligned}\lim_{\delta\to 0}\limsup_{n_1\to\infty}\limsup_{n\to\infty}\big(II'_{n_1,n,\delta}\big)^* &\leq \lim_{\delta\to 0}\limsup_{n_1\to\infty}\limsup_{n\to\infty}\mathbb{E}\big(\sup_{\ell>n_1} U_\ell\big|\mathcal{S}_n\big)\\ &= \lim_{\delta\to 0}\limsup_{n_1\to\infty}\mathbb{E}\sup_{\ell>n_1} U_\ell\\ &\leq \lim_{\delta\to 0} 2\sqrt{2}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'_\delta} = 0 \text{ a.s.}\end{aligned}$$

This proves (4.20), hence the theorem.

□

As is the case in $\mathbb{R}$, here too, if the limit in E.4 is $c^2 > 0$, then the bootstrap limit in Theorems 4.3 and 4.4 is cG_{P}.

4.9. EXAMPLES. 1) $W_n(j) = \sum_{i=1}^n I_{U_i\in A(n,j)}$ gives the regular bootstrap. There is some work involved at showing that these weights satisfy E.4 with $c = 1$.

2) $W_n(i) = Y_i/\bar{Y}_n$, where Y_i are i.i.d. random variables with finite variance, gives the Bayesian bootstrap. Here $c^2 = \mathrm{Var}Y_1/(\mathbb{E}Y_1)^2$.

3) $W_n(j) = \frac{n}{m_n}\sum_{i=1}^{m_n} I_{U_i\in A(n,j)}$ gives the regular bootstrap with bootstrap sample size m_n. $c = 1$ if $m_n \to \infty$. So, the above results contain the bootstrap of the empirical process with arbitrary bootstrap sample size $m_n \to \infty$. The proofs in the previous two sections, using Lemmas 4.6 and 4.7, give also this for the bootstrap

in probability, and give it also for the bootstrap a.s. if the sequence m_n is regular. (These two lemmas were in fact noticed by Giné and Zinn to this effect and later communicated to Præstgaard and Wellner, who obviously made better use of them.)

4) $\mathbf{W}_n =_d \text{Multinomial}\Big(n; M_{n1}/n, \ldots, M_{nn}/n\Big)$ conditional on $\mathbf{M}_n = (M_{n1}, \ldots, M_{nn}) =_d \text{Multinomial}(n; 1/n, \ldots, 1/n)$, gives the double bootstrap. Here $c^2 = 2$.

5) The m out of n bootstrap without replacement also falls into this scheme.

Details on these examples can be found in Præstgaard and Wellner (1993) and van der Vaart and Wellner(1996).

2.5. Uniformly pregaussian classes of functions and the bootstrap. The results from the previous sections show that, contrary to the prevailing thought a few years ago, the clt (for empirical processes) does not have to hold uniformly in Q near P, for it to be bootstrappable 'at P'. However, uniformity helps: in fact, when the clt holds uniformly in P, then one can justify all kinds of reasonable resampling procedures. This section contains part of the pertinent theory.

Whereas the previous sections are almost self–contained, here we will rely heavily on two important results from Gaussian process theory namely the Slepian–Fernique comparison theorem and Sudakov's minorization. Fernique's Saint Flour 1974 Lecture notes or Ledoux and Talagrand (1991) are good references for this.

Notation: $\mathcal{P}(S)$ will denote the set of all probability measures on $(S, \mathcal{S})$ and $\mathcal{P}_f(S)$ will denote the set of all the probability measures on $(S, \mathcal{S})$ whose support is finite. G_{P} and Z_{P} will repectively denote versions of the P–Brownian bridge and the P–Brownian motion with sample paths as regular as possible. (Recall that these are centered Gaussian processes defined on $\mathcal{F} \subset \mathcal{L}_2(\mathrm{P})$ whose covariances are respectively those of $\delta_{X_1} - \mathrm{P}$ and $\varepsilon_1 \delta_{X_1}$.) ρ_{P} and e_{P} are the $\mathcal{L}_2$ pseudo–distances associated respectively to G_{P} and Z_{P}. Extrapolating notation from previous sections, given a pseudo–distance e on $\mathcal{F}$, we set

$$\mathcal{F}'(\delta, e) := \mathcal{F}'_{\delta,e} := \big\{f - g : f, g \in \mathcal{F}, e(f,g) \leq \delta\big\},$$

and we will be interested in $\mathcal{F}'(\delta, \rho_{\mathrm{P}})$ and $\mathcal{F}'(\delta, e_{\mathrm{P}})$. It is convenient to keep in mind that if $\mathrm{P} = \sum_{i=1}^m \alpha_i \delta_{x_i}$, then

$$G_{\mathrm{P}} = \sum_{i=1}^m \alpha_i^{1/2} g_i (\delta_{x_i} - \mathrm{P}) \quad \text{and} \quad Z_{\mathrm{P}} = \sum_{i=1}^m \alpha_i^{1/2} g_i \delta_{x_i}, \tag{5.1}$$

where the variables g_i are i.i.d. $N(0,1)$. In general, whether P is discrete or not, a version of Z_{P} is $G_{\mathrm{P}} + g\mathrm{P}$, with g independent of G_{P}, as can be seen by computing covariances.

5.1. DEFINITION. *A uniformly bounded class $\mathcal{F}$ of measurable functions on S is finitely uniformly pregaussian, $\mathcal{F} \in UPG_f$ for short, if*

$$\sup_{\mathrm{P} \in \mathcal{P}_f(S)} \mathbb{E} \|Z_{\mathrm{P}}\|_{\mathcal{F}} < \infty, \tag{5.2}$$

and

$$\lim_{\delta\to 0}\sup_{\mathrm{P}\in\mathcal{P}_f(S)} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} = 0. \tag{5.3}$$

In other words, a class of functions $\mathcal{F}$ is UPG_f if the processes Z_P indexed by $\mathcal{F}$ are well behaved uniformly in $\mathrm{P} \in \mathcal{P}_f(S)$. These classes were introduced by Sheehy and Wellner (1988) and studied by Giné and Zinn (1991) and Sheehy and Wellner (1992).

Note that in (5.2) and (5.3) we could have used any other moments, or even tail probabilities, to obtain an equivalent definition since the median of the sup of a Gaussian process dominates (up to fixed, universal constant factors) all its L_p norms (see e.g. the proof of Claim 3 within the proof of Theorem 5.3 below). We could also have used G_{P} instead of the simpler Z_{P}, but Z_{P} is more convenient when dealing with uniformly bounded classes as in this case $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded iff $(\mathcal{F}, \rho_{\mathrm{P}})$ is, and G_{P} has a Radon law in $\ell^\infty(\mathcal{F})$ iff Z_{P} does.

5.2. EXAMPLES. 1) Recall that $N(T, d, \varepsilon)$, the covering nombers of the pseudo–metric space (T, d), is the smallest number of (closed) d–balls of radius ε and centers in T needed to cover T. Recall also Dudley's entropy bound for the sup of a Gaussian process (e.g., Ledoux and Talagrand, 1991, Sections 11.1 and 12.1). This bound immediately gives that: If the class $\mathcal{F}$ satisfies

$$\sup_{P\in\mathcal{P}_f(S)} \int_0^\infty \sqrt{\log N(\mathcal{F}, e_{\mathrm{P}}, \varepsilon)}\,d\varepsilon < \infty \tag{5.4}$$

and

$$\lim_{\delta\to 0}\sup_{P\in\mathcal{P}_f(S)} \int_0^\delta \sqrt{\log N(\mathcal{F}, e_{\mathrm{P}}, \varepsilon)}\,d\varepsilon = 0, \tag{5.5}$$

then $\mathcal{F} \in UPG_f$. In particular, *uniformly bounded VC–subgraph classes of functions* are UPG_f, and so are the classes Pollard (1990) calls manageable and those he calls Euclidean. This is the statistically most important example of UPG_f. Gaenssler (1987) was first to prove the bootstrap clt for empirical processes based on Euclidean classes of sets. See Dudley (1987) for the definition and universal clt properties of VC–subgraph classes and several other types of classes of functions that satisfy conditions related to (5.4) and (5.5).

2) If $\mathcal{F} = \{f_k\}_{k=2}^\infty$ with $\|f_k\|_\infty = o\big((\log k)^{-1/2}\big)$, then $\mathcal{F} \in UPG_f$. To prove it, set $\alpha_k = (\log k)^{1/2}\|f_k\|_\infty$, which tends to zero by hypothesis, and $\bar\alpha_N = \sup_{\alpha\geq N}\alpha_k$, which also tends to zero. It is classical that if g_i are $N(0,1)$, not necessarily independent, then

$$\mathbb{E}\sup_{k\geq N}\frac{|\alpha_k g_k|}{(\log k)^{1/2}} \leq K\bar\alpha_N,$$

where K is a universal constant. Since

$$Z_{\mathrm{P}}(f_k) = \left(\mathbb{E}_{\mathrm{P}} f_k^2\right)^{1/2} g_k \ \text{ and } \ \left(\mathbb{E}_{\mathrm{P}} f_k^2\right)^{1/2} \leq \frac{\alpha_k}{(\log k)^{1/2}},$$

the previous inequality gives

$$\sup_{P \in \mathcal{P}_f(S)} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}} \leq K\bar{\alpha}_2$$

and

$$\begin{aligned}\mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'(\delta, e_{\mathrm{P}})} &\leq \mathbb{E}\sup_{k \geq N}|Z_{\mathrm{P}}(f_k)| + \mathbb{E}\Big(\sup_{k,\ell \leq N, e_{\mathrm{P}}(f_k, f_\ell) \leq \delta}\big|Z_{\mathrm{P}}(f_k) - Z_{\mathrm{P}}(f_\ell)\big|\Big) \\ &\leq 2K\bar{\alpha}_N + \delta N^2,\end{aligned}$$

so that we also have

$$\lim_{\delta \to 0}\sup_{\mathrm{P} \in \mathcal{P}_f(S)} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'(\delta, e_{\mathrm{P}})} = 0.$$

Some of these classes are not covered by Example 1 (see e.g. Dudley, 1987).

3) Let H be a separable, infinite dimensional Hilbert space and let H_1 be its unit ball centered at zero. Take $S = H_1$, $\mathcal{S}$ the Borel sets of S, and $\mathcal{F} = H_1$ acting on S by inner product. Since bounded random H-valued random vectors satisfy the central limit theorem (see e.g. Araujo and Giné, 1980), it follows that $\mathcal{F}$ is P-Donsker for all P (universal Donsker). However, it is not UPG_f. The reader can verify that (5.3) does not hold by considering the sequence of probability measures $Q_N = \sum_{i=1}^N \delta_{e_i}/N$, $N \in \mathbb{N}$, for an orthonormal basis e_i.

The two main reasons behind Definition 5.1 are that 1) as we will see, empirical processes indexed by UPG_f classes satisfy very strong uniformity in P properties, and so do their limiting Gaussian processes, and 2) Gaussian processes are sufficiently well understood so as to make it feasible, in general, to decide whether a given class satisfies the UPG_f property, and in fact, as the examples above show, there are many.

More notation: We set $\mathcal{F}\mathcal{F} := \{f^2, f : f \in \mathcal{F}\}$ and denote the class of Radon probability measures on $\ell^\infty(\mathcal{F})$ by $\mathcal{R}(\ell^\infty(\mathcal{F}))$. X_i, as usual will be the coordinates $S^{\mathbb{N}} \to S$, P_n the empirical measure based on the first n such coordinates, $\nu_n^{\mathrm{P}} := \sqrt{n}(\mathrm{P}_n - \mathrm{P})$ and, since here we will be interested in ν_n^{P} for all $\mathrm{P} \in \mathcal{P}(S)$, we will use the subindex P on the signs for expectation, probability, the bounded Lipschitz distance, etc. to indicate that we are integrating with respect to the probability $\mathrm{P}^{\mathbb{N}}$ on $S^{\mathbb{N}}$.

The main result in Giné and Zinn (1991) can be stated as follows:

5.3. THEOREM. *Let $\mathcal{F}$ be an image admissible Suslin UPG_f class. Then,*
i) the laws of G_{P} and Z_{P} are Radon in $\ell^\infty(\mathcal{F})$ for all $\mathrm{P} \in \mathcal{P}(S)$; in fact, $\mathcal{F}$ is UPG in the sense that (5.2) and (5.3) hold with $\mathcal{P}_f(S)$ replaced by $\mathcal{P}(S)$, and it also holds with Z_{P} replaced by G_{P};
ii) $\mathcal{F}$ is uniform Donsker in the sense that $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded uniformly in P and

$$\lim_{n \to \infty}\sup_{\mathrm{P} \in \mathcal{P}} d_{BL_{\mathrm{P}}}(\nu_n^{\mathrm{P}}, G_{\mathrm{P}}) = 0 \tag{5.6}$$

and iii), the map

$$G : (\mathcal{P}(S), \|\cdot\|_{\mathcal{F}\mathcal{F}}) \to (\mathcal{R}(\ell^\infty(\mathcal{F})), d_{BL}),$$

given by

$$G(P) = \mathcal{L}(G_{\mathrm{P}}), \tag{5.7}$$

is uniformly continuous.

Before proving the theorem we give a consequence for the bootstrap. (This consequence justifies including this section in these lectures).

5.4. COROLLARY. *Let $\mathcal{F}$ be an image admissible Suslin UPG_f class and let Q_n be random probability measures on $(S, \mathcal{S})$ such that, for some P$\in \mathcal{P}(S)$,*

$$\lim_{n\to\infty} \|\mathrm{Q}_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} = 0 \ \text{ a.s. } \ (\text{in pr.}) \tag{5.8}$$

where we assume $\|\mathrm{Q}_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}}$ to be measurable. Then,

$$\lim_{n\to\infty} d_{BL_{\mathrm{Q}_n}}\big(\nu_n^{\mathrm{Q}_n}, G_{\mathrm{P}}\big) = 0 \ \text{ a.u. } \ (\text{in outer pr.}) \tag{5.9}$$

PROOF. The proof is basically a triangle inequality. Assume the a.s. version of (5.8). Just note

$$d_{BL}\big(\nu_n^{\mathrm{Q}_n}, G_{\mathrm{P}}\big)^* \le d_{BL}\big(\nu_n^{\mathrm{Q}_n}, G_{\mathrm{Q}_n}\big)^* + d_{BL}\big(G_{\mathrm{Q}_n}, G_{\mathrm{P}}\big)^*.$$

Conclusion ii) in Theorem 5.3 implies that there exist $c_n \to 0$ such that $d_{BL}\big(\nu_n^{\mathrm{Q}_n}, G_{\mathrm{Q}_n}\big)^* \le c_n$. Conclusion iii) of the same theorem implies that given $\varepsilon > 0$ there is $\delta > 0$ such that, for all n,

$$d_{BL}\big(G_{\mathrm{Q}_n}, G_{\mathrm{P}}\big) \le \varepsilon I_{\|\mathrm{Q}_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} < \delta} + 2 I_{\|\mathrm{Q}_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} \ge \delta}$$

(recall that d_{BL} is bounded by 2). For each n, the right side is a measurable random variable and the limsup of these random variables as $n \to \infty$ is dominated by ε due to hypothesis (5.8). Hence, $d_{BL}\big(\nu_n^{\mathrm{Q}_n}, G_{\mathrm{P}}\big)^* \to 0$ a.s. The same proof for subsequences gives the in probability version of the result.

□

Corollary 5.4 was obtained by Giné and Zinn (1991, Corollary 2.7). Sheehy and Wellner (1988, Theorem 1.6) have a similar result for Efron's bootstrap with arbitrary bootstrap sample size tending to infinity with n.

5.5. EXAMPLES. 1) As we will see somewhere along the proof of Theorem 5.3, if $\mathcal{F}$ is image admissible Suslin and UPG_f then $\|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} \to 0$ $\mathrm{P}^{\mathbb{N}}$–a.s. for all P$\in \mathcal{P}(S)$. Then, $\mathrm{Q}_n = \mathrm{P}_n$ satisfies condition (5.8) and therefore Corollary 5.4 gives Efron's bootstrap a.s. for the empirical process with arbitrasry bootstrap sample size $m_n \to \infty$, that is, $\nu_n^{\mathrm{Q}_{m_n}} \to_{\mathcal{L}} G_{\mathrm{P}}$ a.u. with $\mathrm{Q}_n = \mathrm{P}_n$ and any m_n tending to infinity. (As mentioned above, this observation for arbitrary sample size m_n was first made by Sheehy and Wellner, 1988.)

2) Let $S = \mathbb{R}^d$ and suppose that the class $\mathcal{F}$ is closed by translations, image admissible Suslin and UPG_f. Let $\lambda_n \in \mathcal{P}(S)$ be smooth probability measures satisfying the 'approximate identity' condition $\|\lambda_n - \delta_0\|_{\mathcal{F}\mathcal{F}} \to 0$. A *smooth bootstrap* consists of sampling not from P_n but from $\mathrm{P}_n * \lambda_n$. Since $\|\mathrm{P}_n * \lambda_n - \mathrm{P} * \lambda_n\|_{\mathcal{F}\mathcal{F}} \le \|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} \to 0$ a.s. and $\|\mathrm{P} * \lambda_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} \le \|\lambda_n - \delta_0\|_{\mathcal{F}\mathcal{F}} \to 0$, it follows that $\|\mathrm{P}_n * \lambda_n - \mathrm{P}\|_{\mathcal{F}\mathcal{F}} \to 0$

a.s. and Corollary 5.4 then gives $\nu_n^{P_n*\lambda_n} \to_{\mathcal{L}} G_P$ a.s., thus justifying the smoothed bootstrap.

3) Suppose we know $P \in \{P_\theta : \theta \in \Theta\}$. For each n, let θ_n be an estimator of θ. The model based bootstrap in this situation consists of resampling not from P_n but from P_{θ_n}. If P_{θ_n} is as close or closer to P in the $\|\cdot\|_{\mathcal{F}\mathcal{F}}$ pseudo–norm as P_n is, then this bootstrap may be better than Efron's, and it is justified by Corollary 5.4 (assuming $\mathcal{F}$ is image admissible Suslin UPG_f).

4) Example 3 is a special case of the following more general situation considered by Romano (1989); see Arcones and Giné (1991) for an additional example. Let $\mathcal{Q}(S) \subset \mathcal{P}(S)$ ($\mathcal{Q}$ could be the set of product probability measures if S is a product space, it could be the set of probability measures which are symmetric about zero if $S = \mathbb{R}^d$, etc.). Let $\tau : \mathcal{P}(S) \to \mathcal{Q}(S)$ be a projection continuous for the pseudo–norm $\|\cdot\|_{\mathcal{F}\mathcal{F}}$. ($\tau$ does not have to be defined on all of $\mathcal{P}(S)$.) The above corollary then shows that resampling from τP_n is consistent whenever $P \in \mathcal{Q}(S)$, that is, $\nu_n^{\tau P_n} \to_{\mathcal{L}} G_P$ a.s., and that in general, $\nu_n^{\tau P_n} \to_{\mathcal{L}} G_{\tau P}$ a.s. (assuming $\mathcal{F}$ is image admissible Suslin and UPG_f). This is important for constructing bootstrap tests of the null hypothesis $P \in \mathcal{Q}(S)$.

Next we prove Theorem 5.3. The long proof is decomposed into several steps.

Before proving Theorem 5.3 we need two lemmas basically giving the conclusions ii) and iii) for $\mathcal{F}$ finite.

5.6. LEMMA. *Let $\mathcal{P}_M^d := \{P : P$ is a Borel probability measure on $\mathbb{R}^d$ with supp $P \subset \{\|x\| \le M\}\}$. For $P \in \mathcal{P}_M^d$, let ξ_i^P be i.i.d.(P) and let $\Phi_P = \mathrm{Cov}(P)$. Then,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_M^d} d_{BL}\Big[\mathcal{L}\Big(\sum_{i=1}^n (\xi_i^P - \mathbb{E}\xi_i^P)/\sqrt{n}\Big), N\big(0,\Phi_P\big)\Big] = 0 \qquad (5.10)$$

where $N(0,\Phi_P)$ denotes the centered normal law of $\mathbb{R}^d$ with covariance Φ_P.

PROOF (Sketch). This follows from standard results on speed of convergence in the multidimensional clt, but an elementary proof obtains along the following lines. The Lindeberg proof of the clt, as e.g. in Araujo and Giné (1980), pages 37 and 67, gives

$$d_3\Big[\mathcal{L}\Big(\sum_{i=1}^n (\xi_i^P - \mathbb{E}\xi_i^P)/\sqrt{n}\Big), N\big(0,\Phi_P\big)\Big] \le \frac{KM(\mathrm{trace}\Phi_P)}{\sqrt{n}}$$

where K is a universal constant and $d_3(\mu,\nu) := \sup\{\big|\int f d(\mu-\nu)\big| : \sum_{|\alpha|\le 3} \|D^\alpha f\|_\infty \le 1\}$, where α is a multi–index and D^α denotes partial derivatives with respect to the variables indicated by the index α (standard notation). For f such that $\|f\|_\infty \le 1$ and $\|f\|_{\mathrm{Lip}} \le 1$, and $\varepsilon > 0$, define

$$f_\varepsilon(x) = \frac{1}{(2\pi)^{d/2}} \int f(x-\varepsilon y) e^{-\|y\|^2/2} dy = \frac{1}{(2\pi\varepsilon^2)^{d/2}} \int f(v) e^{-\|v-x\|/2\varepsilon^2} dv$$

and note that

$$\|f - f_\varepsilon\|_\infty \le \frac{1}{(2\pi)^{d/2}} \int (2 \wedge \varepsilon\|y\|) e^{-\|y\|^2/2} dy \le c(d)\varepsilon,$$

where $c(d)$ depends only on d, whereas

$$\|D^{\alpha} f_{\varepsilon}\|_{\infty} \leq \frac{1}{\varepsilon^{|\alpha|}} \int |D^{\alpha} \varphi| dy$$

φ being the density of the standard normal law in $\mathbb{R}^d$. Using these estimates in a simple triangle inequality and taking into account that the trace of Φ_{P} is bounded by dM^2, we obtain

$$d_{BL}\Big[\mathcal{L}\Big(\sum_{i=1}^{n}(\xi_i^{\mathrm{P}} - \mathbb{E}\xi_i^{\mathrm{P}})/\sqrt{n}\Big), N(0, \Phi_{\mathrm{P}})\Big] \leq \frac{c(d) M^{3/4}}{n^{1/8}},$$

where $c(d)$ is a constant that only depends on d (not necessarily equal to the one above).

□

5.7. LEMMA. *Let Φ and Ψ be two covariances on $\mathbb{R}^d \times \mathbb{R}^d$ and let $N(0, \Phi)$ and $N(0, \Psi)$ be the centered normal laws of $\mathbb{R}^d$ with theses covariances. Set $\|\Phi - \Psi\|_{\infty} := \max_{i,j \leq d} |\Phi(e_i, e_j) - \Psi(e_i, e_j)|$, where $\{e_i\}$ is the canonical basis of $\mathbb{R}^d$. Then,*

$$d_{BL}\big[N(0, \Phi), N(0, \Psi)\big] \leq c(d) \|\Phi - \Psi\|_{\infty}^{1/4}, \tag{5.11}$$

where $c(d)$ is a constant that depends only on d.

PROOF. The approximation arguments in the previous proof show that it suffices to see that

$$d_3\big[N(0, \Phi), N(0, \Psi)\big] \leq c(d) \|\Phi - \Psi\|_{\infty}$$

for some constant $c(d)$ depending on d only. One way to prove this inequality is to proceed as in the proof of Lindeberg's theorem: compare $\mathbb{E} f(\sum_{i=1}^n X_i/\sqrt{n})$ with $\mathbb{E} f(\sum_{i=1}^n Y_i/\sqrt{n})$ where the variables X_i and Y_i are all independent, the X's distributed as $N(0, \Phi)$ and the Y's as $N(0, \Psi)$, by replacing one X by one Y at a time, and developing by Taylor up to the third term, as in Lindeberg's proof of the clt. Then the linear term is zero, the second term is bounded by $c(d)\|\Phi - \Psi\|_{\infty}$ and the third term is bounded by a constant times $n^{-1/2}$, and letting n tend to infinity gives the result.

□

It is also convenient to have at hand the statements of the results on Gaussian processes that we will use. These are the following:

5.8. THE SLEPIAN–FERNIQUE COMPARISON THEOREM. Let $Z_i(t)$, $t \in T$, $i = 1, 2$, be two centered Gaussian processes defined on a countable set T. If

$$\mathbb{E}\big[Z_1(t) - Z_1(s)\big]^2 \leq \mathbb{E}\big[Z_2(t) - Z_2(s)\big]^2$$

for all $s, t \in T$, then

$$\mathbb{E} \sup_{t \in T} Z_1(t) \leq \mathbb{E} \sup_{t \in T} Z_2(t).$$

If the processes have Radon laws in $\ell^{\infty}(T)$, the same is true without any restrictions on T for the bounded ρ_{Z_i} uniformly continuous version of Z_i, $i = 1, 2$ (T is separable for ρ_{Z_i} in this case).

5.9. SUDAKOV'S MINORIZATION. Let Z be a centered Gaussian process on T and let ρ_Z denote the induced L_2 distance. Let $N(T, \rho_Z, \varepsilon)$, $\varepsilon > 0$, denote the covering numbers of the pseudo-metric space (T, ρ_Z), that is, $N(T, \rho_Z, \varepsilon)$ is the smallest number of ρ-balls (closed) of radius ε needed to cover T. Then, there exist a universal constant $K < \infty$ such that

$$\sup_{\varepsilon>0} \varepsilon \sqrt{\log N(T, \rho_Z, \varepsilon)} \leq K \mathbb{E} \sup_{t\in T} Z(t).$$

5.10. PROOF OF THEOREM 5.3. We can assume $\mathcal{F}$ is uniformly bounded by 1.

CLAIM 1. Set $\mathcal{G} = \{f, f^2, f-g, (f-g)^2 : f, g \in \mathcal{F}\}$. Then,

$$\sup_{\mathrm{P}\in\mathcal{P}(S)} \mathbb{E}_{\mathrm{P}} \|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{G}} = O(n^{-1/2}). \tag{5.12}$$

Proof. (The proof of this claim contains the main idea behind the theorem.) We will prove (5.12) only for the smaller class $\mathcal{H} = \{(f-g)^2 : f, g \in \mathcal{F}\}$ since, as it will become apparent, a subset of the proof for this set gives the rest. Let g_i be i.i.d. $N(0,1)$ defined on Ω'. We then have, by Lemma 1.6 c) and the left side of (1.31) (the easy part!):

$$\mathbb{E}_{\mathrm{P}} \|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{H}} \leq 2\mathbb{E}_{\mathrm{P},\varepsilon} \Big\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \delta_{X_i} \Big\|_{\mathcal{H}} \leq \sqrt{\frac{2\pi}{n}} \mathbb{E}_{\mathrm{P},g} \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \delta_{X_i} \Big\|_{\mathcal{H}}. \tag{5.13}$$

Now we will change the index $\mathcal{H}$ to $\mathcal{F}' = \{f - g : f, g \in \mathcal{F}\}$ using the Slepian–Fernique comparison lemma for Gaussian processes. For $\omega \in \Omega$ fixed, consider the two Gaussian processes

$$\Big\{Z_1(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i f^2(X_i(\omega)) : f \in \mathcal{F}'\Big\}, \quad \Big\{Z_2(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i f(X_i(\omega)) : f \in \mathcal{F}'\Big\}.$$

Note then that: i)

$$\begin{aligned} \mathbb{E}\big(Z_1(f_1)) - Z_1(f_2)\big)^2 &= \mathrm{P}_n\big(f_1^2 - f_2^2\big)^2 = \mathrm{P}_n\Big[(f_1 - f_2)^2 (f_1 + f_2)^2\Big] \\ &\leq 16\mathrm{P}_n(f_1 - f_2)^2 = 16\mathbb{E}\big(Z_2(f_1)) - Z_1(f_2)\big)^2, \end{aligned}$$

ii) both processes have uniformly continuous sample paths with respect to their corresponding L_2 distances (recall ω is fixed), and iii) both Z_1 and Z_2 attain the value zero at one of the f's in $\mathcal{F}$ ($f = 0$). Therefore, applying the Slepian–Fernique theorem 5.8 to $\{Z_1, Z_2\}$ and to $\{-Z_1, -Z_2\}$ (as, by iii), $\mathbb{E}_g \sup_{f\in\mathcal{F}'} |Z_1| \leq \mathbb{E}_g \sup_{f\in\mathcal{F}'} Z_1 + \mathbb{E}_g \sup_{f\in\mathcal{F}'}(-Z_1)$), we obtain

$$\mathbb{E}_g \sup_{f\in\mathcal{F}'} |Z_1| \leq 8\mathbb{E}_g \sup_{f\in\mathcal{F}'} |Z_2|. \tag{5.14}$$

(5.13) and (5.14) give

$$\mathbb{E}_{\mathrm{P}} \|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{H}} \leq \sqrt{\frac{2\pi}{n}} \mathbb{E}_{\mathrm{P},g} \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \delta_{X_i} \Big\|_{\mathcal{H}} \leq 8\sqrt{\frac{2\pi}{n}} \mathbb{E}_{\mathrm{P},g} \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i \delta_{X_i} \Big\|_{\mathcal{F}'}$$

$$\leq 16\sqrt{\frac{2\pi}{n}}\mathbb{E}_{\mathrm{P},g}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i\delta_{X_i}\Big\|_{\mathcal{F}} = 16\sqrt{\frac{2\pi}{n}}\mathbb{E}_{\mathrm{P}}\mathbb{E}_g\big\|Z_{\mathrm{P}_n}\big\|_{\mathcal{F}}$$

$$\leq 16\sqrt{\frac{2\pi}{n}}\sup_{Q\in\mathcal{P}_f(S)}\mathbb{E}\big\|Z_Q\big\|_{\mathcal{F}}, \tag{5.15}$$

proving Claim 1.

CLAIM 2. $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded uniformly in $\mathrm{P}\in\mathcal{P}(S)$ and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\mathrm{P}\in\mathcal{P}(S)}\mathrm{P}^{\mathbb{N}}\big\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > \varepsilon\big\} = 0 \tag{5.16}$$

for all $\varepsilon > 0$.

Proof. (5.2) implies

$$\sup\Big\{\mathbb{E}\|Z_{\mathrm{P}_n(\omega)}\|_{\mathcal{F}} : \mathrm{P}\in\mathcal{P}(S),\ \omega\in S^{\mathbb{N}},\ n\in\mathbb{N}\Big\} < \infty$$

and therefore, by Sudakov's minorization, there is $c<\infty$ such that

$$\log N(\mathcal{F}, e_{\mathrm{P}_n(\omega)}, \varepsilon) < \frac{c}{\varepsilon^2} \tag{5.17}$$

for all ε, n and P. Claim 1 implies

$$\sup_{f,g\in\mathcal{F}}\big|e^2_{\mathrm{P}_n(\omega)}(f,g) - e_{\mathrm{P}}(f,g)^2\big| \to 0$$

in $\mathrm{P}^{\mathbb{N}}$ probability. Then we have convergence for at least one ω along a subsequence (for each fixed P), and this and (5.17) imply

$$\sup_{\mathrm{P}\in\mathcal{P}(S)}\log N(\mathcal{F}, e_{\mathrm{P}}, \varepsilon) < \frac{c}{\varepsilon^2}, \tag{5.18}$$

proving $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded for all P, uniformly in P. In order to prove (5.16) we first symmetrize and decompose the resulting probability as follows: Using the symmetrization inequality (1.16) in the first step, we have that, for $\delta\leq 2\varepsilon$,

$$\begin{aligned}
\mathrm{P}^{\mathbb{N}}\big\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > 4\varepsilon\big\} &\leq 4\mathrm{Pr}_{\mathrm{P},\varepsilon}\Big\{\Big\|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > \varepsilon\Big\}\\
&\leq 4\mathrm{Pr}_{\mathrm{P},\varepsilon}\Big\{\Big\|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(2^{1/2}\delta,e_{\mathrm{P}_n})} > \varepsilon\Big\}\\
&\quad + \mathrm{P}^{\mathbb{N}}\Big\{\sup_{f,g\in\mathcal{F}}\big|e^2_{\mathrm{P}_n(\omega)}(f,g) - e_{\mathrm{P}}(f,g)^2\big| > \delta^2\Big\}\\
&:= I_{\mathrm{P},n} + II_{\mathrm{P},n}.
\end{aligned}$$

Now, Claim 1 ((5.12)) directly implies

$$\lim_{n\to\infty}\sup_{\mathrm{P}\in\mathcal{P}(S)} II_{\mathrm{P},n} = 0.$$

Moreover, replacing $\mathcal{H}$ by $\mathcal{H}_\delta = \{(f-g)^2 : f, g \in \mathcal{F}, e_{\mathrm{P}}(f,g) \le \delta\}$, the first line in the string of inequalities (5.15), in the proof of Claim 1, becomes

$$\mathbb{E}\Big\|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(2^{1/2}\delta, e_{\mathrm{P}_n})} \le 16\sqrt{2\pi}\sup_{Q\in\mathcal{P}_f(S)}\mathbb{E}\big\|Z_Q\big\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} \tag{5.19}$$

for all $\delta > 0$ and $\mathrm{P} \in \mathcal{P}(P)$. Then, $\mathcal{F}$ being UPG_f, we obtain

$$\lim_{n\to\infty}\sup_{\mathrm{P}\in\mathcal{P}(S)} I_{\mathrm{P},n} = 0,$$

and Claim 2 follows.

CLAIM 3. $\mathcal{F}$ is UPG in the sense that

$$\sup_{\mathrm{P}\in\mathcal{P}(S)}\mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}} < \infty, \quad \sup_{\mathrm{P}\in\mathcal{P}(S)}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}} < \infty$$

and

$$\lim_{\delta\to 0}\sup_{\mathrm{P}\in\mathcal{P}(S)}\mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} = 0 \text{ and } \lim_{\delta\to 0}\sup_{\mathrm{P}\in\mathcal{P}(S)}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} = 0.$$

Proof. Since the process $G_{\mathrm{P}} + gP$, where g is $N(0,1)$ independent of G_{P}, is a version of Z_{P} for all P, and $\|P\|_{\mathcal{F}} \le 1$ and $\|P\|_{\mathcal{F}'_{\delta,e_{\mathrm{P}}}} \le \delta$, it suffices to prove the claim for G_{P}. Claim 2 and the Portmanteau theorem give

$$\begin{aligned}&\limsup_{\delta\to 0}\sup_{\mathrm{P}\in\mathcal{P}(S)}\Pr\big\{\|G_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > \varepsilon\big\}\\ &\qquad\le \limsup_{\delta\to 0}\sup_{\mathrm{P}\in\mathcal{P}(S)}\liminf_{n\to\infty}\mathrm{P}^{\mathbb{N}}\big\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > \varepsilon\big\} = 0.\end{aligned}$$

By hypercontractivity of Gaussian processes (e.g., Ledoux and Talagrand, 1991, page 65), we have $\big(\mathbb{E}\|G_{\mathrm{P}}\|^4\big)^{1/4} \le 3^{1/2}\big(\mathbb{E}\|G_{\mathrm{P}}\|^2\big)^{1/2}$ and therefore

$$\Pr\Big\{\|G_{\mathrm{P}}\| > \lambda\big(\mathbb{E}\|G_{\mathrm{P}}\|^2\big)^{1/2}\Big\} \ge \Big(\frac{1-\lambda^2}{3}\Big)^2$$

for any $0 < \lambda < 1$ by Paley–Zygmund's argument (Kahane, 1968, page 6: just note that, by Hölder, for all $0 < a < \infty$, $\mathbb{E}\|G_{\mathrm{P}}\|^2 \le a^2 + \big(\mathbb{E}\|G_{\mathrm{P}}\|^4\big)^{1/2}\big[\Pr\{\|G_{\mathrm{P}}\| > a\}\big]^{1/2}$), where the norm refers to any $\mathcal{F}'(\delta, e_{\mathrm{P}})$ for all δ and P. The above limit and inequality then show that for all $\varepsilon > 0$ there is $\tau > 0$ such that

$$\lambda\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} < \varepsilon$$

for all $0 < \delta < \tau$ and for all $\mathrm{P} \in \mathcal{P}(S)$, that is,

$$\lim_{\delta\to 0}\sup_{\mathrm{P}\in\mathcal{P}(S)}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} = 0.$$

Now,

$$\sup_{\mathrm{P}\in\mathcal{P}(S)}\mathbb{E}\|G_{\mathrm{P}}\|_{\mathcal{F}} < \infty$$

follows from this and the fact that $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded uniformly in $\mathrm{P} \in \mathcal{P}(S)$. The claim is proved.

CLAIM 4.

$$\lim_{n\to\infty} \sup_{\mathrm{P}\in\mathcal{P}} d_{BL_{\mathrm{P}}}\big(\nu_n^{\mathrm{P}}, G_{\mathrm{P}}\big) = 0 \tag{5.6}$$

Proof. In the decomposition of $|\mathbb{E}_{\mathrm{P}}^* H(\nu_n^{\mathrm{P}}) - \mathbb{E}H(G_{\mathrm{P}})|$ into $I_{n,\tau} + II_{n,\tau} + III_\tau$ from the proofs of Theorem 1.3, Corollary 1.5 etc. for H Lipschitz in $\ell^\infty(\mathcal{F})$, the three terms can be estimated uniformly in P, the first, by claim 2, the second by Lemma 5.6, and the third by Claim 3. We omit the details in order to avoid repetition.

CLAIM 5. The map given by (5.7) is uniformly continuous.

Proof. Let $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(S)$ and $\tau > 0$. (5.18) shows that there is a universal constant $a < \infty$ and $N(\tau, \mathrm{P}, \mathrm{Q}) \le e^{a/\tau^2}$ subsets A_i of $\mathcal{F}$ that cover $\mathcal{F}$ and for each there is $f_i \in \mathcal{F}$ such that $A_i \subseteq \{f : e_{\mathrm{P}}(f, f_i) \vee e_{\mathrm{Q}}(f, f_i) \le \tau\}$. Let $H : \ell^\infty(\mathcal{F}) \to \mathbb{R}$ be bounded Lipschitz with sup and Lip norms bounded by one. As in previous proofs, define the Gaussian processes $Z_{\mathrm{P},\tau}(f) = Z_{\mathrm{P}}(f_i)$ if $f \in A_i$, and likewise for $Z_{\mathrm{Q},\tau}$. These processes are nothing but centered normal random vectors in $\mathbb{R}^{\exp(a/\tau^2)}$. Let $\Phi_{\mathrm{P},\tau}$ and $\Phi_{\mathrm{Q},\tau}$ be their covariances. Then,

$$\begin{aligned}
\mathbb{E}\big|H(Z_{\mathrm{P}}) - H(Z_{\mathrm{Q}})\big| &\le \mathbb{E}\big|H(Z_{\mathrm{P}}) - H(Z_{\mathrm{P},\tau})\big| + \mathbb{E}\big|H(Z_{\mathrm{Q}}) - H(Z_{\mathrm{Q},\tau})\big| \\
&\qquad + \mathbb{E}\big|H(Z_{\mathrm{P},\tau}) - H(Z_{\mathrm{Q},\tau})\big| \\
&\le \mathbb{E}\big\|Z_{\mathrm{P}}\big\|_{\mathcal{F}'(\tau, e_{\mathrm{P}})} + \mathbb{E}\big\|Z_{\mathrm{Q}}\big\|_{\mathcal{F}'(\tau, e_{\mathrm{Q}})} + c\big(e^{a/\tau^2}\big)\big\|\Phi_{\mathrm{P},\tau} - \Phi_{\mathrm{Q},\tau}\big\|_\infty^{1/4},
\end{aligned}$$

where in the last inequality we apply Lemma 5.7. Since

$$\big\|\Phi_{\mathrm{P},\tau} - \Phi_{\mathrm{Q},\tau}\big\|_\infty = \max_{i,j \le N(\tau,\mathrm{P},\mathrm{Q})} \big|\Phi_{\mathrm{P},\tau}(f_i, f_j) - \Phi_{\mathrm{Q},\tau}(f_i, f_j) \le \|\mathrm{P} - \mathrm{Q}\|_{\mathcal{F}\mathcal{F}},$$

uniform continuity of the map G defined by (5.7) follows from Claim 3. This concludes the proof of the theorem.

□

It is a simple exercise to show that for $\mathcal{F}$ uniformly bounded (in fact $\|\mathrm{P}f\|_{\mathcal{F}} < \infty$ suffices), $\mathcal{F} \in CLT(\mathrm{P})$ if and only if $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathrm{P}^{\mathbb{N}}\big\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta, e_{\mathrm{P}})} > \varepsilon\big\} = 0.$$

(Theorem 1.15 gives this for ρ_{P}, and the reason why it works for e_{P} instead is that if $\|\mathrm{P}f\|_{\mathcal{F}} < \infty$ then $(\mathcal{F}, e_{\mathrm{P}})$ is totally bounded if and only if $(\mathcal{F}, \rho_{\mathrm{P}})$ is.) So, the following corollary is a kind of converse to $UPG_f \Rightarrow ii)$ in Theorem 5.3.

5.11. COROLLARY. *For $\mathcal{F}$ uniformly bounded and image admissible Suslin the following are equivalent:*

a) $\mathcal{F} \in UPG_f$;

b) $(\mathcal{F}, e_{\mathrm{P}})$ *is totally bounded uniformly in* $\mathrm{P} \in \mathcal{P}(S)$ *and*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{\mathrm{P}\in\mathcal{P}(S)} \mathrm{P}^{\mathbb{N}}\big\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta, e_{\mathrm{P}})} > \varepsilon\big\} = 0$$

PROOF. a) implies b) by Claim 2 and b) implies a) by Claim 3 above.

□

Even if we weaken condition b) by not assuming uniformity in P of the total boundedness of $(\mathcal{F}, e_{\mathrm{P}})$, one still has that b) implies a) (Giné and Zinn, 1991, Theorem 2.6).

We should point out that if $\mathcal{F}$ is image admissible Suslin and satisfies condition (5.4) with the sup inside the integral, and if for some class of probability measures $\mathcal{Q}$ it also satisfies $\lim_{\lambda\to 0}\sup_{\mathrm{P}\in\mathcal{Q}}\int F^2 I_{F>\lambda} d\mathrm{P} = 0$, then $\mathcal{F}$ is P–Donsker uniformly in $\mathrm{P}\in\mathcal{Q}$ (Sheehy and Wellner, 1992).

5.12. REMARK. *(Uniform bounds for exponential moments).* The estimate (5.12) can be improved to a bound for exponential moments of the empirical process indexed by a UPG class (less is required), valid for all P and n. For a neater statement, it is convenient to recall that the expression

$$\psi(\xi) := \inf\{c : \mathbb{E}\exp(\xi^2/c^2) \le 2\}$$

is a pseudo–norm on the space of random variables (defined on, say, Ω) such that $\mathbb{E}e^{a\xi^2} < \infty$ for some $a > 0$ (this is the Luxemburg norm corresponding to the Young function $e^{x^2} - 1$). Fernique's integrability theorem for Gaussian processes (Fernique, 1974) implies that there exist $c < \infty$ such that for any centered Gaussian process $Z(t)$, $t \in T$, with bounded sample paths,

$$\psi(\|Z\|_T) \le c\mathbb{E}\|Z\|_T. \tag{5.20}$$

(where the L_1 norm can be replaced by any L_p norm). It is easy to find good estimates for c, e.g. by using concentration inequalities (Section 3.1 in Ledoux and Talagrand, 1991). Inequality (5.20) has the following consequence (Giné and Zinn, 1991):

There exists $C < \infty$ such that if $\mathcal{F}$ be an image admissible Suslin uniformly bounded class of functions satisfying

$$M_{\mathcal{F}} := \sup_{\mathrm{P}\in\mathcal{P}_f(S)} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}} < \infty, \tag{5.21}$$

then

$$\sup_{\mathrm{P}\in\mathcal{P}(S)}\sup_{n\in\mathbb{N}}\psi\big(\|\nu_n^P\|_{\mathcal{F}}\big) \le CM_{\mathcal{F}}. \tag{5.22}$$

Proof. Let $\mathrm{P}\in\mathcal{P}_f(S)$. Let P_n' be an independent copy of P_n (made of coordinates in an enlarged product space). Proceeding as in (5.13) and (5.15), on account of the facts that ψ is a pseudo–norm and that the function $e^{\lambda\|\cdot\|_{\mathcal{F}}^2}$ is convex, (5.20) and (5.21) give

$$\begin{aligned}\psi\big(\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}}\big) &\le \psi\big(\sqrt{n}(\mathrm{P}_n - \mathrm{P}_n'\|_{\mathcal{F}}\big)\\ &\le 2\psi\big(\|\sum_{i=1}^n \varepsilon_i\delta_{X_i}/\sqrt{n}\|_{\mathcal{F}}\big)\end{aligned}$$

$$\le (2\pi)^{1/2}\psi\big(\|\sum_{i=1}^n g_i\delta_{X_i}/\sqrt{n}\|_{\mathcal{F}}\big)$$
$$\le (2\pi)^{1/2} c M_{\mathcal{F}}. \tag{5.23}$$

□

Inequality (5.22) is not best possible: for really sharp exponential inequalities for empirical processes, particularly over VC–subgraph classes, the reader should consult work by Alexander, by Massart and by Talagrand (see e.g. the Ledoux–Talgrand book). However, (5.22) is very easy to prove and applies in slightly more generality than just VC.

2.6. Some remarks on applications. In this section we see some instances of application of the previous theory, particularly in connection with M–estimators. Very often the proof of a bootstrap limit result for a given sequence of statistics will follow the original proof, except that some key steps will only be possible because of the bootstrap theorems presented above. Typically, proving an a.s. bootstrap result will also involve results on the almost sure behavior of empirical processes. We will see this in two instances: for the median and for Z–estimators.

2.6.1. The bootstrap of the median. Bickel and Freedman's (1981) proof of the bootstrap clt for the median depends on KMT. In fact, nothing as deep as KMT is necessary for the bootstrap clt of the median but, certainly, deeper properties of the empirical process than just the limit theorems from Section 2.1 are needed, concretely a result on the a.s. oscillation behavior of the empirical process for the uniform distribution. See Stute (1982) for oscillations of the classical empirical process, and Alexander (1984,1985) for exponential bounds that extend Stute's result to the empirical process over VC–subgraph classes of functions. Here is a less sharp exponential bound for VC–subgraph classes, based on Remark 5.12, that also yields a result on oscillations.

6.1 THEOREM. *Let $\mathcal{F}$ be a uniformly bounded image admissible Suslin class of functions on S such that*

$$M := \sup_{\mathrm{P}\in\mathcal{P}_f(S)} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}} < \infty \tag{5.21}$$

and

$$\sup_{\mathrm{P}\in\mathcal{P}_f(S)} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} = O\big(\delta(\log\delta^{-1})^{\alpha}\big) \tag{6.1}$$

for some $\alpha \ge 0$. Then, there exist positive finite constants c_1, c_2, c_3 such that

$$\sup_{\mathrm{P}\in\mathcal{P}(S)}\sup_{n\in\mathbb{N}} \mathrm{P}^{\mathbb{N}}\Big\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > t\Big\} \le c_1\Big[\exp\Big(-c_2\frac{t^2}{\delta^2(\log\delta^{-1})^{2\alpha}}\Big) + \exp\big(-c_3\delta^4 n\big)\Big] \tag{6.2}$$

for all $t > 0$ and $\delta < (2t)\wedge(1/2)$. As a consequence, for all $0 < \tau \le 1/2$,

$$\lim_{n\to\infty} \|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'((\log\log n)^{-1/2-\tau},e_{\mathrm{P}})} = 0 \ \text{ a.s.} \tag{6.3}$$

and

$$\lim_{n\to\infty} (\log\log n)^{1/2} \|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'((\log\log n)^{-1-\tau}, e_{\mathrm{P}})} = 0 \quad \text{a.s.} \tag{6.4}$$

PROOF. As in the proof of Claim 2 within the proof of Theorem 5.3, the symmetrization inequality (1.16) gives that for $\delta < 2t$,

$$\mathrm{P}^{\mathbb{N}}\{\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})} > 4t\} \le 4\mathrm{Pr}\Big\{\Big\|\frac{1}{n^{1/2}}\sum_{i=1}^{n} \varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(2^{1/2}\delta, e_{\mathrm{P}_n})} > t\Big\} + \mathrm{P}^{\mathbb{N}}\Big\{\|\nu\|_{\mathcal{H}} > \delta^2 n^{1/2}\Big\},$$

where $\mathcal{H} = \{(f-g)^2 : f, g \in \mathcal{F}\}$. By Chebyshev and the definition of the pseudo–norm ψ of Remark 5.12,

$$\mathrm{Pr}\Big\{\Big\|\frac{1}{n^{1/2}}\sum_{i=1}^{n} \varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(2^{1/2}\delta, e_{\mathrm{P}_n})} > t\Big\} \le 2\exp\Big\{-t^2/\psi^2\big(\Big\|\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(2^{1/2}\delta,e_{\mathrm{P}_n})}/\sqrt{n}\big)\Big\}.$$

Now, the parameter M corresponding to $\mathcal{F}'(2^{1/2}\delta, e_{\mathrm{P}_n})$ by (5.21) satisfies $M = O\big(\delta(\log\delta^{-1})^{\alpha}\big)$ so that application of (5.22) for the randomized empirical process $\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}/\sqrt{n}$ indexed by this class (see (5.23)) gives

$$\mathrm{Pr}\Big\{\Big\|\frac{1}{n^{1/2}}\sum_{i=1}^{n} \varepsilon_i\delta_{X_i}\Big\|_{\mathcal{F}'(2^{1/2}\delta, e_{\mathrm{P}_n})} > t\Big\} \le c_1\exp\Big\{-c_2\frac{t^2}{\delta^2\big(\log\delta^{-1}\big)^{2\alpha}}\Big\}$$

for some fixed $0 < c_1, c_2 < \infty$ independent of P, t and δ. Assuming without loss of generality that the funcitons in the class $\mathcal{F}$ are bounded by 1, inequality (5.14), shows that the parameter $M_{\mathcal{H}}$ for $\mathcal{H}$ is $M \le 8\mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'} \le 16\mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}}$, so that, by (5.22) and Chebyshev, we have

$$\mathrm{P}^{\mathbb{N}}\Big\{\|\nu\|_{\mathcal{H}} > \delta^2 n^{1/2}\Big\} \le c_1\exp\big\{-c_3\delta^4 n\big\}.$$

These inequalities prove the oscillations bound (6.2).

Next we prove (6.3). The proof of (6.4), similar, will be omitted. It suffices to show that, for all $\varepsilon > 0$,

$$\sum_{k=3}^{\infty}\mathrm{Pr}\Big\{\max_{2^k<n\le 2^{k+1}}\|\nu_n^{\mathrm{P}}\|_{\mathcal{F}'_n} > \varepsilon\Big\} < \infty,$$

where we set $\mathcal{F}'_n := \mathcal{F}'\big((\log\log n)^{-1/2-\tau}, e_{\mathrm{P}}\big)$. We can first replace $\mathcal{F}'_n$ by the larger $\mathcal{F}'_{2^k}$ and then use Lévy for i.i.d. random vectors (Corollary 1.12) to reduce the problem to showing that

$$\sum_{k=3}^{\infty}\mathrm{Pr}\Big\{\|\nu_{2^{k+1}}^{\mathrm{P}}\|_{\mathcal{F}'_{2^k}} > \varepsilon\Big\} < \infty$$

for all $\varepsilon > 0$. Now, we apply inequality (6.2) to each summand and obtain

$$\Pr\Big\{\|\nu^{\mathrm{P}}_{2^{k+1}}\|_{\mathcal{F}'_{2^k}} > \varepsilon\Big\} \le c_1\Big[\exp\big\{-c_2\varepsilon^2(\log k)^{1+\tau}\big\} + \exp\big\{-c_3(\log k)^{-2-4\tau}2^{k+1}\big\}\Big],$$

which is the general term of a convergent series.

□

(6.2) in Theorem 6.1 comes from Giné and Zinn, 1991, and we take this opportunity to mention that their equation (2.43) is *false* as its right hand side should depend on the rate at which the expected value of the oscillations $\sup_{\mathrm{P}} \mathbb{E}\|Z_{\mathrm{P}}\|_{\mathcal{F}'(\delta,e_{\mathrm{P}})}$ tend to zero with δ. The limits (6.3) and (6.4) below were observed in Arcones and Giné (1992) as consequences of Alexander's (1984) exponential inequalities, but they follow as well from inequality (6.2), easier to prove.

Here is Bickel and Freedman's bootstrap theorem for the median:

6.2. THEOREM. *Let F be a continuous distribution function on $\mathbb{R}$ such that its derivative f exists on a neighborhood of its median m, and is continuous and positive at m. Let X_i be i.i.d.(F), let m_n be a median of the empirical ditribution F_n, and let m_n^b be a median of the bootstrap distribution function F_n^b. Then,*

$$\mathcal{L}^b\Big(n^{1/2}(m_n^b - m_n)\Big) \to_w N\Big(0, \frac{1}{4f^2(m)}\Big). \tag{6.5}$$

PROOF. Under the conditions of the theorem, m is unique, m_n is a.s. unique if n is odd, and is any point between the $n/2$-th and $(n/2+1)$–th data, e.g. the middle point, if n is even. We can also make a measurable choice for m_n^b. We must show

$$\Big\|F^b_{\sqrt{n}(m_n^b - m_n)} - F_{N(0,1/4f^2(m))}\Big\|_\infty \to 0 \ \text{ a.s.},$$

where $F^b_{\sqrt{n}(m_n^b - m_n)}$ is the conditional distribution function of the random variable $n^{1/2}(m_n^b - m_n)$ given the sample $\mathbf{X}$. By separability of $\mathbb{R}$ and Polya's lemma, it suffices to show that

$$F^b_{\sqrt{n}(m_n^b - m_n)}(x) \to F_{N(0,1/4f^2(m))}(x) \ \text{ a.s.} \tag{6.6}$$

for all $x \in \mathbb{R}$. Let us take n odd for concreteness (the proof for n even is only slightly different). Letting F_n^b denote the empirical distribution function of the bootstrap sample, we have, by definition,

$$\begin{aligned}\Pr{}^b\{\sqrt{n}(m_n^b - m_n) \le x\} &= \Pr{}^b\Big\{F_n^b(m_n + n^{-1/2}x) \ge \frac{[n/2]+1}{n}\Big\}\\ &= \Pr{}^b\Big\{\sqrt{n}(F_n^b - F_n)(m_n + n^{-1/2}x)\\ &\qquad \ge \sqrt{n}\Big(\frac{[n/2]+1}{n} - F_n(m_n + n^{-1/2}x)\Big)\Big\}.\end{aligned} \tag{6.7}$$

The class of sets $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$ is image admissible Suslin (it is parametrized by $\mathbb{R}$ and the evaluation map $(x,u) \mapsto I_{(-\infty,x]}(u)$ is jointly measurable) and it is an easy exercise to show that the covering number $N(\mathcal{C}, e_{\mathrm{P}}, \varepsilon)$ is dominated by $3/\varepsilon^2$ for any P discrete; then, by the entropy bound for Gaussian processes we have $\sup_{P\in\mathcal{P}_f(S)} \mathbb{E}\|Z_P\|_{\mathcal{C}} < \infty$ and $\sup_{P\in\mathcal{P}_f(S)} \mathbb{E}\|Z_P\|_{\mathcal{C}'(\delta,e_P)} = O(\delta\sqrt{\log\delta^{-1}})$. In conclusion, $\mathcal{C}$ is P–Donsker for all P (even uniformly in P) and, moreover, the bound (6.3) applies. Then, the a.s. bootstrap clt for empirical processes in Section 2.3 (or the simpler Corollary 5.4) applies and gives

$$\mathcal{L}^b\Big(\sqrt{n}(F_n^b - F_n)\Big) \to \mathcal{L}(G_{\mathcal{F}}) \text{ a.s. in } \ell^\infty(\mathbb{R}),$$

which, in particular, implies that

$$\mathcal{L}^b(\sqrt{n}(F_n^b - F_n)(a_n)) \to_w \mathcal{L}(G_{\mathcal{F}}(a)) \text{ a.s.}$$

whenever $a_n \to a$ a.s. and a_n, a are conditionally constant given the sample $\mathbf{X}$. Hence, since, as it is well known (e.g., Pollard, 1984, page 8), $m_n \to m$ a.s., and since by definition $G_F((-\infty, m])$ is normal with variance $F(m) - F^2(m) = 1/4$, we have

$$\mathcal{L}^b(\sqrt{n}(F_n^b - F_n)(m_n + n^{-1/2}x)) \to_w \mathcal{L}(G_{\mathcal{F}}((-\infty, m])) = N(0, 1/4) \text{ a.s.} \tag{6.8}$$

(6.8) estimates the conditional distribution of the bootstrap variable in the last term of (6.7). We now estimate the 'conditional constant' at its right. Since f is positive on a neighborhood of m, F_n has only jumps of size $1/n$ near m with probability one. Therefore, $|F_n(m_n) - 1/2| \le 1/n$ with probability one, which gives

$$\sqrt{n}\Big|\frac{[n/2]+1}{n} - F_n(m_n)\Big| \le \frac{2}{\sqrt{n}} \text{ a.s.}$$

Also, inequality (6.3) in Theorem 6.1 implies

$$\sqrt{n}((F_n - F)(m_n + n^{-1/2}x) - (F_n - F)(m_n)) \to 0 \text{ a.s.}$$

Finally, since f exists in a neighborhood of m and is continuous at m, and again using $m_n \to m$ a.s., we have

$$\sqrt{n}(F(m_n) - F(m_n + n^{-1/2}x)) \to -xf(m) \text{ a.s.}$$

by the mean value theorem. The last three observations show that

$$\tau_n := \sqrt{n}\Big(\frac{[n/2]+1}{n} - F_n\big(m_n + n^{-1/2}x\big)\Big) \to -xf(m) \text{ a.s.} \tag{6.9}$$

(6.8) and (6.9) then yield

$$\begin{aligned}\lim_{n\to\infty} \Pr{}^b\Big\{\sqrt{n}(F_n^b - F_n)\big(m_n + n^{-1/2}x\big) \ge \tau_n\Big\} &\to N(0,1/4)[-xf(m), \infty) \\ &= N(0,1/4)(-\infty, xf(m)]\end{aligned}$$

for all x, which, by (6.7), proves (6.6) and therefore the theorem.

□

It is somewhat pedantic to use Corollay 5.4 and Theorem 6.1 in the previous proof given that the bootstrap clt for $\sqrt{n}(F_n - F)$, already proved by Bickel and Freedman (1981), and the oscillation inequalities for the classical empirical process obtained by Stute (1982), do the same job. It should be pointed out, however, that the proofs of the more general theorems used here are not more difficult than the proofs of their real line analogues.

The proof of Theorem 6.2 given here is adapted from an unpublished manuscript of Sheehy and Wellner, 1988.

2.6.2. The delta method. This is treated very nicely in the book by van der Vaart and Wellner (1996). Other references on the bootstrap in connection with the delta method: Bickel and Freedman (1981), Dudley (1990) and Arcones and Giné (1992'). Regarding the delta method, it should be mentioned that recent work of Dudley tends to replace Hadamard differentiability with respect to $\|F_n - F\|_\infty$ by Fréchet differentiability with respect to $\|F_n - F\|_{p-\text{variation}}$.

2.6.3. M-estimators. M-estimators can be formally defined in two ways, either by maximization (minimization) of $P_n g(\cdot, \theta)$ for some criterion function g, or as solutions of $P_n h(\cdot, \theta) = 0$ (or almost 0). Van der Vaart and Wellner (1996) call the latter Z–estimators (Z for zero). Pollard (1985) gave an excellent treatment of the clt for M–estimators of the first type based on empirical processes. It turns out that, with Theorems 2.2, 3.2 and 6.1 at our disposal, just the natural changes on Pollard's proof give a bootstrap version of his theorem, with no surprises involved (Arcones and Giné, 1992'). For instance, the reader can recognize a proof of the original clt for the median in the proof of Theorem 6.2, and see that the changes to be made for the bootstrap are not striking at all. The bootstrap in probability version of Pollard's theorem is even more straightforward than the a.s. bootstrap version since it does not require Theorem 6.1 (and it is in a sense better because it holds under the exact same hypotheses of the original, non–bootstrap, result, whereas the bootstrap a.s. seems to require some strenghtening of the hypotheses). Z–estimators are easier to treat than M–estimators (in the strict sense) because the proof of the original clt for them is also simpler. See, however, Wellner and Zhan (1996) for the bootstrap in probability of Z estimators in an infinite dimensional setting, with applications. We illustrate these comments with a proof of the a.s. bootstrap clt for Z-estimators (Arcones and Giné, 1992').

Some notation: We will use the notation $\mathrm{Q}h(\cdot, \theta)$ to indicate $\int h(x, \theta) d\mathrm{Q}(x)$.

Let $\Theta \subset \mathbb{R}^d$ with $0 \in \Theta^\circ$, $\mathrm{P} \in \mathcal{P}(S)$, and let $h : S \times \Theta \to \mathbb{R}^d$ be a jointly measurable function. The following are our conditions for the bootstrap clt for Z–estimators (Arcones and Giné, 1992'):

(Z.1) $Ph(\cdot, 0) = 0$ and $P(h(\cdot, 0))^2 < \infty$.

(Z.2) $H(\theta) := \mathrm{P}h(\cdot, \theta)$ is "strongly" differentiable at zero with non-degerenate first derivative. Assuming (without loss of generality) $H'(0) = I$ the differentiability condition is as follows:

$$H(\theta) - H(\theta') = \theta - \theta' + o(|\theta - \theta'|) \text{ as } \theta \to 0 \text{ and } \theta' \to 0. \qquad (6.10)$$

(Z.3) If h_i, $i = 1, \ldots, d$ denote the coordinates of h, the classes of functions $\mathcal{F} = \{h_i(\cdot, \theta) - h_i(\cdot, 0) : |\theta| \le M\}$, for some $M > 0$, satisfy conditions

(5.21) and (6.1), and

$$\mathrm{P}\big[h_i(\cdot,\theta)-h_i(\cdot,0)\big]^2 \le \frac{c}{\big(\log\big|\log|\theta|\big|\big)^{1+\delta}} \tag{6.11}$$

for some $c<\infty$, $\delta>0$, all θ with $|\theta|\le M$, and $i=1,\dots,d$.

(Z.4) (Existence and consistency of the Z–estimators) There exist symmetric measurable functions $\theta(x_1,\dots,x_n)$ defined on the support of $\mathrm{P}^n, n\in N$, such that if $\theta_n:=\theta(X_1,\dots,X_n)$, then

$$\sqrt{n}\mathrm{P}_n h(\cdot,\theta_n)\to_{a.s.} 0 \ \text{ and } \ \theta_n\to_{a.s.} 0. \tag{6.12}$$

(Z.5) (existence and consistency of the bootstrap Z–estimators) For almost every $\omega\in\Omega$, there exist symmetric random variables $\theta_n^b(\omega)=\theta_n^b\big(X_{n,1}^b(\omega),\dots,X_{n,n}^b(\omega)\big)$ such that

$$\sqrt{n}\mathrm{P}_n^b h(\cdot,\theta_n^b)\to_{pr^b} 0 \ \text{ a.s. and } \ \theta_n^b-\theta_n\to_{pr^b} 0 \ \text{ a.s.} \tag{6.13}$$

The existence and consistency of the estimators (original and bootstrap) is a simpler issue and will not be treated; perhaps it should be mentioned that the conditions for consistency of Huber (1967) for case B (which is the Z case) not only give the consistency (Z.4) but also the bootstrap consistency (Z.5). [Regarding laws of large numbers, we have not proved, in these notes, that the law of large numbers uniform in $\mathcal{F}$ for empirical processes can be bootstrapped the same as the clt. This is easier than the bootstrap clt and may be taken as an exercise by the reader –alternatively, see Giné and Zinn, 1990.]

6.3. THEOREM. *Under (Z.1)-(Z.5),*

$$\lim_{n\to\infty}\mathcal{L}^b(\sqrt{n}(\theta_n^b-\theta_n))=\lim_{n\to\infty}\mathcal{L}(\sqrt{n}\theta_n)=N(0,\mathrm{Cov}_{\mathrm{P}}h(\cdot,0)) \ \text{ a.s.} \tag{6.14}$$

PROOF (Sketch). Suppose we prove

$$\sqrt{n}\Big(H\big(\theta_n^b\big)-H\big(\theta_n\big)\Big)+\sqrt{n}\big(\mathrm{P}_n^b-\mathrm{P}_n\big)h(\cdot,0)\to_{pr^b} 0 \ \text{ a.s.} \tag{6.15}$$

Then the theorem follows from the bootstrap clt for the mean in $\mathbb{R}^d$. To prove (6.15), we decompose the left side into the bootstrap and the non–bootstrap terms. Since $\mathrm{P}h(\cdot,0)=0$ and $\sqrt{n}\mathrm{P}_n h(\cdot,\theta_n)\to 0$ a.s., we can write the non–bootstrap terms as follows:

$$\sqrt{n}\Big|H(\theta_n)-\mathrm{P}_n h(\cdot,0)\Big|\simeq\sqrt{n}\Big|\big(\mathrm{P}_n-\mathrm{P}\big)\big(h(\cdot,0)-h(\cdot,\theta_n)\big)\Big| \ \text{ a.s.} \tag{6.16}$$

(as $n\to\infty$). Now, we need to show that θ_n is a.s. small so that we can apply Theorem 6.1. For this, we compare θ_n to $H(\theta_n)$: By consistency of θ_n ((6.12)) and the differentiability hypothesis (6.10) (in less than its full force), there is $c>0$ such that for all n large enough (depending on ω),

$$c\sqrt{n}|\theta_n|\le\sqrt{n}|H(\theta_n)|\le\sqrt{n}\big|(\mathrm{P}_n-\mathrm{P})h(\cdot,\theta_n)\big|+\sqrt{n}\big|\mathrm{P}_n h(\cdot,\theta_n)\big| \ \text{ a.s.}$$

Now, the last summand tends to zero a.s. by (6.12) (i.e., by definition) and the first is a.s. of the order of $(\log\log n)^{1/2}$ by the law of the iterated logarithm for empirical processes over bounded P–Donkser classes –e.g., Dudley and Philipp (1983), Theorem 4.1. Therefore,

$$|\theta_n| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{a.s. as } n \to \infty. \tag{6.17}$$

Now, (6.17) and hypothesis (Z.3) (including (6.11)) allow us to apply Theorem 6.1 and conclude

$$\sqrt{n}\Big|(\mathrm{P}_n - \mathrm{P})\big(h(\cdot,0) - h(\cdot,\theta_n)\big)\Big| \to 0 \ \text{ a.s.,}$$

so that, by (6.16), the non–bootstrap terms of the random variable at the left in (6.15) do indeed converge to zero a.s. Next, we deal with the bootstrap terms of (6.15):

$$\begin{aligned}\sqrt{n}\Big(\mathrm{P}h(\cdot,\theta_n^b) - \mathrm{P}_n^b h(\cdot,0)\Big) &\simeq \sqrt{n}(\mathrm{P}_n^b - \mathrm{P})\big(h(\cdot,\theta_n^b) - h(\cdot,0)\big)\\ &= \sqrt{n}(\mathrm{P}_n^b - \mathrm{P}_n)\big(h(\cdot,\theta_n^b) - h(\cdot,0)\big)\\ &\quad \sqrt{n}(\mathrm{P}_n - \mathrm{P})\big(h(\cdot,\theta_n^b) - h(\cdot,0)\big) \ \text{ in pr}^{\mathrm{b}}, \ \text{a.s.}\end{aligned} \tag{6.16b}$$

as $n \to \infty$ since $\mathrm{P}h(\cdot,0) = 0$ ((Z.1)) and $\sqrt{n}\mathrm{P}_n^b h(\cdot,\theta_n^b) \to 0$ in pr^b, a.s. Now, since by bootstrap consistency $\theta_n^b \to 0$ in pr^b a.s. ((6.12) and (6.13)), applying the asymptotic equicontinuity condition associated to the bootstrap clt for for each ω fixed, we obtain

$$\lim_{n\to\infty} \sqrt{n}(\mathrm{P}_n^b - \mathrm{P}_n)\big(h(\cdot,\theta_n^b) - h(\cdot,0)\big) = 0 \ \text{ in pr}^{\mathrm{b}}, \ \text{ a.s.}$$

and it is the last summand in (6.16b) that requires Theorem 6.1, as before. For this, we need to estimate the size of θ_n. Using the full force of the differentiability condition (6.10), we find that for some $c > 0$ and for almost every ω, the following inequality holds with bootstrap probability tending to 1 as $n \to \infty$:

$$\begin{aligned}c\sqrt{n}|\theta_n^b| \le \sqrt{n}\big|H(\theta_n^b)\big| \le \sqrt{n}\big|(\mathrm{P}_n^b - \mathrm{P}_n)h(\cdot,\theta_n^b)\big| + \sqrt{n}\big|(\mathrm{P}_n - \mathrm{P})h(\cdot,\theta_n^b)\big|\\ + \sqrt{n}\big|\mathrm{P}_n^b h(\cdot,\theta_n^b)\big|.\end{aligned}$$

Now, the first summand is $O_{pr^b}(1)$ a.s. by the bootstrap clt (a.s.); the last summand is $o_{pr^b}(1)$ a.s. because $\theta_n^b \to 0$ in pr^b a.s.; and the second summand is $O\Big(\sqrt{(\log\log n)}\Big)$ a.s. by the law of the iterated logarithm for empirical processes. We conclude

$$|\theta_n| = O_{pr^b}\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{a.s. as } n \to \infty. \tag{6.16b}$$

Then, as before, hypothesis (Z.3), (6.16b) and Theorem 6.1 give that

$$\sqrt{n}\big|(\mathrm{P}_n - \mathrm{P})h(\cdot,\theta_n^b)\big| \to 0 \ \text{ in pr}^{\mathrm{b}}, \ \text{ a.s.}$$

This shows that the bootstrap terms in (6.15) tend to zero in bootstrap probability a.s., concluding the proof of the theorem.

$\sqcap$

Theorem 6.3 applies to Huber's (1964) location parameters with a significant simplification of the hypotheses (Arcones and Giné, loc. cit.).

6.4. THEOREM. *Let* $h : \mathbb{R} \to \mathbb{R}$ *be a bounded monotone function and let* P *be a probability measure on* $\mathbb{R}$. *We let* $h(x, \theta) := h(x - \theta)$, $\theta \in \mathbb{R}$. *Assume:*

(H.1) Letting $H(\theta) = \mathrm{P}h(\cdot, \theta)$, *we have* $H(0) = 0, H'(0) = 1$ *and*

$$\lim_{r,s \to 0} \frac{(H(r) - H(s))}{(r - s)} = H'(0). \tag{6.18}$$

(H.2) There is a neighborhood U *of* 0 *such that for all* $\theta \in U$,

$$\mathrm{P}\big[h(\cdot, \theta) - h(\cdot, 0)\big]^2 \le \frac{c}{(\log|\log|\theta||)^{1+\delta}}$$

for some $\delta > 0$ *and* $c < \infty$.

(H.3) P *is continuous on* C_δ *for some* $\delta > 0$, *where* C_δ *denotes the open* δ *neighborhood of the set* C *of discontinuity points of* $h(x, \theta(P)) = h(x, 0)$.

Then

$$\lim_{n \to \infty} \mathcal{L}\big(\sqrt{n}(\theta_n^b - \theta_n)\big) = \lim_{n \to \infty} \mathcal{L}\big(\sqrt{n}\theta_n\big) = N(0, \mathrm{Var}_{\mathrm{P}} h) \ \text{ a.s.} \tag{6.19}$$

where

$$\theta_n = \inf\{\theta : \mathrm{P}_n h(\cdot, \theta) > 0\} \ \text{ and } \ \theta_n^b = \inf\{\theta : \mathrm{P}_n^b h(\cdot, \theta) > 0\}.$$

PROOF. By (H.1), 0 is the only zero of the function $H(\theta)$. Since $\mathrm{P}_n h(\cdot, \varepsilon) \to H(\varepsilon) > 0$ a.s. for all $\varepsilon > 0$ it follows that eventually $\theta_n \le \varepsilon$ a.s. Likewise $\theta \ge -\varepsilon$ a.s., i.e.

$$\theta_n \to 0 \ \text{ a.s.} \tag{6.20}$$

The same argument using the bootstrap law of large numbers gives

$$\theta_n^b - \theta_n \to 0 \ \text{ in } \mathrm{pr}^b \ \text{ a.s.} \tag{6.21}$$

Let $|\theta| < \delta/2$. The sample points X_i satisfying $X_i - \theta \in C_{\delta/2}$ are all a.s. different by continuity of P on C_δ. Moreover, $h(x)$ is continuous at $x = X_i - \theta$ if $X_i - \theta \notin C_{\delta/2}$. Hence the function $\mathrm{P}_n h(\cdot, \theta)$ has a jump at θ of size at most $2\|h\|_\infty / n$ a.s. This proves, by (6.20) and (6.21), that

$$\sqrt{n}\mathrm{P}_n h(\cdot, \theta_n) \to 0 \ \text{ a.s. and } \ \sqrt{n}\mathrm{P}_n^b h(\cdot, \hat{\theta}_n) \to 0 \ \text{ in } \mathrm{pr}^b \ \text{ a.s.}$$

So, conditions (Z.4) and (Z.5) are satisfied.

(H.1) is just (Z.1) and (Z.2). (H.2) is part of (Z.3). Finally the rest of (Z.3) is satisfied because of the monotonicity of h: any class of sets ordered by inclusion is Vapnik–Červonenkis and therefore satisfies conditions (5.21) and (6.1) (see Example 5.2, 1)).

□

Theorem 6.3 and its corollary Theorem 6.4 contain the bootstrap CLT for the most usual M-estimators in particular for the median, Huber's estimators, etc. For instance, Theorem 6.4 applies to

$$h(x) = -kI_{(-\infty,-k)}(x) + xI_{[-k,k]}(x) + kI_{(k,\infty)}(x)$$

under minimal conditions on P, namely that $\mathrm{P}\{k\} = \mathrm{P}\{-k\} = 0$ and $\mathrm{P}(-k,k) \neq 0$ (assuming $\mathrm{P}h(\cdot,0) = 0$).

References

DE ACOSTA, A., ARAUJO, A. AND GINÉ, E. (1978). On Poisson measures, Gaussian measures and the central limit theorem in Banach spaces. In *Probability on Banach spaces* (J. Kuelbs Ed.), Advances in Probability and Related Topics, Vol. 4, pp. 1–68.

ALEXANDER, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probability* **12** 1041-1067.

ALEXANDER, K. S. (1985). Rates of growth for weighted empirical processes. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L.M. LeCam and R. A. Olshen, eds.) **2** 475-493. Wadsworth, Monterey, CA.

ANDERSEN, N. T. (1985). *The Calculus of Non-Measurable Functions and Sets.* Various Publications Series, Aarhus Universitet, Matematisk Institut, No. 36. Aarhus.

ANDERSEN, N. T. AND DOBRIĆ, V. (1987). The central limit theorem for stochastic processes. *Ann. Probab.* **15** 164–177.

ARAUJO, A. AND GINÉ, E. (1980). *The central limit theorem for real and Banach valued random variables.* Wiley, New York.

ARCONES, M. AND GINÉ, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Ann. Inst. H. Poincaré* **25** 457-481.

ARCONES, M. AND GINÉ, E. (1991). Addition and Correction to 'The bootstrap of the mean with arbitrary bootstrap sample size'. *Ann. Inst. H. Poincaré* **27** 583–595.

ARCONES, M. AND GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674.

ARCONES, M. AND GINÉ, E. (1992'). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the limits of bootstrap* (R. LePage and L. Billard eds.) pp. 13-48. Wiley, New York.

ARCONES, M. AND GINÉ, E. (1993). Limit theorems for U–processes. *Ann. Probab.* **21** 1494–1542.

ARCONES, M. AND GINÉ, E. (1994). U–processes indexed by Vapnik–Červonenkis classes of functions with applications to asymptotics and bootstrap of U–statistics with estimated parameters. *Stochastic Proc. Appl.* **52** 17–38.

ARENAL–GUTIÉRREZ, E. AND MATRÁN, C. (1996). A zero–one law approach to the central limit theorem for the weighted bootstrap mean. *Ann. Probab.* **24** 532–540.

ATHREYA, K. B. (1985). Bootstrap of the mean in the infinite variance case, II. Tech. Report 86–21, Dept. Statistics, Iowa State University. Ames, Iowa.

BERTAIL, P. (1994). Second order properties of a corrected bootstrap without replacement, under weak assumptions. Document de Travail # 9306, CORELA et HEDM. INRA, Ivry–sur-Seine.

BICKEL, P. J. AND FREEDMAN, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196-1216.

BICKEL, P., GÖTZE, F. AND VAN ZET, W. R. (1994). Resampling fewer than n observations: gains, losses and remedies for losses. Preprint.

BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincaré* **19** 281–296.

CHEN, J. AND RUBIN, H. (1984). A note on the behavior of sample statistics when the population mean is infinite. *Ann. Probab.* **12** 256–261.

CSÖRGŐ, S. AND MASON, D. (1989). Bootstrapping empirical functions. *Ann. Statist.* **17** 1447–1471.

DAVYDOV, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.* **13** 691–696.

DEHEUVELS, P., MASON, D. AND SHORACK, G. R. (1992). Some results on the influence of extremes on the bootstrap. *Ann. Inst. H. Poincaré* 29 83–103.

DOUKHAN, P., MASSART, P. AND RIO, E. (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré* **30** 63–82.

DUDLEY, R. M. (1978). Central limit theorem for empirical measures. *Ann. Probability* **6** 899-929.

DUDLEY, R. M. (1984). A course on empirical processes. *Lecture Notes in Math.* **1097** 1-142. Springer, New York.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Lecture Notes in Math.* **1153** 141–178. Springer, Berlin. DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326.

DUDLEY, R. M. (1989). *Real Analysis and Probability.* Wadsworth, Pacific Grove, California.

DUDLEY, R. M. (1990). Nonlinear functions of empirical measures and the bootstrap. In *Probability in Banach spaces, 7* pp. 63–82. (E. Eberlein, J. Kuelbs, M. B. Marcus eds.) Birkhäuser, Boston.

DUDLEY, R. M. AND PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. v. Geb.* **62** 509–552.

FERNIQUE, X. (1975). Régularité des trajectoires des fonctions aléatoires gaussiennes. Ecole d'Été de Probabilités de Saint Flour, 1974. *Lecture Notes in Math.* **480** 1–96. 1995, Springer, Berlin.

FILIPPOVA, A. A. (1961). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7** 24–57.

GAENSSLER, P. (1987). Bootstrapping empirical measures indexed by Vapnik-Červonenkis classes of sets. In *Probability theory and Math. Statist.* **1** 467-481. VNU Science Press, Utrecht.

GINÉ, E. (1996). Empirical processes and applications: an overview. *Bernoulli* **2** 1–28.

GINÉ, E. AND ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probability* **12** 929-989.

GINÉ, E. AND ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes in Math.* **1221** 50-113. Springer, Berlin.

GINÉ, E. AND ZINN, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17** 684–691.

GINÉ, E. AND ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probability* **18** 851-869.

GINÉ, E. AND ZINN, J. (1991). Gaussian characterization of uniform Donsker clases of functions. *Ann. Probability* **19** 758-782.

GINÉ, E. AND ZINN, J. (1992). Marcinkiewicz type laws of large numbers and convergence of moments for U–statistics. In *Probability in Banach Spaces* **8** 273–291.

GÖTZE, F. (1993). Abstract in *Bulletin of the IMS.*

HALL, P. (1988). Rate of convergence in bootstrap approximations. *Ann. Probab.* **16** 1665–1684.

HALL, P. (1990). Asymptotic properties of the bootstrap for heavy–tailed distributions. *Ann. Probab.* **18** 1342–1360.

HOEFFDING, W. (1948). A non–parametric test of independence. *Ann. Math. Statist.* **19** 546–557.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOFFMANN-JØRGENSEN, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–189.

HOFFMANN-JØRGENSEN, J. (1984). *Stochastics processes on Polish spaces*, Aarhus Universitet, Matematisk Inst., Various Publication Series No. 39, 1991. Aarhus.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 221-233. University of California Press, Berkeley.

HUSKOVÁ, M. AND JANSSEN, P. (1993). Consistency for the generalized bootstrap for degenerate U–statistics. *Ann. Statist.* **21** 1811–1823.

IBRAGIMOV, I. A. AND LINNIK, Y.V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.

KAHANE, J. P. (1968). *Some random series of functions.* Heath, Lexington, Massachusetts.

KESTEN, H. (1971). Sums of random variables with infinite expectation. Solution to Advanced Problem # 5716. *Amer. Math. Monthly* **78** 305–308.

KLAASSEN, C. A. J. AND WELLNER, J. A. (1992). Kac empirical processes and the bootstrap. In *Probability in Banach Spaces* **8** 411–429. (R. Dudley, M. Hahn, J. Kuelbs eds.) Birkhäuser, Boston.

KUNSCH, H. (1989). The jacknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

KWAPIEŃ , S. AND WOYCZYNSKI, W. (1992). *Random series and stochastic integrals: Single and multiple.* Birkhäuser, Basel and Boston.

LE CAM, L. (1970). Remarques sur le théorème limite central dans les espaces localement convexes. A *Les Probabilités sur les Structures Algébriques*, Clermont-Ferrand 1969. Colloque CNRS, Paris, 233-249.

LEDOUX, M. AND TALAGRAND, M. (1986). Conditions d'integrabilité pour les multiplicateurs dans le TLC banachique. *Ann. Probability* **14** 916-921.

LEDOUX, M. AND TALAGRAND, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Prob. Theory Rel. Fields* **77** 29-47.

LEDOUX, M. AND TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer-Verlag, New York.

LIU, R. AND SINGH, K. (1992). Moving blocks jacknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap* pp. 225–248. (R. LePage and L. Billard eds.) Wiley, New York.

MASON, D. AND NEWTON, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* **20** 1611–1624.

MONTGOMERY–SMITH, S. J. (1994). Comparison of sums of independent identically dsitributed random variables. *Prob. Math. Statist.* **14** 281–285.

O'BRIEN, G. L. (1980). A limit theorem for sample maxima and heavy branches in Galton–Watson trees. *J. Appl. Probab.* **17** 539–545.

PELIGRAD, M. (1996). On the blockwise bootstrap for empirical processes for stationary sequences. Preprint.

PISIER, G. AND ZINN, J. (1978). On the limit theorems for random variables with values in the spaces L_p $(2 \leq p < \infty)$. *Z. Wahrsch. v. Geb.* **41** 289–304.

POLITIS, D. N. AND ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.

POLLARD, D. (1984). *Convergence of stochastic processes.* Springer, Berlin.

POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1** 295-314.

POLLARD, D. (1990). *Empirical processes: Theory and Applications.* NSF–CBMS Regional Conference Series in Probability and Statistics, 2. IMS and ASA.

PRÆSTGAARD, J. AND WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probability* **21** 2053–2086.

RADULOVIĆ, D. (1996). The bootstrap of the mean for strong mixing sequences under minimal conditions. *Statist. and Probab. Lett.* **28** 65–72.

RADULOVIĆ, D. (1996b). On the bootstrap for empirical processes based on stationary observations. To appear in *Stochastic Proc. Appl.*

RADULOVIĆ, D. (1996c). Private communications.

ROMANO, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17** 141–159.

SEN, P. K. (1974). On L_p–convergence of U–statistics. *Ann. Inst. Statist. Math.* **26** 55–60.

SEPANSKI, S. J. (1993). Almost sure bootstrap of the mean under random normalization. *Ann. Probab.* **21** 917–925.

SHEEHY, A. AND WELLNER, J. A. (1988). Uniformity in P of some limit theorems for empirical measures and processes. Tech. Rep. # 134, Department of Statistics, University of Washington, Seattle, Washington.

SHEEHY, A. AND WELLNER, J. A. (1988'). Almost sure consistency of the bootstrap for the median. Unpublished manuscript.

SHEEHY, A. AND WELLNER, J. A. (1992). Uniform Donsker classes of functions. *Ann. Probability* **20** 1983-2030.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

STOUT, W. F. (1974). *Almost sure convergence.* Cademic Press, New York.

STROBL, F. (1994). *Zur Theorie empirischer Prozesse.* Doctoral Dissertation, University of Munich, Faculty of Mathematics.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

VAN DER VAART, A. AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, Berlin.

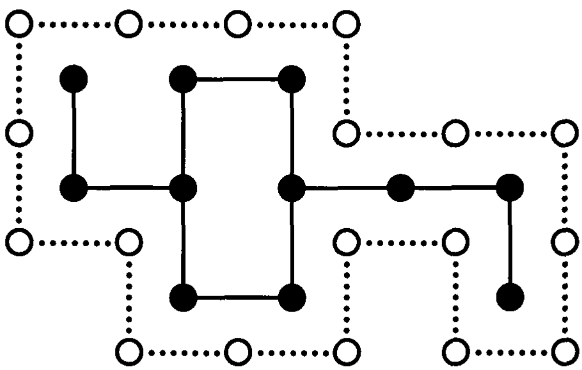
WELLNER, J. A. AND ZHAN, Y. (1996). Bootstrapping Z–estimators. Preprint.

YURINSKII, Y. Y. (1974). Exponential bounds for large deviations. *Theor. Probab. Appl.* **19** 154–155.

ZIEGLER, K. (1994). *On Functional Central Limit Theorems and Uniform Laws of Large Numbers for Sums of Independent Processes.* Doctoral Dissertation, University of Munich, Faculty of Mathematics.

University of Connecticut
Department of Mathematics and
Department of Statistics
Storrs, CT 06269, USA

PERCOLATION AND DISORDERED SYSTEMS



Geoffrey GRIMMETT

PREFACE

This course aims to be a (nearly) self-contained account of part of the mathematical theory of percolation and related topics. The first nine chapters summarise rigorous results in percolation theory, with special emphasis on results obtained since the publication of my book [156] entitled 'Percolation', and sometimes referred to simply as [G] in these notes. Following this core material are chapters on random walks in random labyrinths, and fractal percolation. The final two chapters include material on Ising, Potts, and random-cluster models, and concentrate on a 'percolative' approach to the associated phase transitions.

The first target of this course is to draw a picture of the mathematics of percolation, together with its immediate mathematical relations. Another target is to present and summarise recent progress. There is a considerable overlap between the first nine chapters and the contents of the principal reference [G]. On the other hand, the current notes are more concise than [G], and include some important extensions, such as material concerning strict inequalities for critical probabilities, the uniqueness of the infinite cluster, the triangle condition and lace expansion in high dimensions, together with material concerning percolation in slabs, and conformal invariance in two dimensions. The present account differs from that of [G] in numerous minor ways also. It does not claim to be comprehensive. A second edition of [G] is planned, containing further material based in part on the current notes.

A special feature is the bibliography, which is a fairly full list of papers published in recent years on percolation and related mathematical phenomena. The compilation of the list was greatly facilitated by the kind responses of many individuals to my request for lists of publications.

Many people have commented on versions of these notes, the bulk of which have been typed so superbly by Sarah Shea-Simonds. I thank all those who have contributed, and acknowledge particularly the suggestions of Ken Alexander, Carol Bezuidenhout, Philipp Hiemer, Anthony Quas, and Alan Stacey, some of whom are mentioned at appropriate points in the text. In addition, these notes have benefited from the critical observations of various members of the audience at St Flour.

Members of the 1996 summer school were treated to a guided tour of the library of the former seminary of St Flour. We were pleased to find there a copy of the *Encyclopédie, ou Dictionnaire Raisonné des Sciences, des Arts et des Métiers*, compiled by Diderot and D'Alembert, and published in Geneva around 1778. Of the many illuminating entries in this substantial work, the following definition of a probabilist was not overlooked.

PROBABILISTE, s. m. (*Gram. Théol.*) celui qui tient pour la doctrine abominable des opinions rendues probables par la décision d'un casuiste, & qui assure l'innocence de l'action faite en conséquence. Pascal a foudroyé ce systême, qui ouvroit la porte au crime en accordant à l'autorité les prérogatives de la certitude, à l'opinion & la sécurité qui n'appartient qu'à la bonne conscience.

This work was aided by partial financial support from the European Union under contract CHRX-CT93-0411, and from the Engineering and Physical Sciences Research Council under grant GR/L15425.

Geoffrey R. Grimmett
Statistical Laboratory
University of Cambridge
16 Mill Lane
Cambridge CB2 1SB
United Kingdom.

October 1996

CONTENTS

1.	**Introductory Remarks**	**158**
1.1	Percolation	158
1.2	Some possible questions	158
1.3	History	159
2.	**Notation and Definitions**	**161**
2.1	Graph terminology	161
2.2	Probability	162
2.3	Geometry	162
2.4	A partial order	164
2.5	Site percolation	165
3.	**Phase Transition**	**166**
3.1	Percolation probability	166
3.2	Existence of phase transition	166
3.3	A question	168
4.	**Inequalities for Critical Probabilities**	**169**
4.1	Russo's formula	169
4.2	Strict inequalities for critical probabilities	171
4.3	The square and triangular lattices	172
4.4	Enhancements	176
5.	**Correlation Inequalities**	**178**
5.1	FKG inequality	178
5.2	Disjoint occurrence	181
5.3	Site and bond percolation	182
6.	**Subcritical Percolation**	**185**
6.1	Using subadditivity	185
6.2	Exponential decay	188
6.3	Ornstein–Zernike decay	198
7.	**Supercritical Percolation**	**199**
7.1	Uniqueness of the infinite cluster	199
7.2	Percolation in slabs	201
7.3	Limit of slab critical points	202
7.4	Percolation in half-spaces	215
7.5	Percolation probability	216
7.6	Cluster-size distribution	218

8. Critical Percolation **219**
8.1 Percolation probability 219
8.2 Critical exponents 219
8.3 Scaling theory 219
8.4 Rigorous results 221
8.5 Mean-field theory 221

9. Percolation in Two Dimensions **229**
9.1 The critical probability is $\frac{1}{2}$ 229
9.2 RSW technology 230
9.3 Conformal invariance 233

10. Random Walks in Random Labyrinths **236**
10.1 Random walk on the infinite percolation cluster 236
10.2 Random walks in two-dimensional labyrinths 240
10.3 General labyrinths 248

11. Fractal Percolation **254**
11.1 Random fractals 254
11.2 Percolation 255
11.3 A morphology 258
11.4 Relationship to Brownian Motion 260

12. Ising and Potts Models **261**
12.1 Ising model for ferromagnets 261
12.2 Potts models 262
12.3 Random-cluster models 263

13. Random-Cluster Models **265**
13.1 Basic properties 265
13.2 Weak limits and phase transitions 266
13.3 First and second order transitions 268
13.4 Exponential decay in the subcritical phase 269
13.5 The case of two dimensions 273

References **281**

1. INTRODUCTORY REMARKS

1.1. Percolation

We will focus our ideas on a specific percolation process, namely 'bond percolation on the cubic lattice', defined as follows. Let $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ be the hypercubic lattice in d dimensions, where $d \geq 2$. Each edge of $\mathbb{L}^d$ is declared *open* with probability p, and *closed* otherwise. Different edges are given independent designations. We think of an open edge as being open to the transmission of disease, or to the passage of water. Now concentrate on the set of open edges, a random set. Percolation theory is concerned with ascertaining properties of this set.

The following question is considered central. If water is supplied at the origin, and flows along the open edges only, can it reach infinitely many vertices with strictly positive probability? It turns out that the answer is *no* for small p, and *yes* for large p. There is a *critical probability* p_c dividing these two phases. Percolation theory is particularly concerned with understanding the geometry of open edges in the *subcritical phase* (when $p < p_c$), the *supercritical phase* (when $p > p_c$), and when p is near or equal to p_c (the *critical* case).

As an illustration of the concrete problems of percolation, consider the function $\theta(p)$, defined as the probability that the origin lies in an infinite cluster of open edges (this is the probability referred to above, in the discussion of p_c). It is believed that θ has the general appearance sketched in Figure 1.1.

- θ should be smooth on $(p_c, 1)$. It is known to be infinitely differentiable, but there is no proof known that it is real analytic for all d.
- Presumably θ is continuous at p_c. No proof is known which is valid for all d.
- Perhaps θ is concave on $(p_c, 1]$, or at least on $(p_c, p_c + \delta)$ for some positive δ.
- As $p \downarrow p_c$, perhaps $\theta(p) \sim a(p - p_c)^\beta$ for some constant a and some 'critical exponent' β.

We stress that, although each of the points raised above is unproved in general, there are special arguments which answer some of them when either $d = 2$ or d is sufficiently large. The case $d = 3$ is a good one on which to concentrate.

1.2 Some Possible Questions

Here are some apparently reasonable questions, some of which turn out to be feasible.

- What is the value of p_c?
- What are the structures of the subcritical and supercritical phases?
- What happens when p is near to p_c?
- Are there other points of phase transition?
- What are the properties of other 'macroscopic' quantities, such as the mean size of the open cluster containing the origin?
- What is the relevance of the choice of dimension or lattice?
- In what ways are the large-scale properties different if the states of nearby edges are allowed to be dependent rather than independent?

There is a variety of reasons for the explosion of interest in the percolation model, and we mention next a few of these.

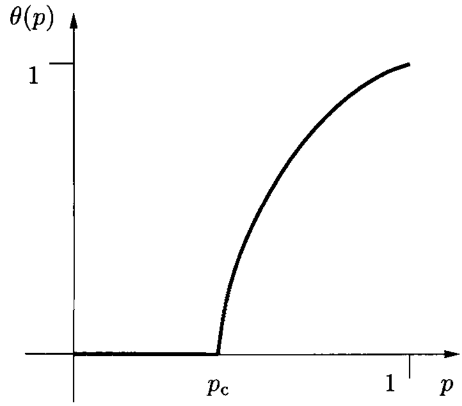


Fig. 1.1. It is generally believed that the percolation probability $\theta(p)$ behaves roughly as indicated here. It is known, for example, that θ is infinitely differentiable except at the critical point p_c. The possibility of a jump discontinuity at p_c has not been ruled out when $d \geq 3$ but d is not too large.

- The problems are simple and elegant to state, and apparently hard to solve.
- Their solutions require a mixture of new ideas, from analysis, geometry, and discrete mathematics.
- Physical intuition has provided a bunch of beautiful conjectures.
- Techniques developed for percolation have applications to other more complicated spatial random processes, such as epidemic models.
- Percolation gives insight and method for understanding other physical models of spatial interaction, such as Ising and Potts models.
- Percolation provides a 'simple' model for porous bodies and other 'transport' problems.

The rate of publication of papers on percolation and its ramifications is very high in the physics journals, although substantial mathematical contributions are rare. The depth of the 'culture chasm' is such that few (if anyone) can honestly boast to understand all the major mathematical and physical ideas which have contributed to the subject.

1.3 History

In 1957, Simon Broadbent and John Hammersley [81] presented a model for a disordered porous medium which they called the *percolation model*. Their motivation was perhaps to understand *flow* through a discrete disordered system, such as particles flowing through the filter of a gas mask, or fluid seeping through the interstices of a porous stone. They proved in [81, 174, 175] that the percolation model has a phase transition, and they developed some technology for studying the two phases of the process.

These early papers were followed swiftly by a small number of high quality articles by others, particularly [138, 181, 339], but interest flagged for a period beginning around 1964. Despite certain appearances to the contrary, some individuals realised that a certain famous conjecture remained unproven, namely that the critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. Fundamental rigorous progress towards this conjecture was made around 1976 by Russo [326] and Seymour and Welsh [338], and the conjecture was finally resolved in a famous paper of Kesten [200]. This was achieved by a development of a sophisticated mechanism for studying percolation in two dimensions, relying in part on path-intersection properties which are not valid in higher dimensions. This mechanism was laid out more fully by Kesten in his monograph [202].

Percolation became a subject of vigorous research by mathematicians and physicists, each group working in its own vernacular. The decade beginning in 1980 saw the rigorous resolution of many substantial difficulties, and the formulation of concrete hypotheses concerning the nature of phase transition.

The principal progress was on three fronts. Initially mathematicians concentrated on the 'subcritical phase', when the density p of open edges satisfies $p < p_c$ (here and later, p_c denotes the critical probability). It was in this context that the correct generalisation of Kesten's theorem was discovered, valid for all dimensions (i.e., two or more). This was achieved independently by Aizenman and Barsky [13] and Menshikov [268, 269].

The second front concerned the 'supercritical phase', when $p > p_c$. The key question here was resolved by Grimmett and Marstrand [165] following work of Barsky, Grimmett, and Newman [50].

The critical case, when p is near or equal to the critical probability p_c, remains largely unresolved by mathematicians (except when d is sufficiently large). Progress has certainly been made, but we seem far from understanding the beautiful picture of the phase transition, involving scaling theory and renormalisation, which is displayed before us by physicists. This multifaceted physical image is widely accepted as an accurate picture of events when p is near to p_c, but its mathematical verification is an open challenge of the first order.

2. NOTATION AND DEFINITIONS

2.1 Graph Terminology

We shall follow the notation of [G] whenever possible (we refer to [156] as [G]). The number of dimensions is d, and we assume throughout that $d \geq 2$. We write $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ for the integers, and $\mathbb{Z}^d$ for the set of all vectors $x = (x_1, x_2, \dots, x_d)$ of integers. For $x \in \mathbb{Z}^d$, we generally denote by x_i the ith coordinate of x. We use two norms on $\mathbb{Z}^d$, namely

$$|x| = \sum_{i=1}^{d} |x_i|, \quad \|x\| = \max\{|x_i| : 1 \leq i \leq d\}, \tag{2.1}$$

and note that

$$\|x\| \leq |x| \leq d\|x\|. \tag{2.2}$$

We write

$$\delta(x, y) = |y - x|. \tag{2.3}$$

Next we turn $\mathbb{Z}^d$ into a graph, called the d-dimensional cubic lattice, by adding edges $\langle x, y\rangle$ between all pairs $x, y \in \mathbb{Z}^d$ with $\delta(x, y) = 1$. This lattice is denoted $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$. We use the usual language of graph theory. Vertices x, y with $\delta(x, y) = 1$ are called *adjacent*, and an edge e is *incident* to a vertex x if x is an endpoint of e. We write $x \sim y$ if x and y are adjacent, and we write $\langle x, y\rangle$ for the corresponding edge. The origin of $\mathbb{L}^d$ is written as the zero vector 0, and e_1 denotes the unit vector $e_1 = (1, 0, 0, \dots, 0)$.

A *path* of $\mathbb{L}^d$ is an alternating sequence $x_0, e_0, x_1, e_1, \dots$ of distinct vertices x_i and edges $e_i = \langle x_i, x_{i+1}\rangle$. If the path terminates at some vertex x_n, it is said to connect x_0 to x_n, and to have length n. If the path is infinite, it is said to connect x_0 to ∞. A *circuit* of $\mathbb{L}^d$ is an alternating sequence $x_0, e_0, x_1, e_1, \dots, e_{n-1}, x_n, e_n, x_0$ such that $x_0, e_0, \dots, e_{n-1}, x_n$ is a path and $e_n = \langle x_n, x_0\rangle$; such a circuit has length $n+1$. Two subgraphs of $\mathbb{L}^d$ are called *edge-disjoint* if they have no edges in common, and *disjoint* if they have no vertices in common.

A *box* is a subset of $\mathbb{Z}^d$ of the form

$$B(a, b) = \prod_{i=1}^{d} [a_i, b_i] \qquad \text{for } a, b \in \mathbb{Z}^d$$

where $[a_i, b_i]$ is interpreted as $[a_i, b_i] \cap \mathbb{Z}$ and it is assumed that $a_i \leq b_i$ for all i. Such a box $B(a, b)$ may be turned into a graph by the addition of all relevant edges from $\mathbb{L}^d$. A useful expanding sequence of boxes is given by

$$B(n) = [-n, n]^d = \{x \in \mathbb{Z}^d : \|x\| \leq n\}.$$

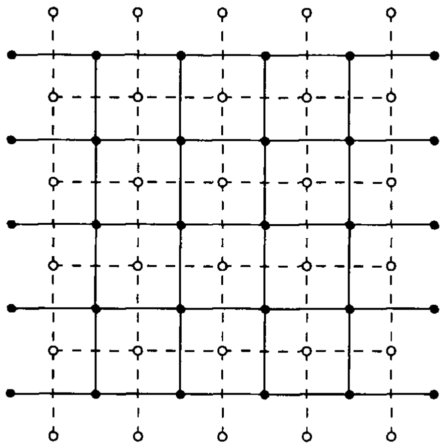


Fig. 2.1. Part of the square lattice $\mathbb{L}^2$ and its dual.

The case of two-dimensional percolation turns out to have a special property, namely that of *duality*. Planar duality arises as follows. Let G be a planar graph, drawn in the plane. The *planar dual* of G is the graph constructed in the following way. We place a vertex in every face of G (including the infinite face if it exists) and we join two such vertices by an edge if and only if the corresponding faces of G share a boundary edge. It is easy to see that the dual of the square lattice $\mathbb{L}^2$ is a copy of $\mathbb{L}^2$, and we refer therefore to the square lattice as being *self-dual*. See Figure 2.1.

2.2 Probability

Let p and q satisfy $0 \leq p = 1 - q \leq 1$. We declare each edge of $\mathbb{L}^d$ to be *open* with probability p, and closed otherwise, different edges having independent designations. The appropriate sample space is the set $\Omega = \{0,1\}^{\mathbb{E}^d}$, points of which are represented as $\omega = (\omega(e) : e \in \mathbb{E}^d)$ called *configurations*. The value $\omega(e) = 1$ corresponds to e being open, and $\omega(e) = 0$ to e being closed. Our σ-field $\mathcal{F}$ is that generated by the finite-dimensional cylinders of Ω, and the probability measure is product measure P_p having density p. In summary, our probability space is $(\Omega, \mathcal{F}, P_p)$, and we write E_p for the expectation operator corresponding to P_p.

2.3 Geometry

Percolation theory is concerned with the study of the geometry of the set of open edges, and particularly with the question of whether or not there is an infinite cluster of open edges.

Let $\omega \in \Omega$ be a configuration. Consider the graph having $\mathbb{Z}^d$ as vertex set, and as edge set the set of open edges. The connected components of this graph are called *open clusters*. We write $C(x)$ for the open cluster containing the vertex x, and call $C(x)$ the *open cluster at x*. Using the translation-invariance of P_p, we see that the distribution of $C(x)$ is the same as that of the open cluster $C = C(0)$ at the origin.

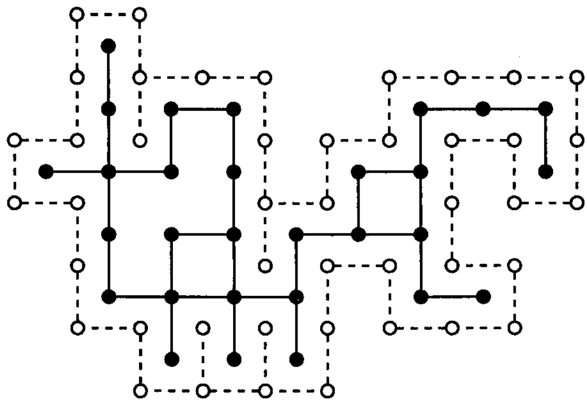


Fig. 2.2. An open cluster, surrounded by a closed circuit in the dual.

We shall be interested in the *size* of a cluster $C(x)$, and write $|C(x)|$ for the number of vertices in $C(x)$.

If A and B are sets of vertices, we write '$A \leftrightarrow B$' if there is an open path (i.e., a path all of whose edges are open) joining some member of A to some member of B. The negation of such a statement is written '$A \not\leftrightarrow B$'. We write '$A \leftrightarrow \infty$' to mean that some vertex in A lies in an infinite open path. Also, for a set D of vertices (resp. edges), '$A \leftrightarrow B$ off D' means that there is an open path joining A to B using no vertex (resp. edge) in D.

We return briefly to the discussion of graphical duality at the end of Section 2.1. Recall that $\mathbb{L}^2$ is self-dual. For the sake of definiteness, we take as vertices of this dual lattice the set $\{x + (\frac{1}{2}, \frac{1}{2}) : x \in \mathbb{Z}^2\}$ and we join two such neighbouring vertices by a straight line segment of $\mathbb{R}^2$. There is a one-one correspondence between the edges of $\mathbb{L}^2$ and the edges of the dual, since each edge of $\mathbb{L}^2$ is crossed by a unique edge of the dual. We declare an edge of the dual to be open or closed depending respectively on whether it crosses an open or closed edge of $\mathbb{L}^2$. This assignment gives rise to a bond percolation process on the dual lattice with the same edge-probability p.

Suppose now that the open cluster at the origin of $\mathbb{L}^2$ is finite, and see Figure 2.2 for a sketch of the situation. We see that the origin is surrounded by a necklace of closed edges which are blocking off all possible routes from the origin to infinity. We may satisfy ourselves that the corresponding edges of the dual contain a closed circuit in the dual which contains the origin of $\mathbb{L}^2$ in its interior. This is best seen by inspecting Figure 2.2 again. It is somewhat tedious to formulate and prove such a statement with complete rigour, and we shall not do so here; see [202, p. 386] for a more careful treatment. The converse holds similarly: if the origin is in the interior of a closed circuit of the dual lattice, then the open cluster at the origin is finite. We summarise these remarks by saying that $|C| < \infty$ if and only if the origin of $\mathbb{L}^2$ is in the interior of a closed circuit of the dual.

2.4 A Partial Order

There is a natural partial order on Ω, namely $\omega_1 \leq \omega_2$ if and only if $\omega_1(e) \leq \omega_2(e)$ for all e. This partial order allows us to discuss orderings of probability measures on $(\Omega, \mathcal{F})$. We call a random variable X on $(\Omega, \mathcal{F})$ *increasing* if

$$X(\omega_1) \leq X(\omega_2) \quad \text{whenever} \quad \omega_1 \leq \omega_2,$$

and *decreasing* if $-X$ is increasing. We call an event A (i.e., a set in $\mathcal{F}$) *increasing* (resp. *decreasing*) if its indicator function 1_A, given by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$

is increasing (resp. decreasing).

Given two probability measures μ_1 and μ_2 on $(\Omega, \mathcal{F})$ we say that μ_1 *dominates* μ_2, written $\mu_1 \geq \mu_2$, if $\mu_1(A) \geq \mu_2(A)$ for all increasing events A. Using this partial order on measures, it may easily be seen that the probability measure P_p is non-decreasing in p, which is to say that

$$P_{p_1} \geq P_{p_2} \quad \text{if} \quad p_1 \geq p_2. \tag{2.4}$$

General sufficient conditions for such an inequality have been provided by Holley [192] and others (see Holley's inequality, Theorem 5.5), but there is a simple direct proof in the case of product measures. It makes use of the following elementary device.

Let $\big(X(e) : e \in \mathbb{E}^d\big)$ be a family of independent random variables each being uniformly distributed on the interval $[0,1]$, and write P_p for the associated (product) measure on $[0,1]^{\mathbb{E}^d}$. For $0 \leq p \leq 1$, define the random variable $\eta_p = \big(\eta_p(e) : e \in \mathbb{E}^d\big)$ by

$$\eta_p(e) = \begin{cases} 1 & \text{if } X(e) < p, \\ 0 & \text{if } X(e) \geq p. \end{cases}$$

It is clear that:

(a) the vector η_p has distribution given by P_p,

(b) if $p_1 \geq p_2$ then $\eta_{p_1} \geq \eta_{p_2}$.

Let A be an increasing event, and $p_1 \geq p_2$. Then

$$\begin{aligned} P_{p_1}(A) &= P(\eta_{p_1} \in A) \geq P(\eta_{p_2} \in A) \quad \text{since } \eta_{p_1} \geq \eta_{p_2} \\ &= P_{p_2}(A), \end{aligned}$$

whence $P_{p_1} \geq P_{p_2}$.

2.5 Site Percolation

In bond percolation, it is the *edges* which are designated open or closed; in site percolation, it is the *vertices*. In a sense, site percolation is more general than bond percolation, since a bond model on a lattice $\mathcal{L}$ may be transformed into a site model on its 'line' (or 'covering') lattice $\mathcal{L}'$ (obtained from $\mathcal{L}$ by placing a vertex in the middle of each edge, and calling two such vertices adjacent whenever the corresponding edges of $\mathcal{L}$ share an endvertex). See [138]. In practice, it matters little whether we choose to work with site or bond percolation, since sufficiently many methods work equally well for both models.

In a more general 'hypergraph' model, we are provided with a hypergraph on the vertex set $\mathbb{Z}^d$, and we declare each hyperedge to be open with probability p. We then study the existence of infinite paths in the ensuing open hypergraph.

We shall see that a percolation model necessarily has a 'critical probability' p_c. Included in Section 5.3 is some information about the relationship between the critical probabilities of site and bond models on a general graph G.

3. PHASE TRANSITION

3.1 Percolation Probability

One of the principal objects of study is the *percolation probability*

$$\theta(p) = P_p(0 \leftrightarrow \infty), \tag{3.1}$$

or alternatively $\theta(p) = P_p(|C| = \infty)$ where $C = C(0)$ is, as usual, the open cluster at the origin. The event $\{0 \leftrightarrow \infty\}$ is increasing, and therefore θ is non-decreasing (using (2.4)), and it is natural to define the *critical probability* $p_c = p_c(\mathbb{L}^d)$ by

$$p_c = \sup\{p : \theta(p) = 0\}.$$

See Figure 1.1 for a sketch of the function θ.

3.2 Existence of Phase Transition

It is easy to show that $p_c(\mathbb{L}) = 1$, and therefore the case $d = 1$ is of limited interest from this point of view.

Theorem 3.2. *If $d \geq 2$ then $0 < p_c(\mathbb{L}^d) < 1$.*

Actually we shall prove that

$$\frac{1}{\mu(d)} \leq p_c(\mathbb{L}^d) \leq 1 - \frac{1}{\mu(2)} \quad \text{for } d \geq 2 \tag{3.3}$$

where $\mu(d)$ is the connective constant of $\mathbb{L}^d$.

Proof. Since $\mathbb{L}^d$ may be embedded in $\mathbb{L}^{d+1}$, it is 'obvious' that $p_c(\mathbb{L}^d)$ is non-increasing in d (actually it is *strictly* decreasing). Therefore we need only to show that

$$p_c(\mathbb{L}^d) > 0 \qquad \text{for all } d \geq 2, \tag{3.4}$$

$$p_c(\mathbb{L}^2) < 1. \tag{3.5}$$

The proof of (3.4) is by a standard 'path counting' argument. Let $N(n)$ be the number of open paths of length n starting at the origin. The number of such paths cannot exceed a theoretical upper bound of $2d(2d-1)^{n-1}$. Therefore

$$\begin{aligned}\theta(p) &\leq P_p\big(N(n) \geq 1\big) \leq E_p\big(N(n)\big)\\ &\leq 2d(2d-1)^{n-1}p^n\end{aligned}$$

which tends to 0 as $n \to \infty$ if $p < (2d-1)^{-1}$. Hence $p_c(\mathbb{L}^d) \geq (2d-1)^{-1}$. By estimating $N(n)$ more carefully, this lower bound may be improved to

$$p_c(\mathbb{L}^d) \geq \mu(d)^{-1}. \tag{3.6}$$

We use a 'Peierls argument' to obtain (3.5). Let $M(n)$ be the number of closed circuits of the dual, having length n and containing 0 in their interior. Note that $|C| < \infty$ if and only if $M(n) \geq 1$ for some n. Therefore

$$
\begin{aligned}
(3.7) \qquad 1 - \theta(p) = P_p(|C| < \infty) &= P_p\Big(\sum_n M(n) \geq 1\Big) \\
&\leq E_p\Big(\sum_n M(n)\Big) = \sum_{n=4}^{\infty} E_p(M(n)) \\
&\leq \sum_{n=4}^{\infty} (n4^n)(1-p)^n,
\end{aligned}
$$

where we have used the facts that the shortest dual circuit containing 0 has length 4, and that the total number of dual circuits, having length n and surrounding the origin, is no greater than $n4^n$. The final sum may be made strictly smaller than 1 by choosing p sufficiently close to 1, say $p > 1 - \epsilon$ where $\epsilon > 0$. This implies that $p_c(\mathbb{L}^2) < 1 - \epsilon$.

This upper bound may be improved to obtain $p_c(\mathbb{L}^2) \leq 1 - \mu(2)^{-1}$. Here is a sketch. Let F_m be the event that there exists a closed dual circuit containing the box $B(m)$ in its interior, and let G_m be the event that all edges of $B(m)$ are open. These two events are independent, since they are defined in terms of disjoint sets of edges. Now,

$$
P_p(F_m) \leq P_p\Big(\sum_{n=4m}^{\infty} M(n) \geq 1\Big) \leq \sum_{n=4m}^{\infty} n a_n (1-p)^n
$$

where a_n is the number of paths of $\mathbb{L}^2$ starting at the origin and having length n. It is the case that $n^{-1} \log a_n \to \log \mu(2)$ as $n \to \infty$. If $1 - p < \mu(2)^{-1}$, we may find m such that $P_p(F_m) < \frac{1}{2}$. However,

$$
\theta(p) \geq P_p(\overline{F_m} \cap G_m) = P_p(\overline{F_m}) P_p(G_m) \geq \tfrac{1}{2} P_p(G_m) > 0
$$

if $1 - p < \mu(2)^{-1}$. □

Issues related to this theorem include:

- The counting of self-avoiding walks (SAWS).
- The behaviour of $p_c(\mathbb{L}^d)$ as a function of d.
- In particular, the behaviour of $p_c(\mathbb{L}^d)$ for large d.

Kesten [200] proved that $p_c(\mathbb{L}^2) = \frac{1}{2}$. This very special calculation makes essential use of the self-duality of $\mathbb{L}^2$ (see Chapter 9). There are various ways of proving the *strict* inequality

$$
p_c(\mathbb{L}^d) - p_c(\mathbb{L}^{d+1}) > 0 \qquad \text{for } d \geq 2,
$$

and good recent references include [20, 158].

On the third point above, we point out that

$$
p_c(\mathbb{L}^d) = \frac{1}{2d} + \frac{1}{(2d)^2} + \frac{7}{2}\frac{1}{(2d)^3} + O\left(\frac{1}{(2d)^4}\right) \qquad \text{as } d \to \infty.
$$

See [178, 179, 180], and earlier work of [151, 209].

We note finally the canonical arguments used to establish that $0 < p_c(\mathbb{L}^d) < 1$. The first inequality was proved by counting paths, and the second by counting circuits in the dual. These approaches are fundamental to proofs of the existence of phase transition in a multitude of settings.

3.3 A Question

The definition of p_c entails that

$$\theta(p)\begin{cases} = 0 & \text{if } p < p_c, \\ > 0 & \text{if } p > p_c, \end{cases}$$

but *what happens when* $p = p_c$?

Conjecture 3.8. $\theta(p_c) = 0$.

This conjecture is known to be valid when $d = 2$ (using duality, see Section 9.1) and for sufficiently large d, currently $d \geq 19$ (using the 'bubble expansion', see Section 8.5). Concentrate your mind on the case $d = 3$.

Let us turn to the existence of an infinite open cluster, and set

$$\psi(p) = P_p(|C(x)| = \infty \text{ for some } x).$$

By using the usual zero-one law (see [170], p. 290), for any p either $\psi(p) = 0$ or $\psi(p) = 1$. Using the fact that $\mathbb{Z}^d$ is countable, we have that

$$\psi(p) = 1 \quad \text{if and only if} \quad \theta(p) > 0.$$

The above conjecture may therefore be written equivalently as $\psi(p_c) = 0$.

There has been progress towards this conjecture: see [50, 165]. It is proved that, when $p = p_c$, no half-space of $\mathbb{Z}^d$ (where $d \geq 3$) can contain an infinite open cluster. Therefore we are asked to eliminate the following absurd possibility: there exists a.s. an infinite open cluster in $\mathbb{L}^d$, but any such cluster is a.s. cut into finite parts by the removal of all edges of the form $\langle x, x+e \rangle$, as x ranges over a hyperplane of $\mathbb{L}^d$ and where e is a unit vector perpendicular to this hyperplane.

4. INEQUALITIES FOR CRITICAL PROBABILITIES

4.1 Russo's Formula

There is a fundamental formula, known in this area as Russo's formula but developed earlier in the context of reliability theory. Let E be a finite set, and let $\Omega_E = \{0,1\}^E$. For $\omega \in \Omega_E$ and $e \in E$, we define the configurations ω^e, ω_e by

$$\omega^e(f) = \begin{cases} \omega(f) & \text{if } f \neq e, \\ 1 & \text{if } f = e, \end{cases} \qquad \omega_e(f) = \begin{cases} \omega(f) & \text{if } f \neq e, \\ 0 & \text{if } f = e. \end{cases} \tag{4.1}$$

Let A be a subset of Ω_E, i.e., an event. For $\omega \in \Omega_E$, we call e *pivotal* for A if

$$\text{either} \quad \omega^e \in A,\ \omega_e \notin A \quad \text{or} \quad \omega^e \notin A,\ \omega_e \in A,$$

which is to say that the occurrence or not of A depends on the state of the edge e. Note that the set of pivotal edges for A depends on the choice of ω. We write N_A for the number of pivotal edges for A (so that N_A is a random variable). Finally, let $N : \Omega_E \to \mathbb{R}$ be given by

$$N(\omega) = \sum_{e \in E} \omega(e),$$

the 'total number of open edges'.

Theorem 4.2. *Let* $0 < p < 1$.
(a) *For any event* A,

$$\frac{d}{dp} P_p(A) = \frac{1}{p(1-p)} \operatorname{cov}_p(N, 1_A).$$

(b) *For any increasing event* A,

$$\frac{d}{dp} P_p(A) = E_p(N_A).$$

Here, P_p and E_p are the usual product measure and expectation on Ω_E, and cov_p denotes covariance.

Proof. We have that

$$P_p(A) = \sum_{\omega} p^{N(\omega)}(1-p)^{|E|-N(\omega)} 1_A(\omega)$$

whence

$$\begin{aligned} \frac{d}{dp} P_p(A) &= \sum_{\omega} \left(\frac{N(\omega)}{p} - \frac{|E| - N(\omega)}{1-p} \right) 1_A(\omega) P_p(\omega) \\ &= \frac{1}{p(1-p)} E_p\big(\{N - p|E|\} 1_A\big), \end{aligned}$$

as required for part (a).

Turning to (b), assume A is increasing. Using the definition of N, we have that

$$\text{cov}_p(N, 1_A) = \sum_{e \in E} \{P_p(A \cap J_e) - pP_p(A)\} \tag{4.3}$$

where $J_e = \{\omega(e) = 1\}$. Now, writing $\{\text{piv}\}$ for the event that e is pivotal for A,

$$P_p(A \cap J_e) = P_p(A \cap J_e \cap \{\text{piv}\}) + P_p(A \cap J_e \cap \{\text{not piv}\}).$$

We use the important fact that J_e is independent of $\{\text{piv}\}$, which holds since the latter event depends only on the states of edges f other than e. Since $A \cap J_e \cap \{\text{piv}\} = J_e \cap \{\text{piv}\}$, the first term on the right side above equals

$$P_p(J_e \cap \{\text{piv}\}) = P_p(J_e \mid \text{piv})P_p(\text{piv}) = pP_p(\text{piv}),$$

and similarly the second term equals (since J_e is independent of the event $A \cap \{\text{not piv}\}$)

$$P_p(J_e \mid A \cap \{\text{not piv}\})P_p(A \cap \{\text{not piv}\}) = pP_p(A \cap \{\text{not piv}\}).$$

Returning to (4.3), the summand equals

$$\begin{aligned}
\{pP_p(\text{piv}) + pP_p(&A \cap \{\text{not piv}\})\} - p\{P_p(A \cap \{\text{piv}\}) + P_p(A \cap \{\text{not piv}\})\} \\
&= pP_p(\overline{A} \cap \{\text{piv}\}) = pP_p(\overline{J_e} \mid \text{piv})P_p(\text{piv}) \\
&= p(1-p)P_p(\text{piv}).
\end{aligned}$$

Insert this into (4.3) to obtain part (b) from part (a). An alternative proof of part (b) may be found in [G]. □

Although the above theorem was given for a finite product space Ω_E, the conclusion is clearly valid for the infinite space Ω so long as the event A is finite-dimensional.

The methods above may be used further to obtain formulae for the higher derivatives of $P_p(A)$. First, Theorem 4.2(b) may be generalised to obtain that

$$\frac{d}{dp}E_p(X) = \sum_{e \in E} E_p(\delta_e X),$$

where X is any given random variable on Ω and $\delta_e X$ is defined by $\delta_e X(\omega) = X(\omega^e) - X(\omega_e)$. It follows that

$$\frac{d^2}{dp^2}E_p(X) = \sum_{e,f \in E} E_p(\delta_e \delta_f X).$$

Now $\delta_e \delta_e X = 0$, and for $e \neq f$

$$\delta_e \delta_f X(\omega) = X(\omega^{ef}) - X(\omega_f^e) - X(\omega_e^f) + X(\omega_{ef}).$$

Let $X = 1_A$ where A is an increasing event. We deduce that

$$\frac{d^2}{dp^2}P_p(A) = \sum_{\substack{e,f\in E\\ e\neq f}} \Big\{1_A(\omega^{ef})(1-1_A(\omega_e^f))(1-1_A(\omega_f^e)) \\ - 1_A(\omega_f^e)1_A(\omega_e^f)(1-1_A(\omega_{ef})\Big\}$$
$$= E_p(N_A^{\text{ser}}) - E_p(N_A^{\text{par}})$$

where N_A^{ser} (resp. N_A^{par}) is the number of distinct ordered pairs e, f of edges such that $\omega^{ef} \in A$ but $\omega_e^f, \omega_f^e \notin A$ (resp. $\omega_e^f, \omega_f^e \in A$ but $\omega_{ef} \notin A$). (The superscripts here are abbreviations for 'series' and 'parallel'.) This argument may be generalised to higher derivatives.

4.2 Strict Inequalities for Critical Probabilities

If $\mathcal{L}$ is a sublattice of the lattice $\mathcal{L}'$ (written $\mathcal{L} \subseteq \mathcal{L}'$) then clearly $p_c(\mathcal{L}) \geq p_c(\mathcal{L}')$, but when does the *strict* inequality $p_c(\mathcal{L}) > p_c(\mathcal{L}')$ hold? The question may be quantified by asking for non-trivial lower bounds for $p_c(\mathcal{L}) - p_c(\mathcal{L}')$.

Similar questions arise in many ways, not simply within percolation theory. More generally, consider any process indexed by a continuously varying parameter T and enjoying a phase transition at some point $T = T_c$. In many cases of interest, enough structure is available to enable us to conclude that certain systematic changes to the process can change T_c but that any such change must push T_c in one particular direction (thereby *increasing* T_c, say). The question then is to understand which systematic changes change T_c *strictly*. In the context of the previous paragraph, the systematic changes in question involve the 'switching on' of edges lying in $\mathcal{L}'$ but not in $\mathcal{L}$.

A related percolation question is that of 'entanglements'. Consider bond percolation on $\mathbb{L}^3$, and examine the box $B(n)$. Think about the open edges as being solid connections made of elastic, say. Try to 'pull apart' a pair of opposite faces of $B(n)$. If $p > p_c$, then you will generally fail because, with large probability (tending to 1 as $n \to \infty$), there is an open path joining one face to the other. Even if $p < p_c$ then you may fail, owing to an 'entanglement' of open paths (a necklace of necklaces, perhaps, see Figure 4.1). It may be seen that there is an 'entanglement transition' at some critical point p_e satisfying $p_e \leq p_c$. Is it the case that strict inequality holds, i.e., $p_e < p_c$?

A technology has been developed for approaching such questions of strict inequality. Although, in particular cases, ad hoc arguments can be successful, there appears to be only one general approach. We illustrate this approach in the next section, by sketching the details in a particular case.

Important references include [20, 157, 158, 269]. See also [75].

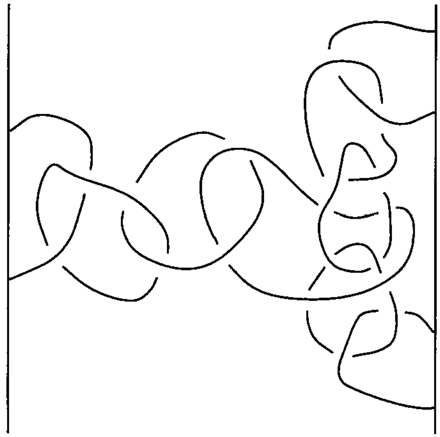


Fig. 4.1. An entanglement between opposite sides of a cube in three dimensions. Note the necklace of necklaces on the right.

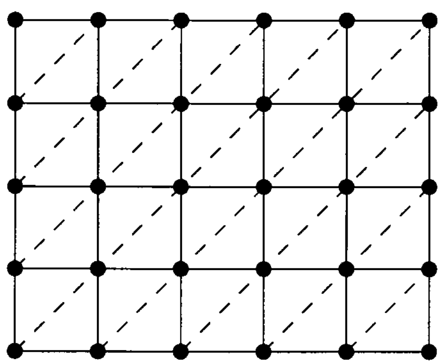


Fig. 4.2. The triangular lattice may be obtained from the square lattice by the addition of certain diagonals.

4.3 The Square and Triangular Lattices

The triangular lattice $\mathbb{T}$ may be obtained by adding diagonals across the squares of the square lattice $\mathbb{L}^2$, in the manner of Figure 4.2. Since any infinite open cluster of $\mathbb{L}^2$ is also an infinite open cluster of $\mathbb{T}$, it follows that $p_c(\mathbb{T}) \le p_c(\mathbb{L}^2)$, but does strict inequality hold? There are various ways of proving the strict inequality. Here we adopt the canonical argument of [20], as an illustration of a general technique.

Before embarking on this exercise, we point out that, *for this particular case*, there is a variety of ways of obtaining the result, by using special properties of the square and triangular lattices. The attraction of the method described here is its generality, relying as it does on essentially no assumptions about graph-structure or number of dimensions.

First we embed the problem in a two-parameter system. Let $0 \le p, s \le 1$. We declare each edge of $\mathbb{L}^2$ to be open with probability p, and each *further edge* of $\mathbb{T}$

(i.e., the dashed edges in Figure 4.2) to be open with probability s. Writing $P_{p,s}$ for the associated measure, define

$$\theta(p, s) = P_{p,s}(0 \leftrightarrow \infty).$$

We propose to prove differential inequalities which imply that $\partial\theta/\partial p$ and $\partial\theta/\partial s$ are comparable, uniformly on any closed subset of the interior $(0,1)^2$ of the parameter space. This cannot itself be literally achieved, since we have insufficient information about the differentiability of θ. Therefore we approximate θ by a finite-volume quantity θ_n, and we then work with the partial derivatives of θ_n.

For any set A of vertices, we define the 'interior boundary' ∂A by

$$\partial A = \{a \in A : a \sim b \text{ for some } b \notin A\}.$$

Let $B(n) = [-n, n]^d$, and define

$$\theta_n(p, s) = P_{p,s}\big(0 \leftrightarrow \partial B(n)\big). \tag{4.4}$$

Note that θ_n is a polynomial in p and s, and that

$$\theta_n(p, s) \downarrow \theta(p, s) \qquad \text{as } n \to \infty.$$

Lemma 4.5. *There exists a positive integer L and a continuous strictly positive function $g : (0,1)^2 \to (0, \infty)$ such that*

$$g(p, s)^{-1}\frac{\partial}{\partial p}\,\theta_n(p, s) \geq \frac{\partial}{\partial s}\,\theta_n(p, s) \geq g(p, s)\frac{\partial}{\partial p}\,\theta_n(p, s) \tag{4.6}$$

for $0 < p, s < 1$, $n \geq L$.

Once this is proved, the main result follows immediately, namely the following.

Theorem 4.7. *It is the case that $p_c(\mathbb{T}) < p_c(\mathbb{L}^2)$.*

Sketch Proof of Theorem 4.7. Here is a rough argument, which needs some rigour. There is a 'critical curve' in (p, s)-space, separating the regime where $\theta(p, s) = 0$ from that when $\theta(p, s) > 0$ (see Figure 4.3). Suppose that this critical curve may be written in the form $h(p, s) = 0$ for some increasing and continuously differentiable function h. It is enough to prove that the graph of h contains no vertical segment. Now

$$\nabla h = \left(\frac{\partial h}{\partial p}, \frac{\partial h}{\partial s}\right)$$

and, by Lemma 4.5,

$$\nabla h \cdot (0, 1) = \frac{\partial h}{\partial s} \geq g(p, s)\frac{\partial h}{\partial p},$$

whence

$$\frac{1}{|\nabla h|}\frac{\partial h}{\partial s} = \left\{\left(\frac{\partial h}{\partial p} \bigg/ \frac{\partial h}{\partial s}\right)^2 + 1\right\}^{-\frac{1}{2}} \geq \frac{g}{\sqrt{g^2 + 1}},$$

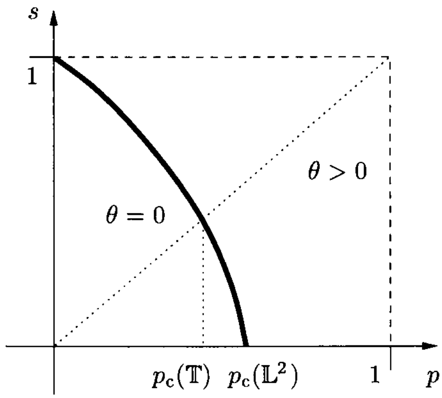


Fig. 4.3. The critical 'surface'. The area beneath the curve is the set of (p, s) for which $\theta(p, s) = 0$.

which is bounded away from 0 on any closed subset of $(0, 1)^2$. This indicates as required that h has no vertical segment.

Here is the proper argument. There is more than one way of defining the critical surface. Let $C_{\text{sub}} = \{(p, s) : \theta(p, s) = 0\}$, and let C_{crit} be the set of all points lying in the closure of both C_{sub} and its complement.

Let η be positive and small, and find γ (> 0) such that $g(p, s) \geq \gamma$ on $[\eta, 1 - \eta]^2$. At the point $(a, b) \in [\eta, 1 - \eta]^2$, the rate of change of $\theta_n(a, b)$ in the direction $(\cos\alpha, -\sin\alpha)$, where $0 \leq \alpha < \frac{\pi}{2}$, is

$$\begin{aligned} \text{(4.8)} \qquad \nabla\theta_n \cdot (\cos\alpha, -\sin\alpha) &= \frac{\partial\theta_n}{\partial a}\cos\alpha - \frac{\partial\theta_n}{\partial b}\sin\alpha \\ &\leq \frac{\partial\theta_n}{\partial a}(\cos\alpha - \gamma\sin\alpha) \leq 0 \end{aligned}$$

so long as $\tan\alpha \geq \gamma^{-1}$.

Suppose $\theta(a, b) = 0$, and $\tan\alpha = \gamma^{-1}$. Let $(a', b') = (a, b) + \epsilon(\cos\alpha, -\sin\alpha)$ where ϵ is small and positive. Then, by (4.8),

$$\theta(a', b') = \lim_{n\to\infty} \theta_n(a', b') \leq \lim_{n\to\infty} \theta_n(a, b) = \theta(a, b) = 0,$$

whence $(a', b') \in C_{\text{sub}}$.

There is quite a lot of information in such a calculation, but we abstract a small amount only. Take $a = b = p_c(\mathbb{T}) - \zeta$ for some small positive ζ. Then choose ϵ large enough so that $a' > p_c(\mathbb{T})$. The above calculation, for small enough ζ, implies that

$$\theta(a', 0) \leq \theta(a', b') = 0,$$

whence $p_c(\mathbb{L}^2) \geq a' > p_c(\mathbb{T})$. □

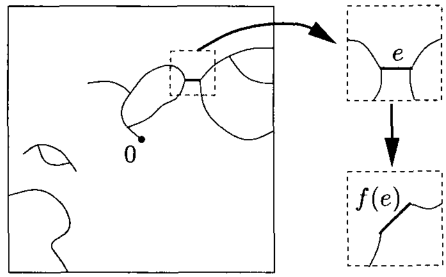


Fig. 4.4. Inside the box $B(n)$, the edge e is pivotal for the event $\{0 \leftrightarrow \partial B(n)\}$. By altering the configuration inside the smaller box, we may construct a configuration in which $f(e)$ is pivotal instead.

Proof of Lemma 4.5. With $\mathbb{E}^2$ the edge set of $\mathbb{L}^2$, and $\mathbb{F}$ the additional edges in the triangular lattice $\mathbb{T}$ (i.e., the diagonals in Figure 4.2), we have by Russo's formula (in a slightly more general version than Theorem 4.2) that

$$\begin{aligned}\frac{\partial}{\partial p}\,\theta_n(p,s) &= \sum_{e\in\mathbb{E}^2} P_{p,s}(e \text{ is pivotal for } A_n),\\ \frac{\partial}{\partial s}\,\theta_n(p,s) &= \sum_{f\in\mathbb{F}} P_{p,s}(f \text{ is pivotal for } A_n),\end{aligned} \tag{4.9}$$

where $A_n = \{0 \leftrightarrow \partial B(n)\}$. The idea now is to show that each summand in the first summation is comparable with some given summand in the second. Actually we shall only prove the second inequality in (4.6), since this is the only one used in proving the theorem, and additionally the proof of the other part is similar.

With each edge e of $\mathbb{E}^2$ we associate a unique edge $f = f(e)$ of $\mathbb{F}$ such that f lies near to e. This may be done in a variety of ways, but in order to be concrete we specify that if $e = \langle u, u+e_1\rangle$ or $e = \langle u, u+e_2\rangle$ then $f = \langle u, u+e_1+e_2\rangle$, where e_1 and e_2 are unit vectors in the directions of the (increasing) x and y axes.

We claim that there exists a function $h(p,s)$, strictly positive on $(0,1)^2$, such that

$$h(p,s)P_{p,s}(e \text{ is pivotal for } A_n) \le P_{p,s}(f(e) \text{ is pivotal for } A_n) \tag{4.10}$$

for all e lying in $B(n)$. Once this is shown, we sum over e to obtain by (4.9) that

$$\begin{aligned}h(p,s)\frac{\partial}{\partial p}\,\theta_n(p,s) &\le \sum_{e\in\mathbb{E}^2} P_{p,s}(f(e) \text{ is pivotal for } A_n)\\ &\le 2\sum_{f\in\mathbb{F}} P_{p,s}(f \text{ is pivotal for } A_n)\\ &= 2\frac{\partial}{\partial s}\,\theta_n(p,s)\end{aligned}$$

as required. The factor 2 arises because, for each f ($\in \mathbb{F}$), there are exactly two edges e with $f(e) = f$.

Finally, we indicate the reason for (4.10). Let us consider the event $\{e \text{ is pivotal for } A_n\}$. We claim that there exists an integer M, chosen uniformly for edges e in $B(n)$ and for all large n, such that

(a) all paths from 0 to $\partial B(n)$ pass through the region $e + B(M)$
(b) by altering the configuration within $e + B(M)$ *only*, we may obtain an event on which $f(e)$ is pivotal for A_n.

This claim is proved by inspecting Figure 4.4. A special argument may be needed when the box $e + B(M)$ either contains the origin or intersects $\partial B(n)$, but such special arguments pose no substantial difficulty. Once this geometrical claim is accepted, (4.10) follows thus. Write E_g for the event that the edge g is pivotal for A_n. For $\omega \in E_e$, let $\omega' = \omega'(\omega)$ be the configuration obtained as above, so that ω' agrees with ω off $e + B(M)$, and furthermore $\omega' \in E_{f(e)}$. Then

$$P_{p,s}(E_e) = \sum_{\omega \in E_e} P_{p,s}(\omega) \leq \sum_{\omega \in E_e} \frac{1}{\alpha^R} P_{p,s}(\omega') \leq \left(\frac{2}{\alpha}\right)^R P_{p,s}(E_{f(e)})$$

where $\alpha = \min\{p, s, 1-p, 1-s\}$ and R is the number of edges of $\mathbb{T}$ in $e + B(M)$. □

4.4 Enhancements

An 'enhancement' is loosely defined as a systematic addition of connections according to local rules. Enhancements may involve further coin flips. Can an enhancement create an infinite cluster when previously there was none?

Clearly the answer can be negative. For example the rule may be of the type: join any two neighbours of $\mathbb{Z}^d$ with probability $\frac{1}{2}p_c$, whenever they have no incident open edges. Such an enhancement creates extra connections but (a.s.) no extra infinite cluster.

Here is a proper definition. Consider bond percolation on $\mathbb{L}^d$ with parameter p, and consider enhancements of the following type. Let $R > 0$, and let f be a function which associates to each configuration on the box $B(R)$ a graph on $\mathbb{Z}^d$ with finitely many edges. For each $x \in \mathbb{Z}^d$, we observe the configuration ω on the box $x + B(R)$, and we write $f(x, \omega)$ for the associated evaluation of f. The enhanced configuration is the graph

$$G(\text{enh}) = G(\omega) \cup \left\{ \bigcup_{x : H(x) = 1} \{x + f(x, \omega)\} \right\}$$

where $G(\omega)$ is the graph of open edges, and $\{H(x) : x \in \mathbb{Z}^d\}$ is a family of Bernoulli random variables, each taking the value 1 with probability s (independently of everything else). The parameter s is the 'density' of the enhancement. In writing the union of graphs, we mean the graph with vertex set $\mathbb{Z}^d$ having the union of the appropriate edge sets.

We call such an enhancement *essential* if there exists a percolation configuration ω containing no doubly-infinite open path but such that $G(\omega) \cup f(0, \omega)$ does contain such a path. The following theorem is taken from [20] and may be proved in a manner similar to the proof given in the last section.

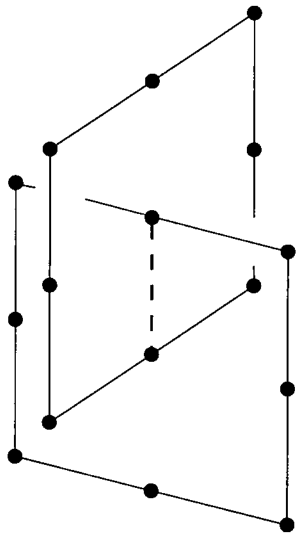


Fig. 4.5. A sketch of the enhancement which adds an edge between any two interlocking 2×2 squares in $\mathbb{L}^3$.

Theorem 4.11. *Let $s > 0$. For any essential enhancement, there exists a non-empty interval $(\pi(s), p_c)$ such that*

$$P\big(G(\text{enh}) \text{ contains an infinite cluster }\big) > 0$$

when $\pi(s) < p \leq p_c$.

That is, essential enhancements shift the critical point *strictly*. Here is such an enhancement relevant to the entanglement transition in $\mathbb{L}^3$. Whenever we see two interlinking 2×2 open squares, then we join them by an edge (see Figure 4.5). It is easy to see that this enhancement is essential, and therefore it shifts the critical point downwards. Hence the entanglement critical point p_e satisfies $p_e < p_c$. See [20, 198].

Finally we note that one may find explicit functions g in Lemma 4.5, whence the mechanism of the method leads to numerical lower bounds on the change in critical value.

5. CORRELATION INEQUALITIES

5.1 FKG Inequality

The FKG inequality for percolation processes was discovered by Harris [181], and is often named now after the authors of [144] who proved a more general version which is the subject of this section.

Let E be a finite set, and $\Omega_E = \{0,1\}^E$ as usual. We write $\mathcal{F}_E$ for the set of all subsets of Ω_E, and call a probability measure μ on $(\Omega_E, \mathcal{F}_E)$ *positive* if $\mu(\omega) > 0$ for all $\omega \in \Omega_E$.

Theorem 5.1 (FKG Inequality). *Let μ be a positive probability measure on $(\Omega_E, \mathcal{F}_E)$ such that*

$$\mu(\omega_1 \vee \omega_2)\mu(\omega_1 \wedge \omega_2) \geq \mu(\omega_1)\mu(\omega_2) \qquad \textit{for all } \omega_1, \omega_2 \in \Omega_E. \tag{5.2}$$

Then

$$\mu(fg) \geq \mu(f)\mu(g) \tag{5.3}$$

for all increasing random variables $f, g : \Omega_E \to \mathbb{R}$.

Here, $\omega_1 \vee \omega_2$ and $\omega_1 \wedge \omega_2$ are defined as the maximum and minimum configurations,

$$\omega_1 \vee \omega_2(e) = \max\{\omega_1(e), \omega_2(e)\}, \quad \omega_1 \wedge \omega_2(e) = \min\{\omega_1(e), \omega_2(e)\},$$

for all $e \in E$. In (5.3), we have used μ to denote *expectation* as well as *probability*.

Specialising to the indicator functions $f = 1_A$, $g = 1_B$, inequality (5.3) implies that

$$\mu(A \cap B) \geq \mu(A)\mu(B) \quad \text{for increasing events } A, B. \tag{5.4}$$

It is easily checked that the product measure P_p satisfies the hypotheses of the theorem (when $0 < p < 1$), and therefore P_p satisfies the FKG inequality (5.3). This inequality may be proved directly in the special case of product measure (see [G], p. 26). Here we shall prove the more general theorem given above. The proof proceeds by first proving a theorem about stochastic orderings of measures, usually called Holley's inequality after [192].

Theorem 5.5 (Holley's Inequality). *Let μ_1 and μ_2 be positive probability measures on $(\Omega_E, \mathcal{F}_E)$ such that*

$$\mu_1(\omega_1 \vee \omega_2)\mu_2(\omega_1 \wedge \omega_2) \geq \mu_1(\omega_1)\mu_2(\omega_2) \qquad \textit{for all } \omega_1, \omega_2 \in \Omega_E. \tag{5.6}$$

Then

$$\mu_1(f) \geq \mu_2(f) \qquad \textit{for all increasing } f : \Omega_E \to \mathbb{R},$$

which is to say that $\mu_1 \geq \mu_2$.

Proof of Theorem 5.5. The theorem is 'merely' a numerical inequality involving a finite number of positive reals. It may be proved in a totally elementary manner, using essentially no general mechanism. Nevertheless, in a more useful (and remarkable) proof we construct Markov chains and appeal to the ergodic theorem. This requires a mechanism, but the method is beautiful, and in addition yields a structure which finds applications elsewhere.

The main step is the proof that μ_1 and μ_2 can be 'coupled' in such a way that the component with marginal measure μ_1 lies above (in the sense of sample realisations) that with marginal measure μ_2. This is achieved by constructing a certain Markov chain with the coupled measure as unique invariant measure.

Here is a preliminary calculation. Let μ be a positive probability measure on $(\Omega_E, \mathcal{F}_E)$. We may construct a time-reversible Markov chain with state space Ω_E and unique invariant measure μ, in the following way. We do this by choosing a suitable generator (or 'Q-matrix') satisfying the detailed balance equations. The dynamics of the chain involve the 'switching on or off' of components of the current state. For $\omega \in \Omega_E$, let ω^e and ω_e be given as in (4.1). Define the function $G : \Omega_E^2 \to \mathbb{R}$ by

$$G(\omega_e, \omega^e) = 1, \quad G(\omega^e, \omega_e) = \frac{\mu(\omega_e)}{\mu(\omega^e)}, \tag{5.7}$$

for all $\omega \in \Omega_E$, $e \in E$; define $G(\omega, \omega') = 0$ for all other pairs ω, ω' with $\omega \neq \omega'$. The diagonal elements are chosen so that

$$\sum_{\omega'} G(\omega, \omega') = 0 \qquad \text{for all } \omega \in \Omega_E.$$

It is elementary that

$$\mu(\omega) G(\omega, \omega') = \mu(\omega') G(\omega', \omega) \qquad \text{for all } \omega, \omega' \in \Omega_E,$$

and therefore G generates a time-reversible Markov chain on the state space Ω_E. This chain is irreducible (using (5.7)), and therefore has a unique invariant measure μ (see [170], p. 208).

We next follow a similar route for *pairs* of configurations. Let μ_1 and μ_2 satisfy the hypotheses of the theorem, and let S be the set of all pairs (π, ω) of configurations in Ω_E satisfying $\pi \leq \omega$. We define $H : S \times S \to \mathbb{R}$ by

$$H(\pi_e, \omega; \pi^e, \omega^e) = 1, \tag{5.8}$$

$$H(\pi, \omega^e; \pi_e, \omega_e) = \frac{\mu_1(\omega_e)}{\mu_1(\omega^e)}, \tag{5.9}$$

$$H(\pi^e, \omega^e; \pi_e, \omega^e) = \frac{\mu_2(\pi_e)}{\mu_2(\pi^e)} - \frac{\mu_1(\omega_e)}{\mu_1(\omega^e)}, \tag{5.10}$$

for all $(\pi, \omega) \in S$ and $e \in E$; all other off-diagonal values of H are set to 0. The diagonal terms are chosen so that

$$\sum_{\pi', \omega'} H(\pi, \omega; \pi', \omega') = 0 \qquad \text{for all } (\pi, \omega) \in S.$$

Equation (5.8) specifies that, for $\pi \in \Omega_E$ and $e \in E$, the edge e is acquired by π (if it does not already contain it) at rate 1; any edge so acquired is added also to ω if it does not already contain it. (Here, we speak of a configuration ψ containing an edge e if $\psi(e) = 1$.) Equation (5.9) specifies that, for $\omega \in \Omega_E$ and $e \in E$ with $\omega(e) = 1$, the edge e is removed from ω (and also from π if $\pi(e) = 1$) at the rate given in (5.9). For e with $\pi(e) = 1$, there is an additional rate given in (5.10) at which e is removed from π but not from ω. We need to check that this additional rate is indeed non-negative. This poses no problem, since the required inequality

$$\mu_1(\omega^e)\mu_2(\pi_e) \geq \mu_1(\omega_e)\mu_2(\pi^e) \quad \text{where } \pi \leq \omega$$

follows from assumption (5.6).

Let $(X_t, Y_t)_{t\geq 0}$ be a Markov chain on S having generator H, and set $(X_0, Y_0) = (0, 1)$, where 0 (resp. 1) is the state of all 0's (resp. 1's). By examination of (5.8)–(5.10) we see that $X = (X_t)_{t\geq 0}$ is a Markov chain with generator given by (5.7) with $\mu = \mu_2$, and that $Y = (Y_t)_{t\geq 0}$ arises similarly with $\mu = \mu_1$.

Let κ be an invariant measure for the paired chain $(X_t, Y_t)_{t\geq 0}$. Since X and Y have (respective) unique invariant measures μ_2 and μ_1, it follows that the marginals of κ are μ_2 and μ_1. We have by construction that

$$\kappa\big(\{(\pi, \omega) : \pi \leq \omega\}\big) = 1,$$

and κ is the required 'coupling' of μ_1 and μ_2.

Let $(\pi, \omega) \in S$ be chosen according to the measure κ. Then

$$\mu_1(f) = \kappa\big(f(\omega)\big) \geq \kappa\big(f(\pi)\big) = \mu_2(f),$$

for any increasing function f. Therefore $\mu_1 \geq \mu_2$. $\square$

Proof of Theorem 5.1. Assume that μ satisfies (5.2), and let f and g be increasing functions. By adding a constant to the function g, we see that it suffices to prove (5.3) under the extra hypothesis that g is strictly positive. We assume this holds. Define positive probability measures μ_1 and μ_2 on $(\Omega_E, \mathcal{F}_E)$ by $\mu_2 = \mu$ and

$$\mu_1(\omega) = \frac{g(\omega)\mu(\omega)}{\sum_{\omega'} g(\omega')\mu(\omega')} \qquad \text{for } \omega \in \Omega_E.$$

Since g is increasing, (5.6) follows from (5.2). By Holley's inequality,

$$\mu_1(f) \geq \mu_2(f),$$

which is to say that

$$\frac{\sum_\omega f(\omega)g(\omega)\mu(\omega)}{\sum_{\omega'} g(\omega')\mu(\omega')} \geq \sum_\omega f(\omega)\mu(\omega)$$

as required. $\square$

5.2 Disjoint Occurrence

Van den Berg has suggested a converse to the FKG inequality, namely that, for some interpretation of the binary operation $\circ$,

$$P_p(A \circ B) \leq P_p(A)P_p(B) \quad \text{for all increasing events } A, B.$$

The correct interpretation of $A \circ B$ turns out to be 'A and B occur disjointly'. We explain this statement next.

As usual, E is a finite set, $\Omega_E = \{0,1\}^E$, and so on. For $\omega \in \Omega_E$, let

$$K(\omega) = \{e \in E : \omega(e) = 1\},$$

so that there is a one–one correspondence between configurations ω and sets $K(\omega)$. For increasing events A, B, let

$$A \circ B = \Big\{\omega : \text{for some } H \subseteq K(\omega), \text{ we have that } \omega' \in A \text{ and } \omega'' \in B, \\ \text{where } K(\omega') = H \text{ and } K(\omega'') = K(\omega) \setminus H\Big\},$$

and we call $A \circ B$ the event that A and B *occur disjointly*.

The canonical example of disjoint occurrence in percolation theory concerns the existence of disjoint open paths. If $A = \{u \leftrightarrow v\}$ and $B = \{x \leftrightarrow y\}$, then $A \circ B$ is the event that are two edge-disjoint paths, one joining u to v, and the other joining x to y.

Theorem 5.11 (BK Inequality [67]). *If A and B are increasing events, then*

$$P_p(A \circ B) \leq P_p(A)P_p(B).$$

Proof. The following sketch can be made rigorous (see [58], and [G], p. 32). For the sake of being concrete, we take E to be the edge-set of a finite graph G, and consider the case when $A = \{u \leftrightarrow v\}$ and $B = \{x \leftrightarrow y\}$ for four distinct vertices u, v, x, y.

Let e be an edge of E. In the process of 'splitting' e, we replace e by two copies e' and e'' of itself, each of which is open with probability p (independently of the other, and of all other edges). Having split e, we look for disjoint paths from u to v, and from x to y, but with the following difference: the path from u to v is not permitted to use e'', and the path from x to y is not permitted to use e'.

The crucial observation is that this splitting *cannot decrease* the chance of finding the required open paths.

We split each edge in turn, the appropriate probability being non-decreasing at each stage. After every edge has been split, we are then looking for two paths within two independent copies of G, and this probability is just $P_p(A)P_p(B)$. Therefore

$$P_p(A \circ B) \leq \cdots \leq P_p(A)P_p(B). \qquad \square$$

Van den Berg and Kesten [67] conjectured a similar inequality for arbitrary A and B (not just the monotone events), with a suitable redefinition of the operation

o. Their conjecture rebutted many serious attempts at proof, before 1995. Here is the more general statement.

For $\omega \in \Omega_E$, $K \subseteq E$, define the cylinder event

$$C(\omega, K) = \{\omega' : \omega'(e) = \omega(e) \text{ for } e \in K\}.$$

Now, for events A and B, define

$$A \,\square\, B = \Big\{\omega : \text{for some } K \subseteq E, \text{ we have } C(\omega, K) \subseteq A \text{ and } C(\omega, \overline{K}) \subseteq B\Big\}.$$

Theorem 5.12 (Reimer's Inequality [322]). *For all events A and B,*

$$P_p(A \,\square\, B) \le P_p(A) P_p(B).$$

The search is on for 'essential' applications of this beautiful inequality; such an application may be found in the study of dependent percolation models [65]. Related results may be found in [62, 64].

Note that Reimer's inequality contains the FKG inequality, by using the fact that $A \,\square\, \overline{B} = A \cap \overline{B}$ if A and B are increasing events.

5.3 SITE AND BOND PERCOLATION

Let $G = (V, E)$ be an infinite connected graph with maximum vertex degree Δ. For a vertex x, define $\theta(p, x, \text{bond})$ (resp. $\theta(p, x, \text{site})$) to be the probability that x lies in an infinite open cluster of G in a bond percolation (resp. site percolation) process on G with parameter p. Clearly $\theta(p, x, \text{bond})$ and $\theta(p, x, \text{site})$ are non-decreasing in p. Also, using the FKG inequality,

$$\theta(p, x, \text{bond}) \ge P_p\Big(\{x \leftrightarrow y\} \cap \{y \leftrightarrow \infty\}\Big) \ge P_p(x \leftrightarrow y)\theta(p, y, \text{bond}),$$

with a similar inequality for the site process. It follows that the critical points

$$p_c(\text{bond}) = \sup\{p : \theta(p, x, \text{bond}) = 0\},$$
$$p_c(\text{site}) = \sup\{p : \theta(p, x, \text{site}) = 0\},$$

exist and are independent of the choice of the vertex x.

Theorem 5.13. *We have that*

$$\frac{1}{\Delta - 1} \le p_c(\text{bond}) \le p_c(\text{site}) \le 1 - \big(1 - p_c(\text{bond})\big)^{\Delta}. \tag{5.14}$$

One consequence of this theorem is that $p_c(\text{bond}) < 1$ if and only if $p_c(\text{site}) < 1$. The third inequality of (5.14) may be improved by replacing the exponent Δ by $\Delta - 1$, but we do no prove this here. Also, the methods of Chapter 4 may be used to establish the strict inequality $p_c(\text{bond}) < p_c(\text{site})$. See [169] for proofs of the latter facts.

Proof. The first inequality of (5.14) follows by counting paths, as in the proof of (3.4). We turn to the remaining two inequalities. Let 0 be a vertex of G, called the *origin*. We claim that

$$C'(p, 0, \text{site}) \le C(p, 0, \text{bond}) \tag{5.15}$$

and

$$C(p, 0, \text{bond}) \le C'(p', 0, \text{site}) \quad \text{if } p' \ge 1 - (1-p)^{\Delta}, \tag{5.16}$$

where "$\le$" denotes stochastic ordering, and where $C(p, 0, \text{bond})$ (resp. $C'(p, 0, \text{site})$) has the law of the cluster of bond percolation at the origin (resp. the cluster of site percolation at the origin conditional on 0 being an open site). Since

$$\begin{aligned} \theta(p, 0, \text{bond}) &= \text{Prob}\big(|C(p, 0, \text{bond})| = \infty\big), \\ p^{-1}\theta(p, 0, \text{site}) &= \text{Prob}\big(|C'(p, 0, \text{site})| = \infty\big), \end{aligned}$$

the remaining claims of (5.14) follow from (5.15)–(5.16).

We construct appropriate couplings in order to prove (5.15)–(5.16). Let $\omega \in \{0,1\}^E$ be a realisation of a bond percolation process on $G = (V, E)$ with density p. We may build the cluster at the origin in the following standard manner. Let $e_1, e_2, \dots$ be a fixed ordering of E. At each stage k of the inductive construction, we shall have a pair (A_k, B_k) where $A_k \subseteq V$, $B_k \subseteq E$. Initially we set $A_0 = \{0\}$, $B_0 = \varnothing$. Having found (A_k, B_k) for some k, we define (A_{k+1}, B_{k+1}) as follows. We find the earliest edge e_i in the ordering of E with the following properties: $e_i \notin B_k$, and e_i is incident with exactly one vertex of A_k, say the vertex x. We now set

$$A_{k+1} = \begin{cases} A_k & \text{if } e_i \text{ is closed}, \\ A_k \cup \{y\} & \text{if } e_i \text{ is open}, \end{cases} \tag{5.17}$$

$$B_{k+1} = \begin{cases} B_k \cup \{e_i\} & \text{if } e_i \text{ is closed}, \\ B_k & \text{if } e_i \text{ is open}, \end{cases} \tag{5.18}$$

where $e_i = \langle x, y \rangle$. If no such edge e_i exists, we declare $(A_{k+1}, B_{k+1}) = (A_k, B_k)$. The sets A_k, B_k are non-decreasing, and the open cluster at the origin is given by $A_\infty = \lim_{k\to\infty} A_k$.

We now augment the above construction in the following way. We colour the vertex 0 *red*. Furthermore, on obtaining the edge e_i given above, we colour the vertex y *red* if e_i is open, and *black* otherwise. We specify that each vertex is coloured at most once in the construction, in the sense that any vertex y which is obtained at two or more stages is coloured in perpetuity according to the first colour it receives.

Let $A_\infty(\text{red})$ be the set of points connected to the origin by red paths of G. It may be seen that $A_\infty(\text{red}) \subseteq A_\infty$, and that $A_\infty(\text{red})$ has the same distribution as $C'(p, 0, \text{site})$. Inequality (5.15) follows.

The derivation of (5.16) is similar but slightly more complicated. We start with a directed version of G, namely $\vec{G} = (V, \vec{E})$ obtained from G by replacing each edge $e = \langle x, y \rangle$ by two directed edges, one in each direction, and denoted respectively by

$[x,y\rangle$ and $[y,x\rangle$. We now let $\overrightarrow{\omega} \in \{0,1\}^{\overrightarrow{E}}$ be a realisation of an (oriented) bond percolation process on $\overrightarrow{G}$ with density p.

We colour the origin *green*. We colour a vertex x ($\neq 0$) *green* if at least one edge f of the form $[y,x\rangle$ satisfies $\overrightarrow{\omega}(f) = 1$; otherwise we colour x *black*. Then

$$(5.19) \qquad P_p(x \text{ is green}) = 1 - (1-p)^{\rho(x)} \leq 1 - (1-p)^{\Delta},$$

where $\rho(x)$ is the degree of x, and $\Delta = \max_x \rho(x)$.

We now build a copy A_∞ of $C(p, 0, \text{bond})$ more or less as described above in (5.17)–(5.18). The only difference is that, on obtaining the edge $e_i = \langle x, y\rangle$ where $x \in A_k$, $y \notin A_k$, we declare e_i to be open for the purpose of (5.17)–(5.18) if and only if $\overrightarrow{\omega}([x,y\rangle) = 1$. Finally, we set $A_\infty(\text{green})$ to be the set of points connected to the origin by green paths. It may be seen that $A_\infty(\text{green}) \supseteq A_\infty$. Furthermore, by (5.19), $A_\infty(\text{green})$ is no larger in distribution that $C'(p', 0, \text{site})$ where $p' = 1 - (1-p)^{\Delta}$. Inequality (5.16) follows. □

6. SUBCRITICAL PERCOLATION

6.1 Using Subadditivity

We assume throughout this chapter that $p < p_c$. All open clusters are a.s. finite, and the phase is sometimes called 'disordered' by mathematical physicists, since there are no long-range connections. In understanding the phase, we need to know how fast the tails of certain distributions go to zero, and a rule of thumb is that 'everything reasonable' should have exponentially decaying tails. In particular, the limits

$$\phi(p) = \lim_{n\to\infty} \left\{ -\frac{1}{n} \log P_p\big(0 \leftrightarrow \partial B(n)\big) \right\},$$
$$\zeta(p) = \lim_{n\to\infty} \left\{ -\frac{1}{n} \log P_p(|C| = n) \right\},$$

should exist, and be strictly positive when $p < p_c$. The function $\phi(p)$ measures a 'distance effect' and $\zeta(p)$ a 'volume effect'.

The existence of such limits is a quite different matter from their positiveness. Existence is usually proved by an appeal to subadditivity (see below) via a correlation inequality. To show positiveness usually requires a hard estimate.

Theorem 6.1 (Subadditive Inequality). *If $(x_r : r \geq 1)$ is a sequence of reals satisfying the subadditive inequality*

$$x_{m+n} \leq x_m + x_n \quad \textit{for all } m, n,$$

then the limit

$$\lambda = \lim_{r\to\infty} \left\{ \frac{1}{r} x_r \right\}$$

exists, with $-\infty \leq \lambda < \infty$, and satisfies

$$\lambda = \inf\left\{ \frac{1}{r} x_r : r \geq 1 \right\}.$$

The history here is that the existence of exponents such as $\phi(p)$ and $\zeta(p)$ was shown using the subadditive inequality, and their positiveness was obtained under extra hypotheses. These extra hypotheses were then shown to be implied by the assumption $p < p_c$, in important papers of Aizenman and Barsky [13] and Menshikov [268, 271]. The case $d = 2$ had been dealt with earlier by Kesten [200, 202].

As an example of the subadditive inequality in action, we present a proof of the existence of $\phi(p)$ (and other things ...). The required 'hard estimate' is given in the next section. We denote by e_1 a unit vector in the direction of increasing first coordinate.

Theorem 6.2. *Let* $0 < p < 1$. *The limits*

$$\phi_1(p) = \lim_{n\to\infty}\Big\{-\frac{1}{n}\log P_p\big(0 \leftrightarrow \partial B(n)\big)\Big\}, \tag{6.3}$$

$$\phi_2(p) = \lim_{n\to\infty}\Big\{-\frac{1}{n}\log P_p(0 \leftrightarrow ne_1)\Big\}, \tag{6.4}$$

exist and are equal.

Before proving this theorem, we introduce the important concept of 'correlation length'. Suppose that $p < p_c$. In the next section, we shall see that the common limit $\phi(p)$ in (6.3)–(6.4) is strictly positive (whereas it equals 0 when $p \ge p_c$). At a basic mathematical level, we define the subcritical *correlation length* $\xi(p)$ by

$$\xi(p) = 1/\phi(p) \quad \text{for } p < p_c. \tag{6.5}$$

The physical motivation for this definition may be expressed as follows. We begin with the following statistical question. Given certain information about the existence (or not) of long open paths in the lattice, how may we distinguish between the two hypotheses that $p = p_c$ and that $p < p_c$. In particular, on what 'length-scale' need we observe the process in order to distinguish these two possibilities? In order to be concrete, let us suppose that we are told that the event $A_n = \{0 \leftrightarrow \partial B(n)\}$ occurs. How large must n be that this information be helpful? In performing the classical statistical hypothesis test of $H_0 : p = p_c$ versus $H_1 : p = p'$, where $p' < p_c$, we will reject the null hypothesis if

$$P_{p'}(A_n) > \beta P_{p_c}(A_n) \tag{6.6}$$

where β (< 1) is chosen in order to adjust the significance level of the test.

Now $P_p(A_n)$ is 'approximately' $e^{-n\phi(p)}$, and we shall see in the next section that $\phi(p) > 0$ if and only if $p < p_c$. (The fact that $\phi(p_c) = 0$ is slightly delicate; see [G], equation (5.18).) Inequality (6.6) may therefore be written as $n\phi(p') < O(1)$, which is to say that n should be of no greater order than $\xi(p') = 1/\phi(p')$. This statistical discussion supports the loosely phrased statement that 'in order to distinguish between bond percolation at $p = p_c$ and at $p = p'$, it is necessary to observe the process over a length-scale of at least $\xi(p')$'.

The existence of the function ϕ in Theorem 6.2 will be shown using standard results associated with the subadditive inequality. When such inequalities are explored carefully (see [G], Chapter 5), they yield some smoothness of ϕ, namely that ϕ is continuous and non-increasing on $(0, 1]$, and furthermore that $\phi(p_c) = 0$. Taken together with the fact that $\chi(p) \ge \phi(p)^{-1}$ (see [27, G]), we obtain that

$$\chi(p_c) = \infty. \tag{6.7}$$

Now $\phi(p) = 0$ when $p > p_c$ (since $P_p(A_n) \ge \theta(p) > 0$). Therefore the above discussion needs more thought in this case. In defining the supercritical correlation length, it is normal to work with the 'truncated' probabilities $P_p(A_n,\ |C| < \infty)$. It may be shown ([95, 165]) that the limit

$$\phi(p) = \lim_{n\to\infty}\Big\{-\frac{1}{n}\log P_p\big(0 \leftrightarrow \partial B(n),\ |C| < \infty\big)\Big\} \tag{6.8}$$

exists for all p, and satisfies $\phi(p) > 0$ if and only if $p \neq p_c$. We now define the *correlation length* $\xi(p)$ by

$$\xi(p) = 1/\phi(p) \quad \text{for } 0 < p < 1. \tag{6.9}$$

Proof of Theorem 6.2. Define the (two-point) *connectivity function* $\tau_p(x,y) = P_p(x \leftrightarrow y)$. Using the FKG inequality,

$$\tau_p(x,y) \geq P_p\Big(\{x \leftrightarrow z\} \cap \{z \leftrightarrow y\}\Big) \geq \tau_p(x,z)\tau_p(z,y)$$

for any $z \in \mathbb{Z}^d$. Set $x = 0$, $z = me_1$, $y = (m+n)e_1$, to obtain that $\tau_p(r) = P_p(0 \leftrightarrow re_1)$ satisfies $\tau_p(m+n) \geq \tau_p(m)\tau_p(n)$. Therefore the limit $\phi_2(p)$ exists by the subadditive inequality.

The existence of $\phi_1(p)$ may be shown similarly, using the BK inequality as follows. Note that

$$\{0 \leftrightarrow \partial B(m+n)\} \subseteq \bigcup_{x \in \partial B(m)} \big\{\{0 \leftrightarrow x\} \circ \{x \leftrightarrow x + \partial B(n)\}\big\}$$

(this is geometry). Therefore $\beta_p(r) = P_p\big(0 \leftrightarrow \partial B(r)\big)$ satisfies

$$\beta_p(m+n) \leq \sum_{x \in \partial B(m)} \tau_p(0,x)\beta_p(n).$$

Now $\tau_p(0,x) \leq \beta_p(m)$ for $x \in \partial B(m)$, so that

$$\beta_p(m+n) \leq |\partial B(m)|\beta_p(m)\beta_p(n).$$

With a little ingenuity, and the subadditive inequality, we deduce the existence of $\phi_1(p)$ in (6.3). That $\phi_2(p) \geq \phi_1(p)$ follows from the fact that $\tau_p(0, ne_1) \leq \beta_p(n)$. For the converse inequality, pick $x \in \partial B(n)$ such that

$$\tau_p(0,x) \geq \frac{1}{|\partial B(n)|}\beta_p(n),$$

and assume that $x_1 = +n$. Now

$$\tau_p(0, 2ne_1) \geq P_p\Big(\{0 \leftrightarrow x\} \cap \{x \leftrightarrow 2ne_1\}\Big) \geq \tau_p(0,x)^2$$

by the FKG inequality. □

6.2 Exponential Decay

The target of this section is to prove exponential decay for connectivity functions when $p < p_c$, i.e., that the common limit $\phi(p)$ in (6.3)–(6.4) is strictly positive when $0 < p < p_c$.

Theorem 6.10. *There exists $\psi(p)$, satisfying $\psi(p) > 0$ when $0 < p < p_c$, such that*

$$P_p\big(0 \leftrightarrow \partial B(n)\big) \leq e^{-n\psi(p)} \quad \textit{for all } n. \tag{6.11}$$

It is straightforward to obtain inequality (6.11) with some $\psi(p)$ which is strictly positive when $p < (2d-1)^{-1}$; just follow the proof of (3.4). The problem is to extend the conclusion from 'small positive p' to 'all subcritical values of p'. Such a difficulty is canonical: one may often obtain estimates valid for sufficiently small (resp. large) p, but one may require such estimates all the way up to (resp. down to) the critical value p_c.

We prove Theorem 6.10 via Menshikov's method [268, 271] rather than that of Aizenman–Barsky [13]. The proof given below is essentially a reproduction of that given in [G], but with the correction of a minor error on page 50 of [G]. The equation, theorem, and figure numbers are taken unchanged from [G], pages 47–56[1]. It is a minor convenience here to work with the ball $S(n) = \{x \in \mathbb{Z}^d : \delta(0,x) \leq n\}$ containing all points within graph-theoretic distance n of the origin. Note that $S(n)$ is a 'diamond' (see the forthcoming figure labelled Fig. 3.1), and write $A_n = \{0 \leftrightarrow \partial S(n)\}$.

(The remainder of this section is extracted largely from [G])

Let $S(n,x)$ be the ball of radius n with centre at the vertex x, and let $\partial S(n,x)$ be the surface of $S(n,x)$; thus $S(n,x) = x + S(n)$ and $\partial S(n,x) = x + \partial S(n)$. Similarly, let $A_n(x)$ be the event that there is an open path from the vertex x to some vertex in $\partial S(n,x)$. We are concerned with the probabilities

$$g_p(n) = P_p(A_n) = P_p\big(A_n(x)\big) \qquad \text{for any } x.$$

Now A_n is an increasing event which depends on the edges joining vertices in $S(n)$ only. We apply Russo's formula to $P_p(A_n)$ to obtain

$$g_p'(n) = E_p\big(N(A_n)\big) \tag{3.9}$$

where the prime denotes differentiation with respect to p, and $N(A_n)$ is the number of edges which are pivotal for A_n. It follows as in (2.29)[2] that

$$\begin{aligned} g_p'(n) &= \frac{1}{p} E_p\big(N(A_n); A_n\big) \\ &= \frac{1}{p} E_p\big(N(A_n) \mid A_n\big) g_p(n) \end{aligned}$$

[1] Reproduced with the kind permission of Springer Verlag, which holds the copyright.

[2] See Theorem 4.2 of the current lecture notes.

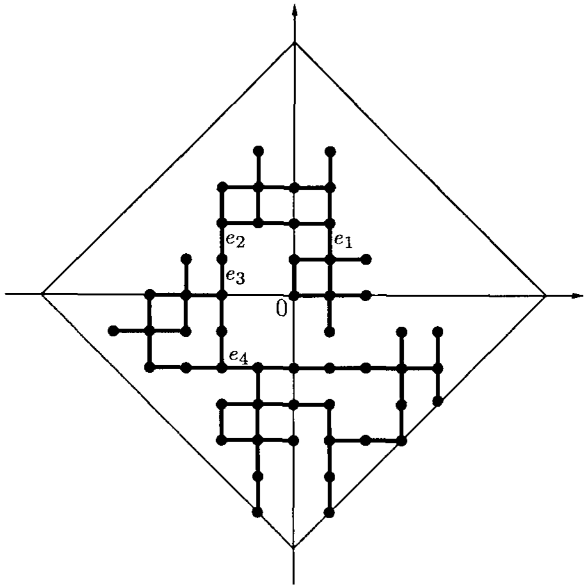


Fig. 3.1. A picture of the open cluster of $S(7)$ at the origin. There are exactly four pivotal edges for A_n in this configuration, and these are labelled e_1, e_2, e_3, e_4.

so that

$$\frac{1}{g_p(n)}\, g_p'(n) = \frac{1}{p}\, E_p\big(N(A_n) \mid A_n\big). \tag{3.10}$$

Let $0 \le \alpha < \beta \le 1$, and integrate (3.10) from $p = \alpha$ to $p = \beta$ to obtain

$$\begin{aligned} g_\alpha(n) &= g_\beta(n) \exp\left(- \int_\alpha^\beta \frac{1}{p}\, E_p\big(N(A_n) \mid A_n\big) dp \right) \\ &\le g_\beta(n) \exp\left(- \int_\alpha^\beta E_p\big(N(A_n) \mid A_n\big) dp \right), \end{aligned} \tag{3.11}$$

as in (2.30). We need now to show that $E_p\big(N(A_n) \mid A_n\big)$ grows roughly linearly in n when $p < p_c$, and then this inequality will yield an upper bound for $g_\alpha(n)$ of the form required in (3.5). The vast majority of the work in the proof is devoted to estimating $E_p\big(N(A_n) \mid A_n\big)$, and the argument is roughly as follows. If $p < p_c$ then $P_p(A_n) \to 0$ as $n \to \infty$, so that for large n we are conditioning on an event of small probability. If A_n occurs, 'but only just', then the connections between the origin and $\partial S(n)$ must be sparse; indeed, there must exist many open edges in $S(n)$ which are crucial for the occurrence of A_n (see Figure 3.1). It is plausible that the number of such pivotal edges in paths from the origin to $\partial S(2n)$ is approximately twice the number of such edges in paths to $\partial S(n)$, since these sparse paths have to traverse twice the distance. Thus the number $N(A_n)$ of edges pivotal for A_n should grow linearly in n.

Suppose that the event A_n occurs, and denote by $e_1, e_2, \ldots, e_N$ the (random) edges which are pivotal for A_n. Since A_n is increasing, each e_j has the property that A_n occurs if and only if e_j is open; thus all open paths from the origin to $\partial S(n)$ traverse e_j, for every j (see Figure 3.1). Let π be such an open path; we assume that the edges $e_1, e_2, \ldots, e_N$ have been enumerated in the order in which they are traversed by π. A glance at Figure 3.1 confirms that this ordering is independent of the choice of π. We denote by x_i the endvertex of e_i encountered first by π, and by y_i the other endvertex of e_i. We observe that there exist at least two edge-disjoint open paths joining 0 to x_1, since, if two such paths cannot be found then, by Menger's theorem (Wilson 1979[3], p. 126), there exists a pivotal edge in π which is encountered prior to x_1, a contradiction. Similarly, for $1 \le i < N$, there exist at least two edge-disjoint open paths joining y_i to x_{i+1}; see Figure 3.2. In the words of the discoverer of this proof, the open cluster containing the origin resembles a chain of sausages.

As before, let $M = \max\{k : A_k \text{ occurs}\}$ be the radius of the largest ball whose surface contains a vertex which is joined to the origin by an open path. We note that, if $p < p_c$, then M has a non-defective distribution in that $P_p(M \ge k) = g_p(k) \to 0$ as $k \to \infty$. We shall show that, conditional on A_n, $N(A_n)$ is at least as large as the number of renewals up to time n of a renewal process whose inter-renewal times have approximately the same distribution as M. In order to compare $N(A_n)$ with such a renewal process, we introduce the following notation. Let $\rho_1 = \delta(0, x_1)$ and $\rho_{i+1} = \delta(y_i, x_{i+1})$ for $1 \le i < N$. The first step is to show that, roughly speaking, the random variables $\rho_1, \rho_2, \ldots$ are jointly smaller in distribution than a sequence $M_1, M_2, \ldots$ of independent random variables distributed as M.

(3.12) Lemma. *Let k be a positive integer, and let $r_1, r_2, \ldots, r_k$ be non-negative integers such that $\sum_{i=1}^{k} r_i \le n - k$. Then, for $0 < p < 1$,*

$$\text{(3.13)} \quad \begin{aligned} &P_p(\rho_k \le r_k,\ \rho_i = r_i \text{ for } 1 \le i < k \mid A_n) \\ &\qquad \ge P_p(M \le r_k) P_p(\rho_i = r_i \text{ for } 1 \le i < k \mid A_n). \end{aligned}$$

Proof. Suppose by way of illustration that $k = 1$ and $0 \le r_1 < n$. Then

$$\text{(3.14)} \qquad \{\rho_1 > r_1\} \cap A_n \subseteq A_{r_1+1} \circ A_n,$$

since if $\rho_1 > r_1$ then the first endvertex of the first pivotal edge lies either outside $S(r_1 + 1)$ or on its surface $\partial S(r_1 + 1)$; see Figure 3.2. However, A_{r_1+1} and A_n are increasing events which depend on the edges within $S(n)$ only, and the BK inequality yields

$$P_p(\{\rho_1 > r_1\} \cap A_n) \le P_p(A_{r_1+1}) P_p(A_n).$$

We divide by $P_p(A_n)$ to obtain

$$P_p(\rho_1 > r_1 \mid A_n) \le g_p(r_1 + 1);$$

however $P_p(M \ge m) = g_p(m)$, and thus we have obtained (3.13) in the case $k = 1$.

[3]Reference [359].

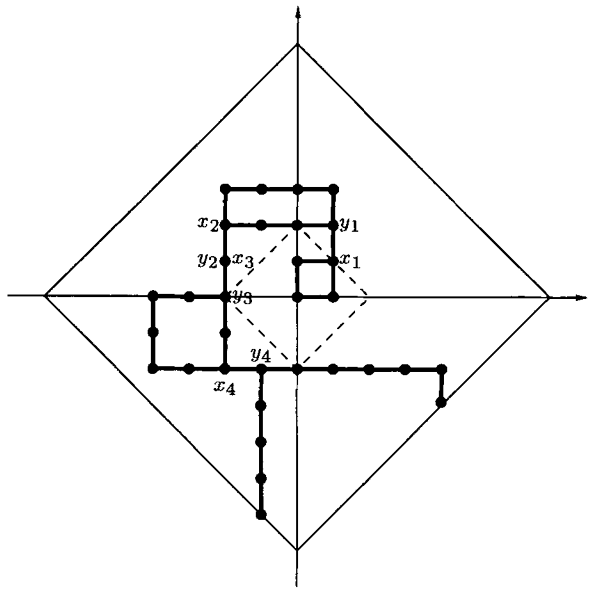


Fig. 3.2. The pivotal edges are $e_i = \langle x_i, y_i \rangle$ for $i = 1, 2, 3, 4$. Note that $x_3 = y_2$ in this configuration. The dashed line is the surface $\partial S(\rho_1)$ of $S(\rho_1)$. Note the two edge-disjoint paths from the origin to $\partial S(\rho_1)$.

We now prove the lemma for general values of k. Suppose that $k \geq 1$, and let $r_1, r_2, \ldots, r_k$ be non-negative integers with sum not exceeding $n - k$. Let N be the number of edges which are pivotal for A_n; we enumerate and label these edges as $e_i = \langle x_i, y_i \rangle$ as before.

(*The following section in italics replaces an incorrect passage in* [G].)

For any edge $e = \langle u, v \rangle$, let D_e be the set of vertices attainable from 0 *along open paths not using e, together with all open edges between such vertices. Let B_e be the event that the following statements hold:*

(a) *exactly one of u or v lies in D_e, say u,*

(b) *e is open,*

(c) *D_e contains no vertex of $\partial S(n)$,*

(d) *the pivotal edges for the event $\{0 \leftrightarrow v\}$ are, taken in order, $\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \ldots, \langle x_{k-2}, y_{k-2} \rangle, \langle x_{k-1}, y_{k-1} \rangle = e$, where $\delta(y_{i-1}, x_i) = r_i$ for $1 \leq i < k$, and $y_0 = 0$.*

We now define the event $B = \bigcup_e B_e$. For $\omega \in A_n \cap B$, there is a unique $e = e(\omega)$ such that B_e occurs.

For $\omega \in B$, we consider the set of vertices and open edges attainable along open paths from the origin without using $e = e(\omega)$; to this graph we append e and its other endvertex $v = y_{k-1}$, and we place a mark over y_{k-1} in order to distinguish it from the other vertices. We denote by $G = D_e$ the resulting (marked) graph, and we write $y(G)$ for the unique marked vertex of G. We condition on G to obtain

$$P_p(A_n \cap B) = \sum_{\Gamma} P_p(B, G = \Gamma) P_p(A_n \mid B, G = \Gamma),$$

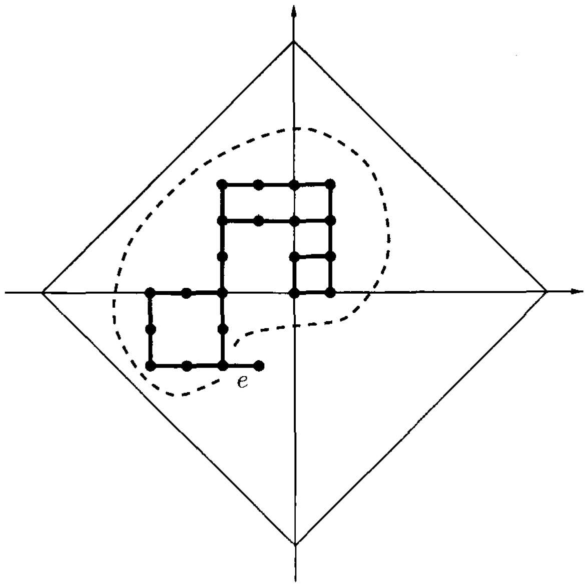


Fig. 3.3. A sketch of the event B_e. The dashed line indicates that the only open 'exit' from the interior is via the edge e. Note the existence of 3 pivotal edges for the event that 0 is connected to an endvertex of e.

where the sum is over all possible values Γ of G. The final term in this summation is the probability that $y(\Gamma)$ is joined to $\partial S(n)$ by an open path which has no vertex other than $y(\Gamma)$ in common with Γ. Thus, in the obvious terminology,

$$P_p(A_n \cap B) = \sum_{\Gamma} P_p(B, G = \Gamma) P_p\big(y(\Gamma) \leftrightarrow \partial S(n) \text{ off } \Gamma\big). \tag{3.15}$$

Similarly,

$$\begin{aligned} &P_p(\{\rho_k > r_k\} \cap A_n \cap B) \\ &\quad = \sum_{\Gamma} P_p(B, G = \Gamma) P_p(\{\rho_k > r_k\} \cap A_n \mid B, G = \Gamma) \\ &\quad = \sum_{\Gamma} P_p(B, G = \Gamma) \\ &\qquad \times P_p\Big(\big\{y(\Gamma) \leftrightarrow \partial S(r_k + 1, y(\Gamma)) \text{ off } \Gamma\big\} \circ \big\{y(\Gamma) \leftrightarrow \partial S(n) \text{ off } \Gamma\big\}\Big). \end{aligned}$$

We apply the BK inequality to the last term to obtain

(3.16)

$$\begin{aligned} &P_p(\{\rho_k > r_k\} \cap A_n \cap B) \\ &\quad \le \sum_{\Gamma} P_p(B, G = \Gamma) P_p\big(y(\Gamma) \leftrightarrow \partial S(n) \text{ off } \Gamma\big) P_p\Big(y(\Gamma) \leftrightarrow \partial S\big(r_k + 1, y(\Gamma)\big) \text{ off } \Gamma\Big) \\ &\quad \le g_p(r_k + 1) P_p(A_n \cap B) \end{aligned}$$

by (3.15) and the fact that, for each possible Γ,

$$\begin{aligned} P_p\Big(y(\Gamma) \leftrightarrow \partial S\big(r_k+1, y(\Gamma)\big) \text{ off } \Gamma\Big) &\le P_p\Big(y(\Gamma) \leftrightarrow \partial S\big(r_k+1, y(\Gamma)\big)\Big) \\ &= P_p(A_{r_k+1}) \\ &= g_p(r_k+1). \end{aligned}$$

We divide each side of (3.16) by $P_p(A_n \cap B)$ to obtain

$$P_p(\rho_k \le r_k \mid A_n \cap B) \ge 1 - g_p(r_k+1),$$

throughout which we multiply by $P_p(B \mid A_n)$ to obtain the result. □

(3.17) Lemma. *For $0 < p < 1$, it is the case that*

$$E_p\big(N(A_n) \mid A_n\big) \ge \frac{n}{\sum_{i=0}^{n} g_p(i)} - 1. \tag{3.18}$$

Proof. It follows from Lemma (3.12) that

$$P_p(\rho_1 + \rho_2 + \cdots + \rho_k \le n-k \mid A_n) \ge P(M_1 + M_2 + \cdots + M_k \le n-k), \tag{3.19}$$

where $k \ge 1$ and $M_1, M_2, \dots$ is a sequence of independent random variables distributed as M. We defer until the end of this proof the minor chore of deducing (3.19) from (3.13). Now $N(A_n) \ge k$ if $\rho_1 + \rho_2 + \cdots + \rho_k \le n-k$, so that

$$P_p\big(N(A_n) \ge k \mid A_n\big) \ge P(M_1 + M_2 + \cdots + M_k \le n-k). \tag{3.20}$$

A minor difficulty is that the M_i may have a defective distribution. Indeed,

$$\begin{aligned} P(M \ge r) &= g_p(r) \\ &\to \theta(p) \qquad \text{as } r \to \infty; \end{aligned}$$

thus we allow the M_i to take the value ∞ with probability $\theta(p)$. On the other hand, we are not concerned with atoms at ∞, since

$$P(M_1 + M_2 + \cdots + M_k \le n-k) = P(M_1' + M_2' + \cdots + M_k' \le n),$$

where $M_i' = 1 + \min\{M_i, n\}$, and we work henceforth with these truncated random variables. Summing (3.20) over k, we obtain

$$\begin{aligned} E_p\big(N(A_n) \mid A_n\big) &\ge \sum_{k=1}^{\infty} P(M_1' + M_2' + \cdots + M_k' \le n) \\ &= \sum_{k=1}^{\infty} P(K \ge k+1) \\ &= E(K) - 1, \end{aligned} \tag{3.21}$$

where $K = \min\{k : M_1' + M_2' + \cdots + M_k' > n\}$. Let $S_k = M_1' + M_2' + \cdots + M_k'$, the sum of independent, identically distributed, bounded random variables. By Wald's equation (see Chow and Teicher 1978[4], pp. 137, 150),

$$n < E(S_K) = E(K)E(M_1'),$$

giving that

$$E(K) > \frac{n}{E(M_1')} = \frac{n}{1 + E(\min\{M_1, n\})} = \frac{n}{\sum_{i=0}^{n} g_p(i)}$$

since

$$E(\min\{M_1, n\}) = \sum_{i=1}^{n} P(M \geq i) = \sum_{i=1}^{n} g_p(i).$$

It remains to show that (3.19) follows from Lemma (3.12). We have that

$$\begin{aligned}
&P_p(\rho_1 + \rho_2 + \cdots + \rho_k \leq n - k \mid A_n) \\
&\quad = \sum_{i=0}^{n-k} P_p(\rho_1 + \rho_2 + \cdots + \rho_{k-1} = i,\ \rho_k \leq n - k - i \mid A_n) \\
&\quad \geq \sum_{i=0}^{n-k} P(M \leq n - k - i) P_p(\rho_1 + \rho_2 + \cdots + \rho_{k-1} = i \mid A_n) \quad \text{by (3.13)} \\
&\quad = P_p(\rho_1 + \rho_2 + \cdots + \rho_{k-1} + M_k \leq n - k \mid A_n),
\end{aligned}$$

where M_k is a random variable which is independent of all edge-states in $S(n)$ and is distributed as M. There is a mild abuse of notation here, since P_p is not the correct probability measure unless M_k is measurable on the usual σ-field of events, but we need not trouble ourselves overmuch about this. We iterate the above argument in the obvious way to deduce (3.19), thereby completing the proof of the lemma. □

The conclusion of Theorem (3.8) is easily obtained from this lemma, but we delay this step until the end of the section. The proof of Theorem (3.4) proceeds by substituting (3.18) into (3.11) to obtain that, for $0 \leq \alpha < \beta \leq 1$,

$$g_\alpha(n) \leq g_\beta(n) \exp\left(- \int_\alpha^\beta \left[\frac{n}{\sum_{i=0}^{n} g_p(i)} - 1 \right] dp \right).$$

It is difficult to calculate the integral in the exponent, and so we use the inequality $g_p(i) \leq g_\beta(i)$ for $p \leq \beta$ to obtain

$$g_\alpha(n) \leq g_\beta(n) \exp\left(-(\beta - \alpha) \left[\frac{n}{\sum_{i=0}^{n} g_\beta(i)} - 1 \right] \right), \tag{3.22}$$

[4]Reference [107].

and it is from this relation that the conclusion of Theorem (3.4) will be extracted. Before continuing, it is interesting to observe that by combining (3.10) and (3.18) we obtain a differential-difference inequality involving the function

$$G(p,n) = \sum_{i=0}^{n} g_p(i);$$

rewriting this equation rather informally as a partial differential inequality, we obtain

$$\frac{\partial^2 G}{\partial p\, \partial n} \geq \frac{\partial G}{\partial n}\left(\frac{n}{G} - 1\right). \tag{3.23}$$

Efforts to integrate this inequality directly have failed so far.

Once we know that

$$E_\beta(M) = \sum_{i=1}^{\infty} g_\beta(i) < \infty \qquad \text{for all } \beta < p_c,$$

then (3.22) gives us that

$$g_\alpha(n) \leq e^{-n\psi(\alpha)} \qquad \text{for all } \alpha < p_c,$$

for some $\psi(\alpha) > 0$, as required. At the moment we know rather less than the finite summability of the $g_p(i)$ for $p < p_c$, knowing only that $g_p(i) \to 0$ as $i \to \infty$. In order to estimate the rate at which $g_p(i) \to 0$, we shall use (3.22) as a mathematical turbocharger.

(3.24) Lemma. *For $p < p_c$, there exists $\delta(p)$ such that*

$$g_p(n) \leq \delta(p) n^{-1/2} \qquad \textit{for } n \geq 1. \tag{3.25}$$

Once this lemma has been proved, the theorem follows quickly. To see this, note that (3.25) implies the existence of $\Delta(p) < \infty$ such that

$$\sum_{i=0}^{n} g_p(i) \leq \Delta(p) n^{1/2} \qquad \text{for } p < p_c. \tag{3.26}$$

Let $\alpha < p_c$, and find β such that $\alpha < \beta < p_c$. Substitute (3.26) with $p = \beta$ into (3.22) to find that

$$\begin{aligned} g_\alpha(n) &\leq g_\beta(n) \exp\left\{-(\beta - \alpha)\left(\frac{n^{1/2}}{\Delta(\beta)} - 1\right)\right\} \\ &\leq \exp\left\{1 - \frac{(\beta - \alpha)}{\Delta(\beta)} n^{1/2}\right\}. \end{aligned}$$

Thus

$$\sum_{n=1}^{\infty} g_\alpha(n) < \infty \qquad \text{for } \alpha < p_c,$$

and the theorem follows from the observations made prior to the statement of Lemma (3.24). We shall now prove this lemma.

Proof. First, we shall show the existence of a subsequence $n_1, n_2, \ldots$ along which $g_p(n)$ approaches 0 rather quickly; secondly, we shall fill in the gaps in this subsequence.

Fix $\beta < p_c$ and a positive integer n. Let α satisfy $0 < \alpha < \beta$ and let $n' \geq n$; later we shall choose α and n' explicitly in terms of β and n. From (3.22),

$$\begin{aligned} g_\alpha(n') &\leq g_\beta(n') \exp\left(1 - \frac{n'(\beta - \alpha)}{\sum_{i=0}^{n'} g_\beta(i)}\right) \\ &\leq g_\beta(n) \exp\left(1 - \frac{n'(\beta - \alpha)}{\sum_{i=0}^{n'} g_\beta(i)}\right) \end{aligned} \tag{3.27}$$

since $n \leq n'$. We wish to write the exponent in terms of $g_\beta(n)$, and to this end we shall choose n' appropriately. We split the summation into two parts corresponding to $i < n$ and $i \geq n$, and we use the monotonicity of $g_\beta(i)$ to find that

$$\begin{aligned} \frac{1}{n'} \sum_{i=0}^{n'} g_\beta(i) &\leq \frac{1}{n'} \{n g_\beta(0) + n' g_\beta(n)\} \\ &\leq 3 g_\beta(n) \qquad \text{if } n' \geq n \lfloor g_\beta(n)^{-1} \rfloor. \end{aligned}$$

We now define

$$n' = n\gamma_\beta(n) \qquad \text{where } \gamma_\beta(n) = \lfloor g_\beta(n)^{-1} \rfloor \tag{3.28}$$

and deduce from (3.27) that

$$g_\alpha(n') \leq g_\beta(n) \exp\left(1 - \frac{\beta - \alpha}{3 g_\beta(n)}\right). \tag{3.29}$$

Next we choose α by setting

$$\beta - \alpha = 3 g_\beta(n)\{1 - \log g_\beta(n)\}. \tag{3.30}$$

Now $g_\beta(m) \to 0$ as $m \to \infty$, so that $0 < \alpha < \beta$ if n has been picked large enough; (3.29) then yields

$$g_\alpha(n') \leq g_\beta(n)^2. \tag{3.31}$$

This conclusion is the basic recursion step which we shall use repeatedly. We have shown that, for $\beta < p_c$, there exists $n_0(\beta)$ such that (3.31) holds for all $n \geq n_0(\beta)$ whenever n' and α are given by (3.28) and (3.30), respectively.

Next, we fix $p < p_c$ and choose π such that $p < \pi < p_c$. We now construct sequences $(p_i : i \geq 0)$ of probabilities and $(n_i : i \geq 0)$ of integers as follows. We set

$p_0 = \pi$ and shall pick n_0 later. Having found $p_0, p_1, \dots, p_i$ and $n_0, n_1, \dots, n_i$, we define

$$n_{i+1} = n_i \gamma_i \quad \text{and} \quad p_i - p_{i+1} = 3g_i(1 - \log g_i) \tag{3.32}$$

where $g_i = g_{p_i}(n_i)$ and $\gamma_i = \lfloor g_i^{-1} \rfloor$. We note that $n_i \leq n_{i+1}$ and $p_i > p_{i+1}$. The recursion (3.32) is valid so long as $p_{i+1} > 0$, and this is indeed the case so long as n_0 has been chosen to be sufficiently large. To see this we argue as follows. From the definition of $p_0, \dots, p_i$ and $n_0, \dots, n_i$ and the discussion leading to (3.31), we find that

$$g_{j+1} \leq g_j^2 \quad \text{for } j = 0, 1, \dots, i-1. \tag{3.33}$$

If a real sequence $(x_j : j \geq 0)$ satisfies $0 < x_0 < 1$, $x_{j+1} = x_j^2$ for $j \geq 0$, then it is easy to check that

$$s(x_0) = \sum_{j=0}^{\infty} 3x_j(1 - \log x_j) < \infty,$$

and furthermore that $s(x_0) \to 0$ as $x_0 \to 0$. We may pick x_0 sufficiently small such that

$$s(x_0) \leq \pi - p \tag{3.34}$$

and then we pick n_0 sufficiently large that $g_0 = g_\pi(n_0) < x_0$. Now $h(x) = 3x(1 - \log x)$ is an increasing function on $[0, x_0]$, giving from (3.32) and (3.33) that

$$\begin{aligned}
p_{i+1} &= p_i - 3g_i(1 - \log g_i) \\
&= \pi - \sum_{j=0}^{i} 2g_j(1 - \log g_j) \\
&\geq \pi - \sum_{j=0}^{\infty} 3x_j(1 - \log x_j) \\
&\geq p \qquad \text{by (3.34).}
\end{aligned}$$

Thus, by a suitable choice of n_0 we may guarantee not only that $p_{i+1} > 0$ for all i but also that

$$\widetilde{p} = \lim_{i \to \infty} p_i$$

satisfies $\widetilde{p} \geq p$. Let us suppose that n_0 has been chosen accordingly, so that the recursion (3.32) is valid and $\widetilde{p} \geq p$. We have from (3.32) and (3.33) that

$$n_k = n_0 \gamma_0 \gamma_1 \cdots \gamma_{k-1} \qquad \text{for } k \geq 1$$

and

$$\begin{aligned}
g_{k-1}^2 &= g_{k-1} g_{k-1} \\
&\leq g_{k-1} g_{k-2}^2 \leq \cdots \\
&\leq g_{k-1} g_{k-2} \cdots g_1 g_0^2 \\
&\leq (\gamma_{k-1} \gamma_{k-2} \cdots \gamma_0)^{-1} g_0 \\
&= \delta^2 n_k^{-1},
\end{aligned} \tag{3.35}$$

where $\delta^2 = n_0 g_0$.

We are essentially finished. Let $n > n_0$, and find an integer k such that $n_{k-1} \leq n < n_k$; this is always possible since $g_k \to 0$ as $k \to \infty$, and therefore $n_{k-1} < n_k$ for all large k. Then

$$\begin{aligned} g_p(n) &\leq g_{p_{k-1}}(n_{k-1}) && \text{since } p \leq p_{k-1} \\ &= g_{k-1} \\ &\leq \delta n_k^{-1/2} && \text{by (3.35)} \\ &\leq \delta n^{-1/2} && \text{since } n < n_k \end{aligned}$$

as required. This is valid for $n > n_0$, but we may adjust the constant δ so that a similar inequality is valid for all $n \geq 1$. □

6.3 Ornstein–Zernike Decay

The connectivity function $\tau_p(x, y) = P_p(x \leftrightarrow y)$ decays exponentially when $p < p_c$, which is to say that the limits

$$\phi(p, x) = \lim_{n\to\infty} \left\{ -\frac{1}{n} \log \tau_p(0, nx) \right\} \tag{6.12}$$

exist and satisfy $\phi(p, x) > 0$ for $0 < p < p_c$ and $x \in \mathbb{Z}^d \setminus \{0\}$ (cf. Theorem 6.2).

In one direction, this observation may lead to a study of the function $\phi(p, \cdot)$. In another, one may ask for finer asymptotics in (6.12). We concentrate on the case $x = e_1$, and write $\phi(p) = \phi(p, e_1)$.

Theorem 6.13 (Ornstein–Zernike Decay). *Suppose that $0 < p < p_c$. There exists a positive function $A(p)$ such that*

$$\tau_p(0, ne_1) = \big(1 + \mathrm{O}(n^{-1})\big) \frac{A(p)}{n^{\frac{1}{2}(d-1)}} e^{-n\phi(p)} \quad \textit{as } n \to \infty.$$

The correction factor $n^{-\frac{1}{2}(d-1)}$ occurs similarly in many other disordered systems, as was proposed by Ornstein and Zernike [301]. Theorem 6.13, and certain extensions, was obtained for percolation by Campanino, Chayes, and Chayes [86].

7. SUPERCRITICAL PERCOLATION

7.1 Uniqueness of the Infinite Cluster

Let I be the number of infinite open clusters.

Theorem 7.1. *For any p, either $P_p(I = 0) = 1$ or $P_p(I = 1) = 1$.*

This result was proved first in [21], then more briefly in [146], and the definitive proof of Burton and Keane [83] appeared shortly afterwards. This last proof is short and elegant, and relies only on the zero–one law and a little geometry.

Proof. Fix $p \in [0, 1]$. The sample space $\Omega = \{0, 1\}^{\mathbb{E}}$ is a product space with a natural family of translations inherited from the translations of the lattice $\mathbb{L}^d$. Furthermore, P_p is a product measure on Ω. Since I is a translation-invariant function on Ω, it is a.s. constant, which is to say that

$$\text{there exists } k \in \{0, 1, \dots\} \cup \{\infty\} \text{ such that } P_p(I = k) = 1. \tag{7.2}$$

Naturally, the value of k depends on the choice of p. Next we show that the k in question satisfies $k \in \{0, 1, \infty\}$. Suppose (7.2) holds with some k satisfying $2 \le k < \infty$. We may find a box B sufficiently large that

$$P_p(B \text{ intersects two or more infinite clusters}) > \tfrac{1}{2}. \tag{7.3}$$

By changing the states of edges in B (by making all such edges open, say) we can decrease the number of infinite clusters (on the event in (7.3)). Therefore $P_p(I = k - 1) > 0$, in contradiction of (7.2). Therefore we cannot have $2 \le k < \infty$ in (7.2).

It remains to rule out the case $k = \infty$. Suppose that $k = \infty$. We will derive a contradiction by using a geometrical argument. We call a vertex x a *trifurcation* if:

(a) x lies in an infinite open cluster, and
(b) the deletion of x splits this infinite cluster into exactly three disjoint infinite clusters and no finite clusters,

and we denote by T_x the event that x is a trifurcation. Now $P_p(T_x)$ is constant for all x, and therefore

$$\frac{1}{|B(n)|} E_p\Biggl(\sum_{x \in B(n)} 1_{T_x} \Biggr) = P_p(T_0). \tag{7.4}$$

(Recall that 1_A denotes the indicator function of an event A.) It is useful to know that the quantity $P_p(T_0)$ is strictly positive, and it is here that we use the assumed multiplicity of infinite clusters. Since $P_p(I = \infty) = 1$ by assumption, we may find a box $B(n)$ sufficiently large that it intersects at least three distinct infinite clusters with probability at least $\frac{1}{2}$. By changing the configuration inside $B(n)$, we

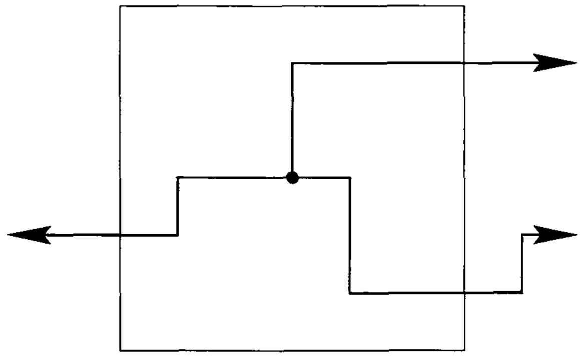


Fig. 7.1. Take a box B which intersects at least three distinct infinite open clusters, and then alter the configuration inside B in order to create a configuration in which 0 is a trifurcation.

may turn the origin into a trifurcation (see Figure 7.1). The corresponding set of configurations has strictly positive probability, so that $P_p(T_0) > 0$ in (7.4).

Before turning to the geometry, we present a lemma concerning partitions. Let Y be a finite set with $|Y| \geq 3$, and define a *3-partition* $\Pi = \{P_1, P_2, P_3\}$ of Y to be a partition of Y into exactly three non-empty sets P_1, P_2, P_3. For 3-partitions $\Pi = \{P_1, P_2, P_3\}$ and $\Pi' = \{P_1', P_2', P_3'\}$, we say that Π and Π' are *compatible* if there exists an ordering of their elements such that $P_1 \supseteq P_2' \cup P_3'$ (or, equivalently, that $P_1' \supseteq P_2 \cup P_3$). A collection $\mathcal{P}$ of 3-partitions is *compatible* if each pair therein is compatible.

Lemma 7.5. *If $\mathcal{P}$ is a compatible family of distinct 3-partitions of Y, then* $|\mathcal{P}| \leq |Y| - 2$.

Proof. There are several ways of doing this; see [83]. For any set $\mathcal{Q}$ of distinct compatible 3-partitions of Y, we define an equivalence relation $\sim$ on Y by $x \sim y$ if, for all $\Pi \in \mathcal{Q}$, x and y lie in the same element of Π. Write $\alpha(\mathcal{Q})$ for the number of equivalence classes of $\sim$. Now, write $\mathcal{P} = (\Pi_1, \Pi_2, \dots, \Pi_m)$ in some order, and let $\alpha_k = \alpha(\Pi_1, \Pi_2, \dots, \Pi_k)$. Evidently $\alpha_1 = 3$ and, using the compatibility of Π_1 and Π_2, we have that $\alpha_2 \geq 4$. By comparing Π_{r+1} with $\Pi_1, \Pi_2, \dots, \Pi_r$ in turn, and using their compatibility, one sees that $\alpha(\Pi_1, \Pi_2, \dots, \Pi_{r+1}) \geq \alpha(\Pi_1, \Pi_2, \dots, \Pi_r)+1$, whence $\alpha_m \geq \alpha_1 + (m-1) = m+2$. However $\alpha_m \leq |Y|$, and the claim of the lemma follows. $\square$

Let K be a connected open cluster of $B(n)$, and write $\partial K = K \cap \partial B(n)$. If x ($\in B(n-1)$) is a trifurcation in K, then the removal of x induces a 3-partition $\Pi_K(x) = \{P_1, P_2, P_3\}$ of ∂K with the properties that

(a) P_i is non-empty, for $i = 1, 2, 3$,

(b) P_i is a subset of a connected subgraph of $B(n) \setminus \{x\}$,

(c) $P_i \nleftrightarrow P_j$ in $B(n)$, if $i \neq j$.

Furthermore, if x and x' are distinct trifurcations of $K \cap B(n-1)$, then $\Pi_K(x)$ and $\Pi_K(x')$ are distinct and compatible; see Figure 7.2.

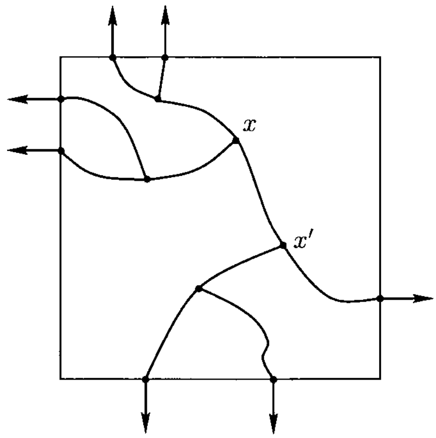


Fig. 7.2. Two trifurcations x and x' belonging to a cluster K of $B(n)$. They induce compatible partitions of ∂K.

It follows by Lemma 7.5 that the number $T(K)$ of trifurcations in $K \cap B(n-1)$ satisfies

$$T(K) \le |\partial K| - 2.$$

We sum this inequality over all connected clusters of $B(n)$, to obtain that

$$\sum_{x \in B(n-1)} 1_{T_x} \le |\partial B(n)|.$$

Take expectations, and use (7.4) to find that

$$|B(n-1)| P_p(T_0) \le |\partial B(n)|,$$

which is impossible for large n since the left side grows as n^d and the right side as n^{d-1}. This contradiction completes the proof. □

7.2 Percolation in Slabs

Many results were proved for subcritical percolation under the hypothesis of 'finite susceptibility', i.e., that $\chi(p) = E_p|C|$ satisfies $\chi(p) < \infty$. Subsequently, it was proved in [13, 268, 271] that this hypothesis is satisfied whenever $p < p_c$. The situation was similar for supercritical percolation, the corresponding hypothesis being that percolation occurs in slabs. We define the slab of thickness k by

$$S_k = \mathbb{Z}^{d-1} \times \{0, 1, \ldots, k\},$$

with critical probability $p_c(S_k)$; we assume here that $d \ge 3$. The decreasing limit $p_c(S) = \lim_{k \to \infty} p_c(S_k)$ exists, and satisfies $p_c(S) \ge p_c$. The hypothesis of 'percolation in slabs' is that $p > p_c(S)$. Here is an example of the hypothesis in action (cf. Theorem 6.2 and equation (6.8)).

Theorem 7.6. *The limit*

$$\sigma(p) = \lim_{n\to\infty}\left\{-\frac{1}{n}\log P_p(0 \leftrightarrow \partial B(n),\ |C| < \infty)\right\}$$

exists. Furthermore $\sigma(p) > 0$ *if* $p > p_c(S)$.

This theorem asserts the exponential decay of a 'truncated' connectivity function when $d \geq 3$. Corresponding results when $d = 2$ may be proved using duality.

Proof. The existence of the limit is an exercise in subadditivity (see [95, G]), and we sketch here only a proof that $\sigma(p) > 0$. Assume that $p > p_c(S)$, so that $p > p_c(S_k)$ for some k; choose k accordingly. Let H_n be the hyperplane containing all vertices x with $x_1 = n$. It suffices to prove that

$$P_p(0 \leftrightarrow H_n,\ |C| < \infty) \leq e^{-\gamma n} \tag{7.7}$$

for some $\gamma = \gamma(p) > 0$. Define the slabs

$$T_i = \{x \in \mathbb{Z}^d : (i-1)k \leq x_1 < ik\}, \quad 1 \leq i < \lfloor n/k \rfloor.$$

Any path from 0 to H_n must traverse every such slab. Since $p > p_c(S_k)$, each slab a.s. contains an infinite open cluster. If $0 \leftrightarrow H_n$ and $|C| < \infty$, then all paths from 0 to H_n must evade all such clusters. There are $\lfloor n/k \rfloor$ slabs to traverse, and a price is paid for each. With a touch of rigour, this argument implies that

$$P_p(0 \leftrightarrow H_n,\ |C| < \infty) \leq \{1 - \theta_k(p)\}^{\lfloor n/k \rfloor}$$

where

$$\theta_k(p) = P_p(0 \leftrightarrow \infty \text{ in } S_k) > 0.$$

For more details, see [G]. □

Grimmett and Marstrand [165] proved that $p_c = p_c(S)$, using ideas similar to those of [49, 50]. This was achieved via a 'block construction' which appears to be central to a full understanding of supercritical percolation and to have further applications elsewhere. The details are presented next.

7.3 Limit of Slab Critical Points

Material in this section is taken from [165]. *We assume that* $d \geq 3$ and that p is such that $\theta(p) > 0$; under this hypothesis, we wish to gain some control of the (a.s.) unique open cluster. In particular we shall prove the following theorem, in which $p_c(A)$ denotes the critical value of bond percolation on the subgraph of $\mathbb{Z}^d$ induced by the vertex set A. In this notation, $p_c = p_c(\mathbb{Z}^d)$.

Theorem 7.8. *If F is an infinite connected subset of $\mathbb{Z}^d$ with $p_c(F) < 1$, then for each $\eta > 0$ there exists an integer k such that*

$$p_c\big(2kF + B(k)\big) \le p_c + \eta.$$

Choosing $F = \mathbb{Z}^2 \times \{0\}^{d-2}$, we have that $2kF + B(k) = \{x \in \mathbb{Z}^d : -k \le x_j \le k \text{ for } 3 \le j \le d\}$. The theorem implies that $p_c(2kF + B(k)) \to p_c$ as $k \to \infty$, which is a stronger statement than the statement that $p_c = p_c(S)$.

In the remainder of this section, we sketch the salient features of the block construction necessary to prove the above theorem. This construction may be used directly to obtain further information concerning supercritical percolation.

The main idea involves working with a 'block lattice' each point of which represents a large box of $\mathbb{L}^d$, these boxes being disjoint and adjacent. In this block lattice, we declare a vertex to be 'open' if there exist certain open paths in and near the corresponding box of $\mathbb{L}^d$. We shall show that, with positive probability, there exists an infinite path of open vertices in the block lattice. Furthermore, this infinite path of open blocks corresponds to an infinite open path of $\mathbb{L}^d$. By choosing sufficiently large boxes, we aim to find such a path within a sufficiently wide slab. Thus there is a probabilistic part of the proof, and a geometric part.

There are two main steps in the proof. In the first, we show the existence of long finite paths. In the second, we show how to take such finite paths and build an infinite cluster in a slab.

The principal parts of the first step are as follows. Pick p such that $\theta(p) > 0$.

1. Let $\epsilon > 0$. Since $\theta(p) > 0$, there exists m such that

   $$P_p\big(B(m) \leftrightarrow \infty\big) > 1 - \epsilon.$$

   This is elementary probability theory.
2. Let $n \ge 2m$, say, and let $k \ge 1$. We may choose n sufficiently large to ensure that, with probability at least $1 - 2\epsilon$, $B(m)$ is joined to at least k points in $\partial B(n)$.
3. By choosing k sufficiently large, we may ensure that, with probability at least $1 - 3\epsilon$, $B(m)$ is joined to some point of $\partial B(n)$, which is itself connected to a copy of $B(m)$, lying 'on' the surface $\partial B(n)$ and every edge of which is open.
4. The open copy of $B(m)$, constructed above, may be used as a 'seed' for iterating the above construction. When doing this, we shall need some control over where the seed is placed. It may be shown that every face of $\partial B(n)$ contains (with large probability) a point adjacent to some seed, and indeed many such points.

Above is the scheme for constructing long finite paths, and we turn to the second step.

5. This construction is now iterated. At each stage there is a certain (small) probability of failure. In order that there be a strictly positive probability of an infinite sequence of successes, we iterate 'in two independent directions'. With care, one may show that the construction dominates a certain supercritical site percolation process on $\mathbb{L}^2$.
6. We wish to deduce that an infinite sequence of successes entails an infinite open path of $\mathbb{L}^d$ within the corresponding slab. There are two difficulties with

this. First, since there is not total control of the positions of the seeds, the actual path in $\mathbb{L}^d$ may leave every slab. This may be overcome by a process of 'steering', in which, at each stage, we choose a seed in such a position as to compensate for any earlier deviation in space.

7. A larger problem is that, in iterating the construction, we carry with us a mixture of 'positive' and 'negative' information (of the form that 'certain paths exist' and 'others do not'). In combining events we cannot use the FKG inequality. The practical difficulty is that, although we may have an infinite sequence of successes, there will generally be breaks in any corresponding open route to ∞. This is overcome by sprinkling down a few more open edges, i.e., by working at edge-density $p + \delta$ where $\delta > 0$, rather than at p.

In conclusion, we show that, if $\theta(p) > 0$ and $\delta > 0$, then there is (with large probability) an infinite $(p + \delta)$-open path in a slice of the form

$$T_k = \{x \in \mathbb{Z}^d : 0 \le x_j \le k \text{ for } j \ge 3\}$$

where k is sufficiently large. This implies that $p + \delta > p_c(T) = \lim_{k\to\infty} p_c(T_k)$ if $p > p_c$, i.e., that $p_c \ge p_c(T)$. Since $p_c(T) \ge p_c$ by virtue of the fact that $T_k \subseteq \mathbb{Z}^d$ for all k, we may conclude that $p_c = p_c(T)$, implying also that $p_c = p_c(S)$.

Henceforth we suppose that $d = 3$; similar arguments are valid when $d > 3$. We begin with some notation and two key lemmas. As usual, $B(n) = [-n, n]^3$, and we shall concentrate on a special face of $B(n)$,

$$F(n) = \{x \in \partial B(n) : x_1 = n\},$$

and indeed on a special 'quadrant' of $F(n)$,

$$T(n) = \{x \in \partial B(n) : x_1 = n,\ x_j \ge 0 \text{ for } j \ge 2\}.$$

For $m, n \ge 1$, let

$$T(m, n) = \bigcup_{j=1}^{2m+1} \{je_1 + T(n)\}$$

where $e_1 = (1, 0, 0)$ as usual.

We call a box $x + B(m)$ a *seed* if every edge in $x + B(m)$ is open. We now set

$$K(m, n) = \Big\{x \in T(n) : \langle x, x + e_1\rangle \text{ is open, and } \\ x + e_1 \text{ lies in some seed lying within } T(m, n)\Big\}.$$

The random set $K(m, n)$ is necessarily empty if $n < 2m$.

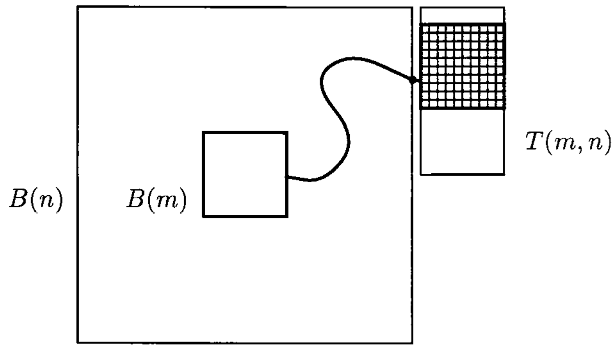


Fig. 7.3. An illustration of the event in (7.10). The hatched region is a copy of $B(m)$ all of whose edges are p-open. The central box $B(m)$ is joined by a path to some vertex in $\partial B(n)$, which is in turn connected to a seed lying on the surface of $B(n)$.

Lemma 7.9. *If $\theta(p) > 0$ and $\eta > 0$, there exists $m = m(p, \eta)$ and $n = n(p, \eta)$ such that $2m < n$ and*

$$P_p\big(B(m) \leftrightarrow K(m, n) \text{ in } B(n)\big) > 1 - \eta. \tag{7.10}$$

The event in (7.10) is illustrated in Figure 7.3.

Proof. Since $\theta(p) > 0$, there exists a.s. an infinite open cluster, whence

$$P_p\big(B(m) \leftrightarrow \infty\big) \to 1 \quad \text{as} \quad m \to \infty.$$

We pick m such that

$$P_p\big(B(m) \leftrightarrow \infty\big) > 1 - (\tfrac{1}{3}\eta)^{24}, \tag{7.11}$$

for a reason which will become clear later.

For $n > m$, let $V(n) = \{x \in T(n) : x \leftrightarrow B(m) \text{ in } B(n)\}$. Pick M such that

$$pP_p\big(B(m) \text{ is a seed}\big) > 1 - (\tfrac{1}{2}\eta)^{1/M}. \tag{7.12}$$

We shall assume for simplicity that $2m+1$ divides $n+1$ (and that $2m < n$), and we partition $T(n)$ into disjoint squares with side-length $2m$. If $|V(n)| \geq (2m+1)^2 M$ then $B(m)$ is joined in $B(n)$ to at least M of these squares. Therefore, by (7.12),

$$\begin{aligned} &P_p\big(B(m) \leftrightarrow K(m, n) \text{ in } B(n)\big) \\ &\quad \geq \Big\{1 - \Big[1 - pP_p\big(B(m) \text{ is a seed}\big)\Big]^M\Big\} P_p\big(|V(n)| \geq (2m+1)^2 M\big) \\ &\quad \geq (1 - \tfrac{1}{2}\eta) P_p\big(|V(n)| \geq (2m+1)^2 M\big). \end{aligned} \tag{7.13}$$

We now bound the last probability from below. Using the symmetries of $\mathbb{L}^3$ obtained by reflections in hyperplanes, we see that the face $F(n)$ comprises four copies of $T(n)$. Now $\partial B(n)$ has six faces, and therefore 24 copies of $T(n)$. By symmetry and the FKG inequality,

$$P_p\big(|U(n)| < 24(2m+1)^2 M\big) \geq P_p\big(|V(n)| < (2m+1)^2 M\big)^{24} \tag{7.14}$$

where $U(n) = \{x \in \partial B(n) : x \leftrightarrow B(m) \text{ in } B(n)\}$. Now, with $l = 24(2m+1)^2 M$,

$$P_p\big(|U(n)| < l\big) \leq P_p\big(|U(n)| < l,\ B(m) \leftrightarrow \infty\big) + P_p\big(B(m) \nleftrightarrow \infty\big), \tag{7.15}$$

and

(7.16)
$$\begin{aligned} P_p\big(|U(n)| < l,\ B(m) \leftrightarrow \infty\big) &\leq P_p\big(1 \leq |U(n)| < l\big) \\ &\leq (1-p)^{-3l} P_p(U(n+1) = \varnothing,\ U(n) \neq \varnothing) \\ &\to 0 \qquad \text{as } n \to \infty. \end{aligned}$$

(Here we use the fact that $U(n+1) = \varnothing$ if every edge exiting $\partial B(n)$ from $U(n)$ is closed.)

By (7.14)–(7.16) and (7.11),

$$P_p\big(|V(n)| < (2m+1)^2 M\big) \leq P_p\big(|U(n)| < l\big)^{1/24} \leq \big(a_n + (\tfrac{1}{3}\eta)^{24}\big)^{1/24}$$

where $a_n \to 0$ as $n \to \infty$. We pick n such that

$$P_p\big(|V(n)| < (2m+1)^2 M\big) \leq \tfrac{1}{2}\eta,$$

and the claim of the lemma follows by (7.13). □

Having constructed open paths from $B(m)$ to $K(m,n)$, we shall need to repeat the construction, beginning instead at an appropriate seed in $K(m,n)$. This is problematic, since we have discovered a mixture of information, some of it negative, about the immediate environs of such seeds. In order to overcome the effect of such negative information, we shall work at edge-density $p+\delta$ rather than p. In preparation, let $(X(e) : e \in \mathbb{E}^d)$ be independent random variables having the uniform distribution on $[0,1]$, and let $\eta_p(e)$ be the indicator function that $X(e) < p$; recall Section 2.3. We say that e is *p-open* if $X(e) < p$ and *p-closed* otherwise, and we denote by P the appropriate probability measure.

For any subset V of $\mathbb{Z}^3$, we define the *exterior boundary* ΔV and *exterior edge-boundary* $\Delta_{\mathrm{e}} V$ by

$$\begin{aligned} \Delta V &= \{x \in \mathbb{Z}^3 : x \notin V,\ x \sim y \text{ for some } y \in V\}, \\ \Delta_{\mathrm{e}} V &= \{\langle x, y\rangle : x \in V,\ y \in \Delta V,\ x \sim y\}. \end{aligned}$$

We write $\mathbb{E}_V$ for the set of all edges of $\mathbb{L}^3$ joining pairs of vertices in V.

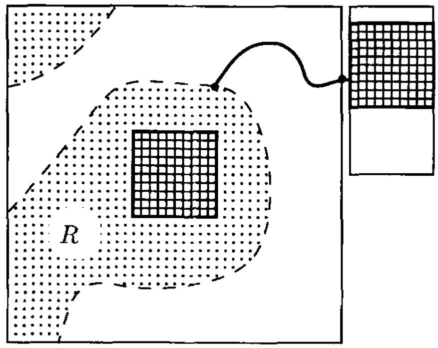


Fig. 7.4. An illustration of Lemma 7.17. The hatched regions are copies of $B(m)$ all of whose edges are p-open. The central box $B(m)$ lies within some (dotted) region R. Some vertex in R is joined by a path to some vertex in $\partial B(n)$, which is in turn connected to a seed lying on the surface of $B(n)$.

We shall make repeated use of the following lemma[5], which is illustrated in Figure 7.4.

Lemma 7.17. *If $\theta(p) > 0$ and $\epsilon, \delta > 0$, there exist integers $m = m(p, \epsilon, \delta)$ and $n = n(p, \epsilon, \delta)$ such that $2m < n$ and with the following property. Let R be such that $B(m) \subseteq R \subseteq B(n)$ and $(R \cup \Delta R) \cap T(n) = \varnothing$, and let $\beta : \Delta_e R \cap \mathbb{E}_{B(n)} \to [0, 1-\delta]$. Define the events*

$$\begin{aligned} G = \{&\text{there exists a path joining } R \text{ to } K(m,n), \text{ this path being } p\text{-open} \\ &\text{outside } \Delta_e R \text{ and } (\beta(e)+\delta)\text{-open at its unique edge } e \text{ lying in } \Delta_e R\}, \\ H = \{&e \text{ is } \beta(e)\text{-closed for all } e \in \Delta_e R \cap \mathbb{E}_{B(n)}\}. \end{aligned}$$

Then $P(G \mid H) > 1 - \epsilon$.

Proof. Assume that $\theta(p) > 0$, and let $\epsilon, \delta > 0$. Pick an integer t so large that

$$(1-\delta)^t < \tfrac{1}{2}\epsilon \tag{7.18}$$

and then choose η (> 0) such that

$$\eta < \tfrac{1}{2}\epsilon(1-p)^t. \tag{7.19}$$

We apply Lemma 7.9 with this value of η, thereby obtaining integers m, n such that $2m < n$ and

$$P_p\big(B(m) \leftrightarrow K(m,n) \text{ in } B(n)\big) > 1 - \eta. \tag{7.20}$$

[5]This lemma is basically Lemma 6 of [165], the difference being that [165] was addressed at *site* percolation. R. Meester and J. Steif have kindly pointed out that, in the case of site percolation, a slightly more general lemma is required than that presented in [165]. The following remarks are directed at the necessary changes to Lemma 6 of [165], and they use the notation of [165]. The proof of the more general lemma is similar to that of the original version. The domain of β is replaced by a general subset S of $B(n) \setminus T(n)$, and G is the event {there exists a path in $B(n)$ from S to $K(m,n)$, this path being p-open in $B(n) \setminus S$ and $(\beta(u)+\delta)$-open at its unique vertex $u \in S$}. In applying the lemma just after (4.10) of [165], we take $S = \Delta C_2 \cap B(n)$ (and similarly later).

Let R and β satisfy the hypotheses of the lemma. Since any path from $B(m)$ to $K(m,n)$ contains a path from ∂R to $K(m,n)$ using no edges of $\mathbb{E}_R$, we have that

$$(7.21) \qquad P_p\big(\partial R \leftrightarrow K(m,n) \text{ in } B(n)\big) > 1 - \eta.$$

Let $K \subseteq T(n)$, and let $U(K)$ be the set of edges $\langle x, y\rangle$ of $B(n)$ such that
(i) $x \in R$, $y \notin R$, and
(ii) there is an open path joining y to K, using no edges of $\mathbb{E}_R \cup \Delta_e R$.
We wish to show that $U(K)$ must be large if $P_p(\partial R \leftrightarrow K \text{ in } B(n))$ is large. The argument centres on the fact that every path from ∂R to K passes through $U(K)$; if $U(K)$ is 'small' then there is substantial uncertainty for the occurrence of the event $\{\partial R \leftrightarrow K \text{ in } B(n)\}$, implying that this event cannot have probability near 1. More rigorously,

$$(7.22) \qquad \begin{aligned} P_p\big(\partial R \not\leftrightarrow K \text{ in } B(n)\big) &= P_p(\text{all edges in } U(K) \text{ are closed}) \\ &\geq (1-p)^t P_p\big(|U(K)| \leq t\big). \end{aligned}$$

We may apply this with $K = K(m,n)$, since $K(m,n)$ is defined on the set of edges exterior to $B(n)$. Therefore, by (7.21), (7.22), and (7.19),

$$(7.23) \qquad \begin{aligned} P_p\Big(\big|U\big(K(m,n)\big)\big| > t\Big) &\geq 1 - (1-p)^{-t} P_p\big(\partial R \not\leftrightarrow K(m,n) \text{ in } B(n)\big) \\ &\geq 1 - (1-p)^{-t}\eta > 1 - \tfrac{1}{2}\epsilon. \end{aligned}$$

We now couple together the percolation processes with different values of p on the same probability space, as described in Section 2.3 and just prior to the statement of Lemma 7.17. We borrow the notation and results derived above by specialising to the p-open edges. Conditional on the set $U = U(K(m,n))$, the values of $X(e)$, for $e \in U$, are independent and uniform on $[0,1]$. Therefore

$$P\Big(\text{every } e \text{ in } U \text{ is } (\beta(e)+\delta)\text{-closed, } |U| > t \,\Big|\, H\Big) \leq (1-\delta)^t,$$

whence, using (7.18) and (7.23),

$$\begin{aligned} P\Big(\text{some } e \text{ in } U \text{ is } (\beta(e)+\delta)\text{-open} \,\Big|\, H\Big) &\geq P(|U| > t \mid H) - (1-\delta)^t \\ &= P_p(|U| > t) - (1-\delta)^t \\ &\geq (1 - \tfrac{1}{2}\epsilon) - \tfrac{1}{2}\epsilon, \end{aligned}$$

and the lemma is proved. □

This completes the two key geometrical lemmas. In moving to the second part of the proof, we shall require a method for comparison of a 'dependent' process and a site percolation process. The argument required at this stage is as follows.

Let F be an infinite connected subset of $\mathbb{L}^d$ for which the associated (*site*) critical probability satisfies $p_c(F, \text{site}) < 1$, and let $\{Z(x) : x \in F\}$ be random variables taking values in $[0,1]$. We construct a connected subset of F in the following recursive manner. Let $e(1), e(2), \dots$ be a fixed ordering of the edges of the graph induced by F. Let $x_1 \in F$, and define the ordered pair $S_1 = (A_1, B_1)$ of subsets of F by

$$S_1 = \begin{cases} (\{x_1\}, \varnothing) & \text{if } Z(x_1) = 1 \\ (\varnothing, \{x_1\}) & \text{if } Z(x_1) = 0. \end{cases}$$

Having defined $S_1, S_2, \dots, S_t = (A_t, B_t)$, for $t \geq 1$, we define S_{t+1} as follows. Let f be the earliest edge in the fixed ordering of the $e(i)$ with the property that one endvertex, x_{t+1} say, lies in A_t and the other endvertex lies outside $A_t \cup B_t$. Then we declare

$$S_{t+1} = \begin{cases} (A_t \cup \{x_{t+1}\}, B_t) & \text{if } Z(x_{t+1}) = 1, \\ (A_t, B_t \cup \{x_{t+1}\}) & \text{if } Z(x_{t+1}) = 0. \end{cases}$$

If no such edge f exists, we declare $S_{t+1} = S_t$. The sets A_t, B_t are non-decreasing, and we set $A_\infty = \lim_{t\to\infty} A_t$, $B_\infty = \lim_{t\to\infty} B_t$. Think about A_∞ as the 'occupied cluster' at x_1, and B_∞ as its external boundary.

Lemma 7.24. *Suppose there exists a constant γ such that $\gamma > p_c(F, \text{site})$ and*

$$P\big(Z(x_{t+1}) = 1 \,\big|\, S_1, S_2, \dots, S_t\big) \geq \gamma \quad \textit{for all } t. \tag{7.25}$$

Then $P(|A_\infty| = \infty) > 0$.

We omit a formal proof of this lemma (but see [165]). Informally, (7.25) implies that, uniformly in the past history, the chance of extending A_t exceeds the critical value of a supercritical site percolation process on F. Therefore A_∞ stochastically dominates the open cluster at x_1 of a supercritical site percolation cluster. The latter cluster is infinite with strictly positive probability, whence $P(|A_\infty| = \infty) > 0$.

Having established the three basic lemmas, we turn to the construction itself. Recall the notation and hypotheses of Theorem 7.8. Let $0 < \eta < p_c$, and choose

$$p = p_c + \tfrac{1}{2}\eta, \quad \delta = \tfrac{1}{12}\eta, \quad \epsilon = \tfrac{1}{24}\big(1 - p_c(F, \text{site})\big). \tag{7.26}$$

Note that $p_c(F, \text{site}) < 1$ since by assumption $p_c(F) = p_c(F, \text{bond}) < 1$ (cf. Theorem 5.13). Since $p > p_c$, we have that $\theta(p) > 0$, and we apply Lemma 7.17 with the above ϵ, δ to find corresponding integers m, n. We define $N = m + n + 1$, and we shall define a process on the blocks of $\mathbb{Z}^3$ having side-length $2N$.

Consider the set $\{4Nx : x \in \mathbb{Z}^d\}$ of vertices, and the associated boxes $B_x(N) = \{4Nx + B(N) : x \in \mathbb{Z}^d\}$; these boxes we call *site-boxes*. A pair $B_x(N), B_y(N)$ of site-boxes is deemed *adjacent* if x and y are adjacent in $\mathbb{L}^d$. Adjacent site-boxes are linked by *bond-boxes*, i.e., boxes $Nz + B(N)$ for $z \in \mathbb{Z}^d$ exactly one component of which is not divisible by 4. If this exceptional component of z is even, the box $Nz + B(N)$ is called a *half-way* box. See Figure 7.5.

We shall examine site-boxes one by one, declaring each to be 'occupied' or 'unoccupied' according to the existence (or not) of certain open paths. Two properties of this construction will emerge.

(a) For each new site-box, the probability that it is occupied exceeds the critical probability of a certain site percolation process. This will imply that, with strictly positive probability, there is an infinite occupied path of site-boxes.

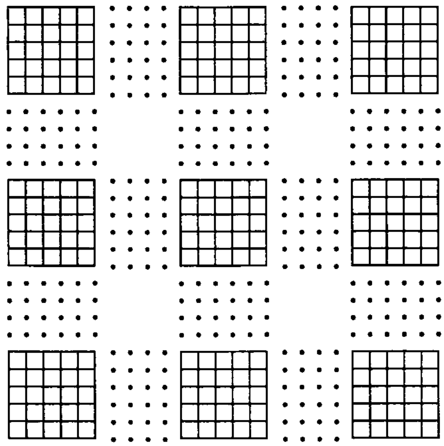


Fig. 7.5. The hatched squares are site-boxes, and the dotted squares are half-way boxes. Each box has side-length $2N$.

(b) The existence of this infinite occupied path necessarily entails an infinite open path of $\mathbb{L}^d$ lying within some restricted region.

The site-boxes will be examined in sequence, the order of this sequence being random, and depending on the past history of the process. Thus, the renormalisation is 'dynamic' rather than 'static'.

As above, let F be an infinite connected subset of $\mathbb{Z}^d$; we shall assume for neatness that F contains the origin 0 (otherwise, translate F accordingly). As above, let $e(1), e(2), \dots$ be a fixed ordering of the edges joining vertices in F. We shall examine the site-boxes $B_x(N)$, for $x \in F$, and determine their states. This we do according to the algorithm sketched before Lemma 7.24, for appropriate random variables $Z(x)$ to be described next.

We begin at the origin, with the site-box $B_0(N) = B(N)$. Once we have explained what is involved in determining the state of $B_0(N)$, most of the work will have been done. (The event $\{B_0(N) \text{ is occupied}\}$ is sketched in Figure 7.7.)

Note that $B(m) \subseteq B(N)$, and say that 'the first step is successful' if every edge in $B(m)$ is p-open, which is to say that $B(m)$ is a 'seed'. (Recall that p and other parameters are given in (7.26).) At this stage we write E_1 for the set of edges of $B(m)$.

In the following sequential algorithm, we shall construct an increasing sequence $E_1, E_2, \dots$ of edge-sets. At each stage k, we shall acquire information about the values of $X(e)$ for certain $e \in \mathbb{E}^3$ (here, the $X(e)$ are independent uniform $[0,1]$-valued random variables, as usual). This information we shall record in the form 'each e is $\beta_k(e)$-closed and $\gamma_k(e)$-open' for suitable functions $\beta_k, \gamma_k : \mathbb{E}^3 \to [0,1]$ satisfying

$$(7.27) \qquad \beta_k(e) \le \beta_{k+1}(e), \quad \gamma_k(e) \ge \gamma_{k+1}(e), \quad \text{for all } e \in \mathbb{E}^3.$$

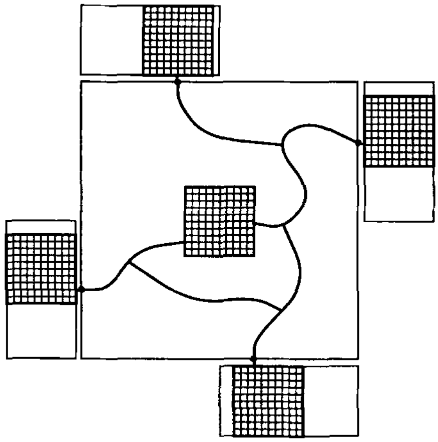


Fig. 7.6. An illustration of the first two steps in the construction of the event $\{0 \text{ is occupied}\}$, when these steps are successful. Each hatched square is a seed.

Having constructed E_1, above, we set

$$\beta_1(e) = 0 \quad \text{for all } e \in \mathbb{E}^3, \tag{7.28}$$

$$\gamma_1(e) = \begin{cases} p & \text{if } e \in E_1, \\ 1 & \text{otherwise.} \end{cases} \tag{7.29}$$

Since we are working with edge-sets E_j rather than with vertex-sets, it will be useful to have some corresponding notation. Two edges e, f are called *adjacent*, written $e \approx f$, if they have exactly one common endvertex. This adjacency relation defines a graph. Paths in this graph are said to be α-open if $X(e) < \alpha$ for all e lying in the path. The exterior edge-boundary $\Delta_e E$ of an edge-set E is the set of all edges $f \in \mathbb{E}^3 \setminus E$ such that $f \approx e$ for some $e \in E$.

For $j = 1, 2, 3$ and $\sigma = \pm$, let L_j^σ be an automorphism of $\mathbb{L}^3$ which preserves the origin and maps $e_1 = (1, 0, 0)$ onto σe_j; we insist that L_1^+ is the identity. We now define E_2 as follows. Consider the set of all paths π lying within the region

$$B_1' = B(n) \cup \left\{ \bigcup_{\substack{1 \le j \le 3 \\ \sigma = \pm}} L_j^\sigma\big(T(m, n)\big) \right\}$$

such that

(a) the first edge f of π lies in $\Delta_e E_1$ and is $(\beta_1(f) + \delta)$-open, and

(b) all other edges lie outside $E_1 \cup \Delta_e E_1$ and are p-open.

We define $E_2 = E_1 \cup F_1$ where F_1 is the set of all edges in the union of such paths π. We say that 'the second step is successful' if, for each $j = 1, 2, 3$ and $\sigma = \pm$, there is an edge in E_2 having an endvertex in $K_j^\sigma(m, n)$, where

$$K_j^\sigma(m, n) = \Big\{ z \in L_j^\sigma\big(T(n)\big) : \langle z, z + \sigma e_j \rangle \text{ is } p\text{-open, and } z + \sigma e_j \text{ lies in some seed lying within } L_j^\sigma\big(T(m, n)\big) \Big\}.$$

The corresponding event is illustrated in Figure 7.6.

Next we estimate the probability that the second step is successful, conditional on the first step being successful. Let G be the event that there exists a path in $B(n)\backslash B(m)$ from $\partial B(m)$ to $K(m,n)$, every edge e of which is p-open off $\Delta_e E_1$ and whose unique edge f in $\Delta_e E_1$ is $(\beta_1(f)+\delta)$-open. We write G_j^σ for the corresponding event with $K(m,n)$ replaced by $L_j^\sigma(K(m,n))$. We now apply Lemma 7.17 with $R = B(m)$ and $\beta = \beta_1$ to find that

$$P\big(G_j^\sigma \,\big|\, B(m) \text{ is a seed}\big) > 1-\epsilon \quad \text{for } j=1,2,3,\ \sigma=\pm.$$

Therefore

$$P\big(G_j^\sigma \text{ occurs for all } j,\sigma \,\big|\, B(m) \text{ is a seed}\big) > 1-6\epsilon, \tag{7.30}$$

so that the second step is successful with conditional probability at least $1-6\epsilon$.

If the second step is successful, then we update the β,γ functions accordingly, setting

$$\beta_2(e) = \begin{cases} \beta_1(e) & \text{if } e \notin \mathbb{E}_{B_1'}, \\ \beta_1(e)+\delta & \text{if } e \in \Delta_e E_1 \backslash E_2, \\ p & \text{if } e \in \Delta_e E_2 \backslash \Delta_e E_1, \\ 0 & \text{otherwise,} \end{cases} \tag{7.31}$$

$$\gamma_2(e) = \begin{cases} \gamma_1(e) & \text{if } e \in E_1, \\ \beta_1(e)+\delta & \text{if } e \in \Delta_e E_1 \cap E_2, \\ p & \text{if } e \in E_2 \backslash (E_1 \cup \Delta_e E_1), \\ 1 & \text{otherwise.} \end{cases} \tag{7.32}$$

Suppose that the first two steps have been successful. We next aim to link the appropriate seeds in each $L_j^\sigma(T(m,n))$ to a new seed lying in the bond-box $2\sigma N e_j + B(N)$, i.e., the half-way box reached by exiting the origin in the direction σe_j. If we succeed with each of the six such extensions, then we terminate this stage of the process, and declare the vertex 0 of the renormalised lattice to be *occupied*; such success constitutes the definition of the term 'occupied'. See Figure 7.7.

We do not present all the details of this part of the construction, since they are very similar to those already described. Instead we concentrate on describing the basic strategy, and discussing any novel aspects of the construction. First, let $B_2 = b_2 + B(m)$ be the earliest seed (in some ordering of all copies of $B(m)$) all of whose edges lie in $E_2 \cap \mathbb{E}_{T(m,n)}$. We now try to extend E_2 to include a seed lying within the bond-box $2Ne_1 + B(N)$. Clearly $B_2 \subseteq Ne_1 + B(N)$. In performing this extension, we encounter a 'steering' problem. It happens (by construction) that all coordinates of b_2 are positive, implying that $b_2 + T(m,n)$ is not a subset of $2Ne_1 + B(N)$. We therefore replace $b_2 + T(m,n)$ by $b_2 + T^*(m,n)$ where $T^*(m,n)$ is given as follows. Instead of working with the 'quadrant' $T(n)$ of the face $F(n)$, we use the set

$$T^*(n) = \big\{x \in \partial B(n) : x_1 = n,\ x_j \le 0 \text{ for } j=2,3\big\}.$$

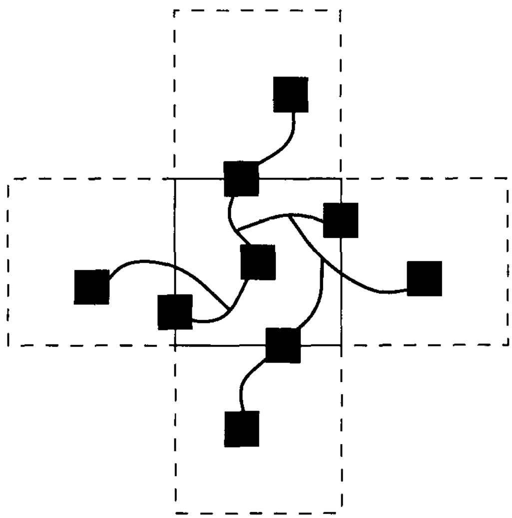


Fig. 7.7. An illustration of the event $\{0 \text{ is occupied}\}$. Each black square is a seed.

We then define

$$T^*(m,n) = \bigcup_{j=1}^{2m+1} \{je_1 + T^*(n)\},$$

and obtain that $b_2 + T^*(m,n) \subseteq 2Ne_1 + B(N)$. We now consider the set of all paths π lying within the region

$$B_2' = b_2 + \{B(n) \cup T^*(m,n)\}$$

such that:

(a) the first edge f of π lies in $\Delta_e E_2$ and is $(\beta_2(f)+\delta)$-open, and

(b) all other edges lie outside $E_2 \cup \Delta_e E_2$ and are p-open.

We set $E_3 = E_2 \cup F_2$ where F_2 is the set of all edges lying in the union of such paths. We call this step successful if E_3 contains an edge having an endvertex in the set

$$b_2 + K^*(m,n) = \Big\{ z \in b_2 + T^*(m,n) : \langle z, z+e_1 \rangle \text{ is } p\text{-open, } z + e_1 \\ \text{lies in some seed lying in } b_2 + T^*(m,n) \Big\}.$$

Using Lemma 7.17, the (conditional) probability that this step is successful exceeds $1-\epsilon$.

We perform similar extensions in each of the other five directions exiting $B_0(N)$. If all are successful, we declare 0 to be *occupied*. Combining the above estimates of success, we find that

(7.33)

$$\begin{aligned} P(0 \text{ is occupied} \mid B(m) \text{ is a seed}) &> (1-6\epsilon)(1-\epsilon)^6 > 1 - 12\epsilon \\ &= \tfrac{1}{2}\big(1 + p_c(F, \text{site})\big) \end{aligned}$$

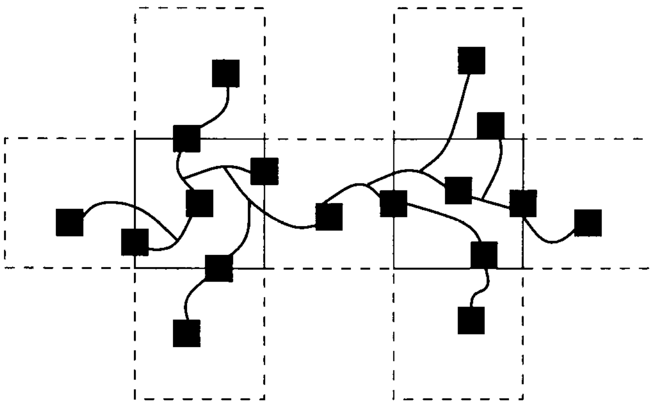


Fig. 7.8. Two adjacent site-boxes both of which are occupied. The construction began with the left site-box $B_0(N)$ and has been extended to the right site-box $B_{e_1}(N)$. The black squares are seeds, as before.

by (7.26).

If 0 is not occupied, we end the construction. If 0 is occupied, then this has been achieved after the definition of a set E_8 of edges. The corresponding functions β_8, γ_8 are such that

$$\beta_8(e) \le \gamma_8(e) \le p + 6\delta \quad \text{for } e \in E_8; \tag{7.34}$$

this follows since no edge lies in more than 7 of the copies of $B(n)$ used in the repeated application of Lemma 7.17. Therefore every edge of E_8 is $(p+\eta)$-open, since $\delta = \frac{1}{12}\eta$ (see (7.26)).

The basic idea has been described, and we now proceed similarly. Assume 0 is occupied, and find the earliest edge $e(r)$ induced by F and incident with the origin; we may assume for the sake of simplicity that $e(r) = \langle 0, e_1 \rangle$. We now attempt to link the seed $b_3 + B(m)$, found as above inside the half-way box $2Ne_1 + B(N)$, to a seed inside the site-box $4Ne_1 + B(N)$. This is done in two steps of the earlier kind. Having found a suitable seed inside the new site-box $4Ne_1 + B(N)$, we attempt to branch-out in the other 5 directions from this site-box. If we succeed in finding seeds in each of the corresponding half-way boxes, then we declare the vertex e_1 of the renormalised lattice to be occupied. As before, the (conditional) probability that e_1 is occupied is at least $\frac{1}{2}(1 + p_c(F, \text{site}))$, and every edge in the ensuing construction is $(p+\eta)$-open. See Figure 7.8.

Two details arise at this and subsequent stages, each associated with 'steering'. First, if $b_3 = (\alpha_1, \alpha_2, \alpha_3)$ we concentrate on the quadrant $T_\alpha(n)$ of $\partial B(n)$ defined as the set of $x \in \partial B(n)$ for which $x_j \alpha_j \le 0$ for $j = 2, 3$ (so that x_j has the opposite sign to α_j). Having found such a $T_\alpha(n)$, we define $T^*_\alpha(m, n)$ accordingly, and look for paths from $b_3 + B(m)$ to $b_3 + T^*_\alpha(m, n)$. This mechanism guarantees that any variation in b_3 from the first coordinate axis is (at least partly) compensated for at the next step.

A further detail arises when branching out from the seed $b^* + B(m)$ reached inside $4Ne_1 + B(N)$. In finding seeds lying in the new half-way boxes abutting

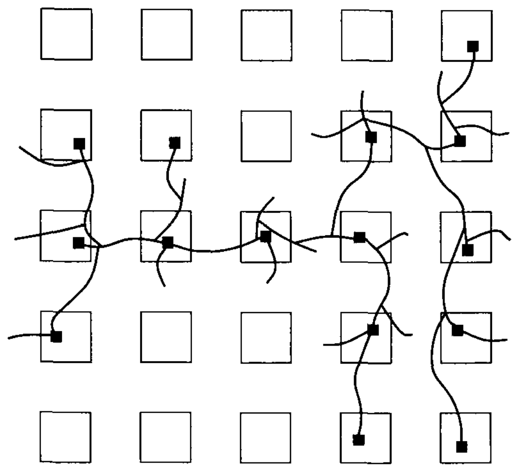


Fig. 7.9. The central seed is $B(m)$, and the connections represent $(p+\eta)$-open paths joining seeds within the site-boxes.

$4Ne_1 + B(N)$, we 'steer away from the inlet branch', by examining seeds lying on the surface of $b^* + B(n)$ with the property that the first coordinates of their vertices are not less than that of b^*. This process guarantees that these seeds have not been examined previously.

We now continue to apply the algorithm presented before Lemma 7.24. At each stage, the chance of success exceeds $\gamma = \frac{1}{2}(1 + p_c(F, \text{site}))$. Since $\gamma > p_c(F, \text{site})$, we have from Lemma 7.24 that there is a strictly positive probability that the ultimate set of occupied vertices of F (i.e., renormalised blocks of $\mathbb{L}^3$) is infinite. Now, on this event, there must exist an infinite $(p + \eta)$-open path of $\mathbb{L}^3$ corresponding to the enlargement of F. This infinite open path must lie within the enlarged set $4NF + B(2N)$, implying that $p_c + \eta \geq p_c(4NF + B(2N))$, as required for Theorem 7.9. See Figure 7.9. The proof is complete. $\square$

7.4 Percolation in Half-Spaces

In the last section, we almost succeeded in proving that $\theta(p_c) = 0$ when $d \geq 3$. The reason for this statement is as follows. Suppose $\theta(p) > 0$ and $\eta > 0$. There is effectively defined in Section 7.3 an event A living in a finite box B such that

(a) $P_p(A) > 1 - \epsilon$, for some prescribed $\epsilon > 0$,

(b) the fact (a) implies that $\theta(p + \eta) > 0$.

Suppose that we could prove this with $\eta = 0$, and that $\theta(p_c) > 0$. Then $P_{p_c}(A) > 1 - \epsilon$, which implies by continuity that $P_{p'}(A) > 1 - \epsilon$ for some $p' < p_c$, and therefore $\theta(p') > 0$ by (b). This would contradict the definition of p_c, whence we deduce by contradiction that $\theta(p_c) = 0$.

The fact that η is strictly positive is vital for the construction, since we need to 'spend some extra money' in order to compensate for negative information acquired earlier in the construction. In a slightly different setting, no extra money is required.

Let $\mathbb{H} = \{0, 1, \ldots\} \times \mathbb{Z}^{d-1}$ be a half-space when $d \geq 3$, and write $p_c(\mathbb{H})$ for its critical probability. It follows from Theorem 7.8 that $p_c(\mathbb{H}) = p_c$, since $\mathbb{H}$ contains slabs of all thicknesses. Let

$$\theta_{\mathbb{H}}(p) = P_p(0 \leftrightarrow \infty \text{ in } \mathbb{H}).$$

Theorem 7.35. *We have that* $\theta_{\mathbb{H}}(p_c) = 0$.

The proof is not presented here, but may be found in [49, 50]. It is closely related to that presented in Section 7.3, but with some crucial differences. The construction of blocks is slightly more complicated, owing to the lack of symmetry of $\mathbb{H}$, but there are compensating advantages of working in a half-space. For amusement, we present in Figure 7.10 two diagrams (relevant to the argument of [50]) depicting the necessary constructions.

As observed in Section 3.3, such a conclusion for half-spaces has a striking implication for the conjecture that $\theta(p_c) = 0$. If $\theta(p_c) > 0$, then there exists a.s. a unique infinite open cluster in $\mathbb{Z}^d$, which is a.s. partitioned into (only) finite clusters by *any* division of $\mathbb{Z}^d$ into two half-spaces.

7.5 Percolation Probability

Although the methods of Chapter 6 were derived primarily in order to study *subcritical* percolation, they involve a general inequality of wider use, namely

$$g'_\pi(n) \geq g_\pi(n)\left(\frac{n}{\sum_{i=0}^{n} g_\pi(i)} - 1\right)$$

where $g_\pi(n) = P_\pi(0 \leftrightarrow \partial S_n)$; see equations (3.10) and (3.18) in Section 6. We argue loosely as follows. Clearly $g_\pi(n) \to \theta(\pi)$ as $n \to \infty$, whence (cross your fingers here)

$$\theta'(\pi) \geq \theta(\pi)\left(\frac{1}{\theta(\pi)} - 1\right)$$

or

$$\theta'(\pi) + \theta(\pi) \geq 1.$$

Integrate this over the interval (p_c, p) to obtain

$$\theta(p)e^p - \theta(p_c)e^{p_c} \geq e^p - e^{p_c}, \qquad p_c \leq p,$$

whence it is an easy exercise to show that

$$\theta(p) - \theta(p_c) \geq a(p - p_c), \qquad p_c \leq p, \tag{7.36}$$

for some positive constant a. The above argument may be made rigorous.

Differential inequalities of the type above are used widely in percolation and disordered systems.

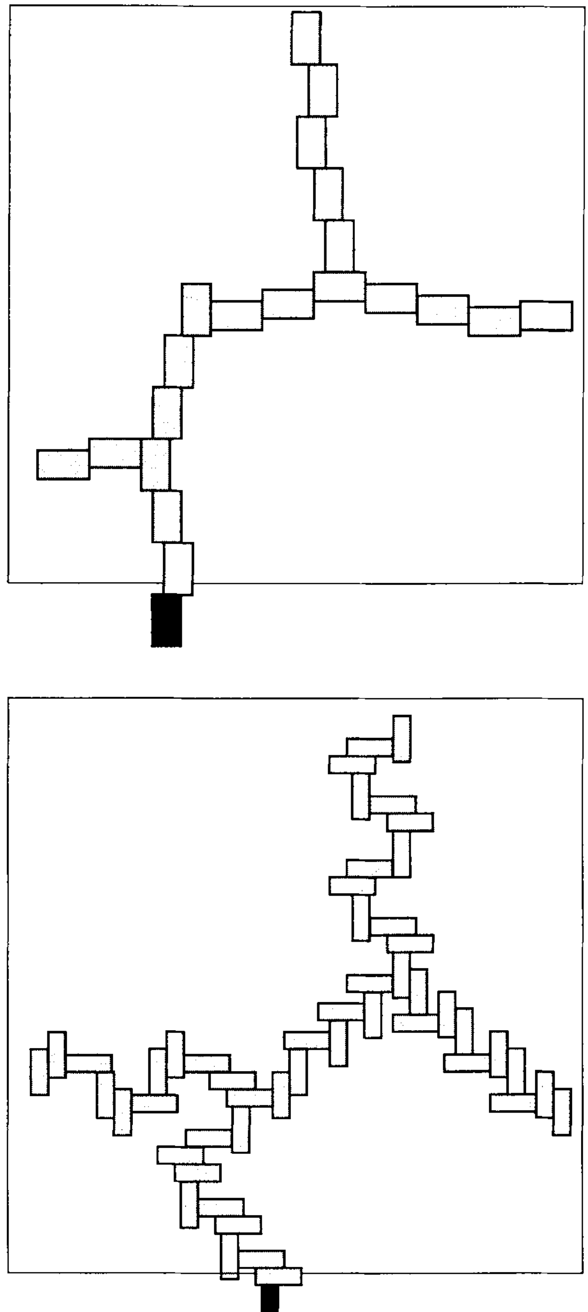


Fig. 7.10. Illustrations of the block construction for the proof of Theorem 7.35 presented in [50]. The grey regions contain open paths joining the black 'inlet' to the three 'outlets'. The fundamental building block is a rectangle rather than a square. We have no control of the aspect ratio of this rectangle, and consequently two cases with somewhat different geometries need to be considered. Compare with Figure 7.7.

7.6 Cluster-Size Distribution

When $p < p_c$, the tail of the cluster-size $|C|$ decays exponentially. Exponential decay is not correct when $p > p_c$, but rather 'stretched exponential decay'.

Theorem 7.37. *Suppose $p_c < p < 1$. There exist positive constants $\alpha(p), \beta(p)$ such that, for all n,*

$$\exp\left(-\alpha n^{(d-1)/d}\right) \le P_p(|C| = n) \le \exp\left(-\beta n^{(d-1)/d}\right). \tag{7.38}$$

See [G] for a proof of this theorem. The reason for the power $n^{(d-1)/d}$ is roughly as follows. It is thought that a large finite cluster is most likely created as a cluster of compact shape, all of whose boundary edges are closed. Now, if a ball has volume n, then its surface area has order $n^{(d-1)/d}$. The price paid for having a surface all of whose edges are closed is $(1-p)^m$ where m is the number of such edges. By the above remark, m should have order $n^{(d-1)/d}$, as required for (7.38).

It is believed that the limit

$$\gamma(p) = \lim_{n\to\infty} \left\{ -\frac{1}{n^{(d-1)/d}} \log P_p(|C| = n) \right\} \tag{7.39}$$

exists, but no proof is known.

Much more is known in two dimensions than for general d. The size and geometry of large finite clusters have been studied in detail in [38], where it was shown that such clusters may be approximated by the so called 'Wulff shape'. This work includes a proof of the existence of the limit in (7.39) when $d = 2$.

8. CRITICAL PERCOLATION

8.1 Percolation Probability

The next main open question is to verify the following.

Conjecture 8.1. *We have that* $\theta(p_c) = 0$.

This is known to hold when $d = 2$ (using results of Harris [181], see Theorem 9.1) and for sufficiently large values of d (by work of Hara and Slade [178, 179]), currently for $d \geq 19$. The methods of Hara and Slade might prove feasible for values of d as small as 6 or 7, but not for smaller d. Some new idea is needed for the general conclusion. As remarked in Section 7.4, we need to rule out the remaining theoretical possibility that there is an infinite cluster in $\mathbb{Z}^d$ when $p = p_c$, but no infinite cluster in any half-space.

8.2 Critical Exponents

Macroscopic functions, such as the percolation probability, have a singularity at $p = p_c$, and it is believed that there is 'power law behaviour' at and near this singularity. The nature of the singularity is supposed to be canonical, in that it is expected to have certain general features in common with phase transitions in other physical systems. These features are sometimes referred to as 'scaling theory' and they relate to 'critical exponents'.

There are two sets of critical exponents, arising firstly in the limit as $p \to p_c$, and secondly in the limit over increasing distances when $p = p_c$. We summarise the notation in Table 7.1.

The asymptotic relation $\approx$ should be interpreted loosely (perhaps via logarithmic asymptotics). The radius of C is defined by $\operatorname{rad}(C) = \max\{n : 0 \leftrightarrow \partial B(n)\}$. The limit as $p \to p_c$ should be interpreted in a manner appropriate for the function in question (for example, as $p \downarrow p_c$ for $\theta(p)$, but as $p \to p_c$ for $\kappa(p)$).

There are eight critical exponents listed in Table 7.1, denoted $\alpha, \beta, \gamma, \delta, \nu, \eta, \rho, \Delta$, but there is no general proof of the existence of any of these exponents.

8.3 Scaling Theory

In general, the eight critical exponents may be defined for phase transitions in a quite large family of physical systems. However, it is not believed that they are independent variables, but rather that they satisfy the following:

(8.2) *Scaling relations*

$$\begin{aligned} 2 - \alpha &= \gamma + 2\beta = \beta(\delta + 1), \\ \Delta &= \delta\beta, \\ \gamma &= \nu(2 - \eta), \end{aligned}$$

Function		Behaviour	Exponent
percolation probability	$\theta(p) = P_p(\|C\| = \infty)$	$\theta(p) \approx (p - p_c)^\beta$	β
truncated mean cluster size	$\chi^f(p) = E_p(\|C\|; \|C\| < \infty)$	$\chi^f(p) \approx \|p - p_c\|^{-\gamma}$	γ
number of clusters per vertex	$\kappa(p) = E_p(\|C\|^{-1})$	$\kappa'''(p) \approx \|p - p_c\|^{-1-\alpha}$	α
cluster moments	$\chi_k^f(p) = E_p(\|C\|^k; \|C\| < \infty)$	$\dfrac{\chi_{k+1}^f(p)}{\chi_k^f(p)} \approx \|p - p_c\|^{-\Delta},\ k \geq 1$	Δ
correlation length	$\xi(p)$	$\xi(p) \approx \|p - p_c\|^{-\nu}$	ν
cluster volume		$P_{p_c}(\|C\| = n) \approx n^{-1-1/\delta}$	δ
cluster radius		$P_{p_c}(\text{rad}(C) = n) \approx n^{-1-1/\rho}$	ρ
connectivity function		$P_{p_c}(0 \leftrightarrow x) \approx \|\|x\|\|^{2-d-\eta}$	η

Table 7.1. Eight functions and their critical exponents.

and, when d is not too large, the

(8.3) *Hyperscaling relations*

$$d\rho = \delta + 1,$$
$$2 - \alpha = d\nu.$$

The *upper critical dimension* is the largest value d_c such that the hyperscaling relations hold for $d \leq d_c$. It is believed that $d_c = 6$ for percolation.

There is no general proof of the validity of the scaling and hyperscaling relations, although certain things are known when $d = 2$ and for large d.

In the context of percolation, there is an analytical rationale behind the scaling relations, namely the 'scaling hypotheses' that

$$P_p(|C| = n) \sim n^{-\sigma} f\big(n/\xi(p)^\tau\big)$$
$$P_p(0 \leftrightarrow x,\ |C| < \infty) \sim \|x\|^{2-d-\eta} g\big(\|x\|/\xi(p)\big)$$

in the double limit as $p \to p_c$, $n \to \infty$, and for some constants σ, τ, η and functions f, g. Playing loose with rigorous mathematics, the scaling relations may be derived from these hypotheses. Similarly, the hyperscaling relations may be shown to be not too unreasonable, at least when d is not too large. For further discussion, see [G].

We make some further points.
Universality. It is believed that the numerical values of critical exponents depend only on the value of d, and are independent of the particular percolation model.
Two dimensions. When $d = 2$, perhaps

$$\alpha = -\tfrac{2}{3}, \quad \beta = \tfrac{5}{36}, \quad \gamma = \tfrac{43}{18}, \quad \delta = \tfrac{91}{5}, \ldots$$

Large dimension. When d is sufficiently large (actually, $d \geq d_c$) it is believed that the critical exponents are the same as those for percolation on a tree (the 'mean-field model'), namely $\delta = 2$, $\gamma = 1$, $\nu = \frac{1}{2}$, $\rho = \frac{1}{2}$, and so on (the other exponents are found to satisfy the scaling relations). Using the first hyperscaling relation, this supports the contention that $d_c = 6$. Such statements are known to hold for $d \geq 19$; see [178, 179] and Section 8.5.

8.4 Rigorous Results

Open challenges include to prove:

- the existence of critical exponents,
- universality,
- the scaling relations,
- the conjectured values when $d = 2$,
- the conjectured values when $d \geq 6$.

Progress towards these goals has been slender, but positive. Most is known in the case of large d, see the next section. For sufficiently large d, exact values are known for many exponents, namely the values from percolation on a regular tree. When $d = 2$, Kesten [204, 205] has proved that, *if two critical exponents exist*, then certain others do also, and certain scaling relations are valid. However, the provocative case when $d = 3$ is fairly open terrain.

Certain partial results are known in generality, yielding inequalities in situations where one expects (asymptotic) equalities. For example, it is known that $\beta \leq 1$, if β exists (cf. (7.36)). In similar vein, we have that $\gamma \geq 1$ and $\delta \geq 2$ for all d.

8.5 Mean-Field Theory

The expression 'mean-field' permits several interpretations depending on context. A narrow interpretation of the term 'mean-field theory' for percolation involves trees rather than lattices. For percolation on a regular tree, it is quite easy to perform exact calculations of many quantities, including the numerical values of critical exponents. That is, $\delta = 2$, $\gamma = 1$, $\nu = \frac{1}{2}$, $\rho = \frac{1}{2}$, and other exponents are given according to the scaling relations (8.2); see [G], Section 8.1.

Turning to percolation on $\mathbb{L}^d$, it is known that the critical exponents agree with those of a regular tree when d is sufficiently large. In fact, this is believed to hold if and only if $d \geq 6$, but progress so far assumes that $d \geq 19$. In the following theorem, taken from [179], we write $f(x) \simeq g(x)$ if there exist positive constants c_1, c_2 such that $c_1 f(x) \leq g(x) \leq c_2 f(x)$ for all x close to a limiting value.

Theorem 8.4. *If $d \geq 19$ then*

$$\theta(p) \simeq (p - p_c)^1 \quad as\ p \downarrow p_c, \tag{8.5}$$

$$\chi(p) \simeq (p_c - p)^{-1} \quad as\ p \uparrow p_c, \tag{8.6}$$

$$\xi(p) \simeq (p_c - p)^{-\frac{1}{2}} \quad as\ p \uparrow p_c, \tag{8.7}$$

$$\frac{\chi^{\mathrm{f}}_{k+1}(p)}{\chi^{\mathrm{f}}_{k}(p)} \simeq (p_c - p)^{-2} \quad as\ p \uparrow p_c,\ for\ k \geq 1. \tag{8.8}$$

Note the strong form of the asymptotic relation $\simeq$, and the identification of the critical exponents $\beta, \gamma, \Delta, \nu$. The proof of Theorem 8.4 centres on a property known as the 'triangle condition'. Define

$$T(p) = \sum_{x,y \in \mathbb{Z}^d} P_p(0 \leftrightarrow x) P_p(x \leftrightarrow y) P_p(y \leftrightarrow 0), \tag{8.9}$$

and introduce the following condition,

(8.10) *Triangle condition:* $\qquad T(p_c) < \infty.$

The triangle condition was introduced by Aizenman and Newman [27], who showed that it implied that $\chi(p) \simeq (p_c - p)^{-1}$ as $p \uparrow p_c$. Subsequently other authors showed that the triangle condition implied similar asymptotics for other quantities. It was Hara and Slade [178] who verified the triangle condition for large d, exploiting a technique known as the 'lace expansion'.

We present no full proof of Theorem 8.4 here, pleading two reasons. First, such a proof would be long and complicated. Secondly, we are unable to do better than is already contained in the existing literature (see [178, 179]). Instead, we (nearly) prove the above Aizenman–Newman result (equation (8.6) above), namely that the triangle condition implies that $\chi(p) \simeq (p_c - p)^{-1}$ as $p \uparrow p_c$; then we present a very brief discussion of the Hara–Slade verification of the triangle condition for large d. We begin with a lemma.

Lemma 8.11. *Let $\tau_p(u, v) = P_p(u \leftrightarrow v)$, and*

$$Q(a, b) = \sum_{v,w \in \mathbb{Z}^d} \tau_p(a, v) \tau_p(v, w) \tau_p(w, b) \quad for\ a, b \in \mathbb{Z}^d.$$

Then Q is a positive-definite form, in that

$$\sum_{a,b} f(a) Q(a, b) \overline{f(b)} \geq 0$$

for all suitable functions $f : \mathbb{Z}^d \to \mathbb{C}$.

Proof. We have that

$$\begin{aligned} \sum_{a,b} f(a) Q(a, b) \overline{f(b)} &= \sum_{v,w} g(v) \tau_p(v, w) \overline{g(w)} \\ &= E_p\Big(\sum_{v,w} g(v) 1_{\{v \leftrightarrow w\}} \overline{g(w)} \Big) \\ &= E_p\Big(\sum_{C} \Big| \sum_{x \in C} g(x) \Big|^2 \Big) \end{aligned}$$

where $g(v) = \sum_a f(a)\tau_p(a,v)$, and the penultimate summation is over all open clusters C. □

We note the consequence of Lemma 8.11, that

$$Q(a,b)^2 \le Q(a,a)Q(b,b) = T(p)^2 \tag{8.12}$$

by Schwarz's inequality.

Theorem 8.13. *If $d \ge 2$ and $T(p_c) < \infty$ then*

$$\chi(p) \simeq (p_c - p)^{-1} \quad \textit{as } p \uparrow p_c.$$

Proof. This is taken from [27]; see also [G] and [179]. The following sketch is incomplete in one important regard, namely that, in the use of Russo's formula, one should first restrict oneself to a finite region Λ, and later pass to the limit as $\Lambda \uparrow \mathbb{Z}^d$; we omit the details of this.

Write $\tau_p(u,v) = P_p(u \leftrightarrow v)$ as before, so that

$$\chi(p) = \sum_{x \in \mathbb{Z}^d} \tau_p(0,x).$$

By (ab)use of Russo's formula,

$$\frac{d\chi}{dp} = \frac{d}{dp} \sum_{x \in \mathbb{Z}^d} \tau_p(0,x) = \sum_{x \in \mathbb{Z}^d} \sum_{e \in \mathbb{E}^d} P_p(e \text{ is pivotal for } \{0 \leftrightarrow x\}). \tag{8.14}$$

If $e = \langle a,b \rangle$ is pivotal for $\{0 \leftrightarrow x\}$, then one of the events $\{0 \leftrightarrow a\} \circ \{b \leftrightarrow x\}$ and $\{0 \leftrightarrow b\} \circ \{a \leftrightarrow x\}$ occurs. Therefore, by the BK inequality,

$$\begin{aligned} \frac{d\chi}{dp} &\le \sum_x \sum_{e=\langle a,b \rangle} \big\{\tau_p(0,a)\tau_p(b,x) + \tau_p(0,b)\tau_p(a,x)\big\} \\ &= \sum_{e=\langle a,b \rangle} \big\{\tau_p(0,a) + \tau_p(0,b)\big\}\chi(p) = 2d\chi(p)^2. \end{aligned} \tag{8.15}$$

This inequality may be integrated, to obtain that

$$\frac{1}{\chi(p_1)} - \frac{1}{\chi(p_2)} \le 2d(p_2 - p_1) \quad \text{for } p_1 \le p_2.$$

Take $p_1 = p < p_c$ and $p_2 > p_c$, and allow the limit $p_2 \downarrow p_c$, thereby obtaining that

$$\chi(p) \ge \frac{1}{2d(p_c - p)} \quad \text{for } p < p_c. \tag{8.16}$$

In order to obtain a corresponding lower bound for $\chi(p)$, we need to obtain a lower bound for (8.14). Let $e = \langle a,b \rangle$ in (8.14), and change variables $(x \mapsto x - a)$ in the summation to obtain that

$$\frac{d\chi}{dp} = \sum_{x,y} \sum_{|u|=1} P_p\big(0 \leftrightarrow x,\ u \leftrightarrow y \text{ off } C_u(x)\big) \tag{8.17}$$

where the second summation is over all unit vectors u of $\mathbb{Z}^d$. The (random set) $C_u(x)$ is defined as the set of all points joined to x by open paths not using $\langle 0,u \rangle$.

In the next lemma, we have a strictly positive integer R, and we let $B = B(R)$. The set $C_B(x)$ is the set of all points reachable from x along open paths using no vertex of B.

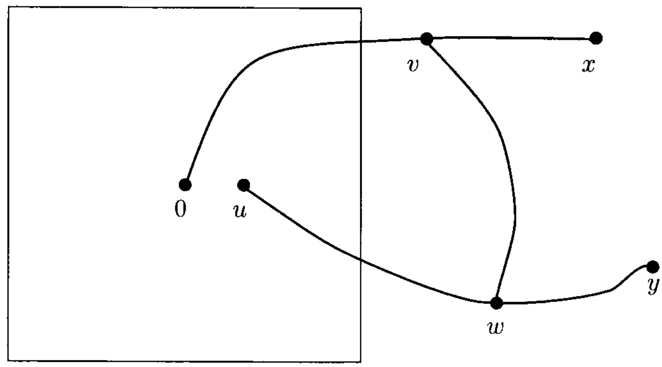


Fig. 8.1. If $0 \leftrightarrow x$, $u \leftrightarrow y$, and $x \leftrightarrow y$ off B, then there exist $v, w \notin B$ such that there are disjoint open paths from x to v, from v to 0, from v to w, from w to u, and from w to y.

Lemma 8.18. *Let u be a unit vector. We have that*

$$P_p\big(0 \leftrightarrow x,\ u \leftrightarrow y \text{ off } C_u(x)\big) \geq \alpha(p) P_p\big(0 \leftrightarrow x,\ u \leftrightarrow y \text{ off } C_B(x)\big),$$

where $\alpha(p) = \{\min(p, 1-p)\}^{2d(2R+1)^d}$.

Proof of Lemma 8.18. Define the following events,

$$\begin{aligned} E &= \big\{0 \leftrightarrow x,\ u \leftrightarrow y \text{ off } C_u(x)\big\}, \\ F &= \big\{0 \leftrightarrow x,\ u \leftrightarrow y \text{ off } C_B(x)\big\}, \qquad (8.19) \\ G &= \big\{B \cap C(x) \neq \varnothing,\ B \cap C(y) \neq \varnothing,\ C_B(x) \cap C_B(y) = \varnothing\big\}, \end{aligned}$$

noting that $E \subseteq F \subseteq G$. Now

$$P_p(E) = P_p(E \mid G) P_p(G) \geq P_p(E \mid G) P_p(F).$$

The event G is independent of all edges lying in the edge-set $\mathbb{E}_B$ of B. Also, for any $\omega \in G$, there exists a configuration $\omega_B = \omega_B(\omega)$ for the edges in $\mathbb{E}_B$ such that the composite configuration (ω off $\mathbb{E}_B$, and ω_B on $\mathbb{E}_B$) lies in E. Since $\mathbb{E}_B$ is finite, and $P_p(\omega_B) \geq \alpha(p)$ whatever the choice of ω_B, we have that $P_p(E \mid G) \geq \alpha(p)$, and the conclusion of the lemma follows. □

The event F in (8.19) satisfies

$$F = \{0 \leftrightarrow x,\ u \leftrightarrow y\} \backslash \{0 \leftrightarrow x,\ u \leftrightarrow y,\ x \leftrightarrow y \text{ off } B\},$$

whence, by the FKG inequality,

$$\begin{aligned} P_p(F) &= P_p(0 \leftrightarrow x,\ u \leftrightarrow y) - P_p(0 \leftrightarrow x,\ u \leftrightarrow y,\ x \leftrightarrow y \text{ off } B) \\ &\geq \tau_p(0, x)\tau_p(u, y) - P_p(0 \leftrightarrow x,\ u \leftrightarrow y,\ x \leftrightarrow y \text{ off } B). \end{aligned}$$

To bound the last term, we use the BK inequality. Glancing at Figure 8.1, we see that, if $0 \leftrightarrow x$, $u \leftrightarrow y$, and $x \leftrightarrow y$ off B, then there exist $v, w \notin B$ such that there are disjoint open paths from x to v, from v to 0, from v to w, from w to u, and from w to y. Applying the BK inequality,

$$P_p(F) \geq \tau_p(0,x)\tau_p(u,y) - \sum_{v,w \notin B} \tau_p(x,v)\tau_p(v,0)\tau_p(v,w)\tau_p(w,u)\tau_p(w,y).$$

Now sum over x and y to obtain via (8.17) and Lemma 8.18 that

$$\frac{d\chi}{dp} \geq 2d\alpha(p)\chi^2\left(1 - \sup_{|u|=1} \sum_{v,w\notin B} \tau_p(0,v)\tau_p(v,w)\tau_p(w,u)\right). \tag{8.20}$$

Now, by (8.12),

$$\sum_{v,w} \tau_p(0,v)\tau_p(v,w)\tau_p(w,u) \leq T(p) \quad \text{for all } u. \tag{8.21}$$

Assuming that $T(p_c) < \infty$, we may choose $B = B(R)$ sufficiently large that

$$\frac{d\chi}{dp} \geq 2d\alpha(p)\chi^2(1 - \tfrac{1}{2}) \quad \text{for } p \leq p_c. \tag{8.22}$$

Integrate this, as for (8.15), to obtain that

$$\chi(p) \leq \frac{1}{\alpha'(p_c - p)} \quad \text{for } p \leq p_c$$

where $\alpha' = \alpha'(p)$ is strictly positive and continuous for $0 < p < 1$ (and we have used the fact (6.7) that $\chi(p_c) = \infty$). □

Finally we discuss the verification of the triangle condition $T(p_c) < \infty$. This has been proved for large d (currently $d \geq 19$) by Hara and Slade [176, 177, 178, 179, 180], and is believed to hold for $d \geq 7$. The corresponding condition for a 'spread-out' percolation model, having large but finite-range links rather than nearest-neighbour only, is known to hold for $d > 6$.

The proof that $T(p_c) < \infty$ is long and technical, and is to be found in [178]; since the present author has no significant improvement on that version, the details are not given here. Instead, we survey briefly the structure of the proof.

The triangle function (8.9) involves convolutions, and it is therefore natural to introduce the Fourier transform of the connectivity function $\tau_p(x,y) = P_p(x \leftrightarrow y)$. More generally, if $f : \mathbb{Z}^d \to \mathbb{R}$ is summable, we define

$$\widehat{f}(\theta) = \sum_{x\in\mathbb{Z}^d} f(x)e^{i\theta\cdot x}, \quad \text{for } \theta = (\theta_1, \dots, \theta_d) \in [-\pi, \pi]^d,$$

where $\theta \cdot x = \sum_{j=1}^d \theta_j x_j$. If f is symmetric (i.e., $f(x) = f(-x)$ for all x), then $\widehat{f}$ is real.

We have now that

$$T(p) = \sum_{x,y} \tau_p(0,x)\tau_p(x,y)\tau_p(y,0) = (2\pi)^{-d} \int_{[-\pi,\pi]^d} \widehat{\tau}_p(\theta)^3 d\theta. \tag{8.23}$$

The proof that $T(p_c) < \infty$ involves an upper bound on $\widehat{\tau}_p$, namely the so called *infra-red bound*

(8.24) $$\widehat{\tau}_p(\theta) \le \frac{c(p)}{|\theta|^2}$$

where $|\theta| = \sqrt{\theta \cdot \theta}$. It is immediate via (8.23) that the infra-red bound (8.24) implies that $T(p) < \infty$. Also, if (8.24) holds for some $c(p)$ which is uniformly bounded for $p < p_c$, then $T(p_c) = \lim_{p \uparrow p_c} T(p) < \infty$.

It is believed that

(8.25) $$\widehat{\tau}_p(\theta) \simeq \frac{1}{|\theta|^{2-\eta}} \quad \text{as } |\theta| \to 0$$

where η is the critical exponent given in the table of Section 8.2.

Theorem 8.26 (Hara–Slade [178]). *There exists D satisfying $D > 6$ such that, if $d \ge D$, then*

$$\widehat{\tau}_p(\theta) \le \frac{c(p)}{|\theta|^2}$$

for some $c(p)$ which is uniformly bounded for $p < p_c$. Also $T(p_c) < \infty$.

The proof is achieved by establishing and using the following three facts:

(a) $T(p)$ and

$$W(p) = \sum_{x \in \mathbb{Z}^d} |x|^2 \tau_p(0,x)^2$$

are continuous for $p \le p_c$;

(b) there exist constants k_T and k_W such that

$$T(p) \le 1 + \frac{k_T}{d}, \quad W(p) \le \frac{k_W}{d}, \qquad \text{for } p \le \frac{1}{2d};$$

(c) for large d, and for p satisfying $(2d)^{-1} \le p < p_c$, we have that

$$T(p) \le 1 + \frac{3k_T}{d}, \quad W(p) \le \frac{3k_W}{d}, \quad p \le \frac{3}{2d}$$

whenever

$$T(p) \le 1 + \frac{4k_T}{d}, \quad W(p) \le \frac{4k_W}{d}, \quad p \le \frac{4}{2d}.$$

Fact (a) is a consequence of the continuity of τ_p and monotone convergence. Fact (b) follows by comparison with a simpler model (the required comparison is successful for sufficiently small p, namely $p \le (2d)^{-1}$). Fact (c) is much harder to prove, and it is here that the 'lace expansion' is used. Part (c) implies that there is a 'forbidden region' for the pairs $(p, T(p))$ and $(p, W(p))$; see Figure 8.2. Since T and W are finite for small p, and continuous up to p_c, part (c) implies that

$$T(p_c) \le 1 + \frac{3k_T}{d}, \quad W(p_c) \le \frac{3k_W}{d}, \quad p_c \le \frac{3}{2d}.$$

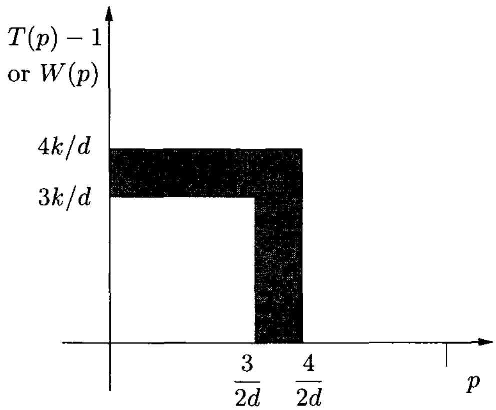


Fig. 8.2. There is a 'forbidden region' for the pairs $(p, T(p) - 1)$ and $(p, W(p))$, namely the shaded region in this figure. The quantity k denotes k_T or k_W as appropriate.

The infra-red bound emerges in the proof of (c), of which there follows an extremely brief account.

We write $x \Leftrightarrow y$, and say that x is 'doubly connected' to y, if there exist two edge-disjoint open paths from x to y. We express $\tau_p(0,x)$ in terms of the 'doubly connected' probabilities $\delta_p(u,v) = P_p(u \Leftrightarrow v)$. In doing so, we encounter formulae involving convolutions, which may be treated by taking transforms. At the first stage, we have that

$$\{0 \leftrightarrow x\} = \{0 \Leftrightarrow x\} \cup \left\{ \bigcup_{\langle u,v\rangle} \{0 \Leftrightarrow (u,v) \leftrightarrow x\} \right\}$$

where $\{0 \Leftrightarrow (u,v) \leftrightarrow x\}$ represents the event that $\langle u,v\rangle$ is the 'first pivotal edge' for the event $\{0 \leftrightarrow x\}$, and that 0 is doubly connected to u. (Similar but more complicated events appear throughout the proof.) Therefore

$$\tau_p(0,x) = \delta_p(0,x) + \sum_{\langle u,v\rangle} P_p\big(0 \Leftrightarrow (u,v) \leftrightarrow x\big). \tag{8.27}$$

Now, with $A(0,u;v,x) = \{v \leftrightarrow x \text{ off } C_{\langle u,v\rangle}(0)\}$,

$$\begin{aligned} P_p\big(0 \Leftrightarrow (u,v) \leftrightarrow x\big) &= pP_p\big(0 \Leftrightarrow u,\ A(0,u;v,x)\big) \\ &= p\delta_p(0,u)\tau_p(v,x) - pE_p\Big(1_{\{0\Leftrightarrow u\}}\big\{\tau_p(v,x) - 1_{A(0,u;v,x)}\big\}\Big) \end{aligned}$$

whence, by (8.27),

$$\tau_p(0,x) = \delta_p(0,x) + \delta_p \star (pI) \star \tau_p(x) - R_{p,0}(0,x) \tag{8.28}$$

where $\star$ denotes convolution, I is the nearest-neighbour function $I(u,v) = 1$ if and only if $u \sim v$, and $R_{p,0}$ is a remainder.

Equation (8.28) is the first step of the lace expansion, In the second step, the remainder $R_{p,0}$ is expanded similarly, and so on. Such further expansions yield the lace expansion: if $p < p_c$ then

$$\tau_p(0,x) = h_{p,N}(0,x) + h_{p,N} \star (pI) \star \tau_p(x) + (-1)^{N+1} R_{p,N}(0,x) \tag{8.29}$$

for appropriate remainders $R_{p,N}$, and where

$$h_{p,N}(0,x) = \delta_p(0,x) + \sum_{j=1}^{N} (-1)^j \Pi_{p,j}(0,x)$$

and the $\Pi_{p,n}$ are appropriate functions (see Theorem 4.2 of [179]) involving nested expectations of quantities related to 'double connections'.

We take Fourier transforms of (8.29), and solve to obtain that

$$\widehat{\tau}_p = \frac{\widehat{\delta}_p + \sum_{j=1}^{N} (-1)^j \widehat{\Pi}_{p,j} + (-1)^{N+1} \widehat{R}_{p,N}}{1 - p\widehat{I}\widehat{\delta}_p - p\widehat{I} \sum_{j=1}^{N} (-1)^j \widehat{\Pi}_{p,j}}. \tag{8.30}$$

The convergence of the lace expansion, and the consequent validity of this formula for $\widehat{\tau}_p$, is obtained roughly as follows. First, one uses the BK inequality to derive bounds for the δ_p, $\Pi_{p,j}$, $R_{p,j}$ in terms of the functions $T(p)$ and $W(p)$. These bounds then imply bounds for the corresponding transforms. In this way, one may obtain a conclusion which is close to point (c) stated above.

9. PERCOLATION IN TWO DIMENSIONS

9.1 The Critical Probability is $\frac{1}{2}$

The famous exact calculation for bond percolation on $\mathbb{L}^2$ is the following, proved originally by Kesten [200]. The proof given here is taken from [G].

Theorem 9.1. *The critical probability of bond percolation on $\mathbb{Z}^2$ equals $\frac{1}{2}$. Furthermore, $\theta(\frac{1}{2}) = 0$.*

Proof. Zhang discovered a beautiful proof that $\theta(\frac{1}{2}) = 0$, using only the uniqueness of the infinite cluster. Set $p = \frac{1}{2}$. Let $T(n)$ be the box $T(n) = [0, n]^2$, and find N sufficiently large that

$$P_{\frac{1}{2}}\big(\partial T(n) \leftrightarrow \infty\big) > 1 - \frac{1}{8^4} \qquad \text{for } n \geq N.$$

We set $n = N + 1$. Writing $A^{\mathrm{l}}, A^{\mathrm{r}}, A^{\mathrm{t}}, A^{\mathrm{b}}$ for the (respective) events that the left, right, top, bottom sides of $T(n)$ are joined to ∞ off $T(n)$, we have by the FKG inequality that

$$\begin{aligned} P_{\frac{1}{2}}\big(T(n) \not\leftrightarrow \infty\big) &= P_{\frac{1}{2}}(\overline{A^{\mathrm{l}}} \cap \overline{A^{\mathrm{r}}} \cap \overline{A^{\mathrm{t}}} \cap \overline{A^{\mathrm{b}}}) \\ &\geq P_{\frac{1}{2}}(\overline{A^{\mathrm{l}}})P(\overline{A^{\mathrm{r}}})P(\overline{A^{\mathrm{t}}})P(\overline{A^{\mathrm{b}}}) \\ &= P_{\frac{1}{2}}(\overline{A^{\mathrm{g}}})^4 \end{aligned}$$

by symmetry, for g = l,r,t,b. Therefore

$$P_{\frac{1}{2}}(A^{\mathrm{g}}) \geq 1 - \Big(1 - P_{\frac{1}{2}}\big(T(n) \leftrightarrow \infty\big)\Big)^{1/4} > \tfrac{7}{8}.$$

Now we move to the dual box, with vertex set $T(n)_{\mathrm{d}} = \{x + (\frac{1}{2}, \frac{1}{2}) : 0 \leq x_1, x_2 < n\}$. Let $A^{\mathrm{l}}_{\mathrm{d}}, A^{\mathrm{r}}_{\mathrm{d}}, A^{\mathrm{t}}_{\mathrm{d}}, A^{\mathrm{b}}_{\mathrm{d}}$ denote the (respective) events that the left, right, top, bottom sides of $T(n)_{\mathrm{d}}$ are joined to ∞ by a *closed* dual path off $T(n)_{\mathrm{d}}$. Since each edge of the dual is closed with probability $\frac{1}{2}$, we have that

$$P_{\frac{1}{2}}(A^{\mathrm{g}}_{\mathrm{d}}) > \tfrac{7}{8} \qquad \text{for g = l,r,t,b.}$$

Consider the event $A = A^{\mathrm{l}} \cap A^{\mathrm{r}} \cap A^{\mathrm{t}}_{\mathrm{d}} \cap A^{\mathrm{b}}_{\mathrm{d}}$, and see Figure 9.1. Clearly $P_{\frac{1}{2}}(\overline{A}) \leq \frac{1}{2}$, so that $P_{\frac{1}{2}}(A) \geq \frac{1}{2}$. However, on A, either $\mathbb{L}^2$ has two infinite open clusters, or its dual has two infinite closed clusters. Each event has probability 0, a contradiction. We deduce that $\theta(\frac{1}{2}) = 0$, implying that $p_{\mathrm{c}} \geq \frac{1}{2}$.

Next we prove that $p_{\mathrm{c}} \leq \frac{1}{2}$. Suppose instead that $p_{\mathrm{c}} > \frac{1}{2}$, so that

$$P_{\frac{1}{2}}\big(0 \leftrightarrow \partial B(n)\big) \leq e^{-\gamma n} \qquad \text{for all } n, \tag{9.2}$$

for some $\gamma > 0$. Let $S(n)$ be the graph with vertex set $\{x \in \mathbb{Z}^2 : 0 \leq x_1 \leq n+1, 0 \leq x_2 \leq n\}$ and edge set containing all edges inherited from $\mathbb{L}^2$ except those in either

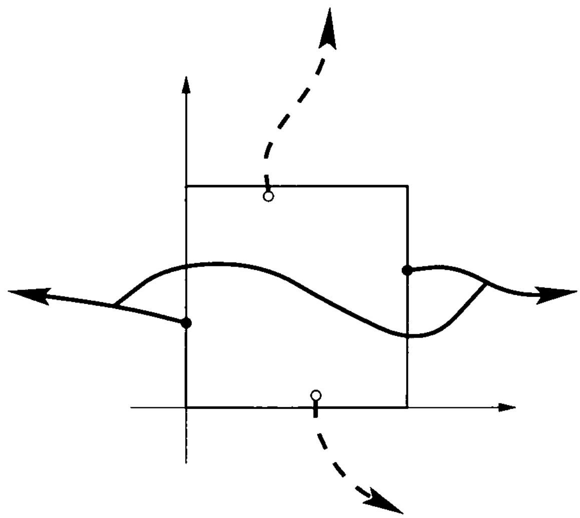


Fig. 9.1. The left and right sides of the box are joined to infinity by open paths of the primal lattice, and the top and bottom sides are joined to infinity by closed dual paths. Using the uniqueness of the infinite open cluster, the two open paths must be joined. This forces the existence of two disjoint infinite closed clusters in the dual.

the left side or the right side of $S(n)$. Denote by A the event that there is an open path joining the left side and right side of $S(n)$. Using duality, if A does not occur, then the top side of the dual of $S(n)$ is joined to the bottom side by a closed dual path. Since the dual of $S(n)$ is isomorphic to $S(n)$, and since $p = \frac{1}{2}$, it follows that $P_{\frac{1}{2}}(A) = \frac{1}{2}$. See Figure 9.2. However, using (9.2),

$$P_{\frac{1}{2}}(A) \leq (n+1)e^{-\gamma n},$$

a contradiction for large n. We deduce that $p_c \leq \frac{1}{2}$. □

9.2 RSW Technology

Substantially more is known about the phase transition in two dimensions than in higher dimensions. The main reason for this lies in the fact that geometrical constraints force the intersection of certain paths in two dimensions, whereas they can avoid one another in three dimensions. Path-intersection properties play a central role in two dimensions, whereas in higher dimensions we have to rely on the more complicated Grimmett–Marstrand construction of Section 7.3.

A basic tool in two dimensions is the RSW lemma, which was discovered independently by Russo [326] and Seymour–Welsh [331]. Consider the rectangle $B(kl, l) = [-l, (2k-1)l] \times [-l, l]$, a rectangle of side-lengths $2kl$ and $2l$; note that $B(l, l) = B(l)$. We write $\mathrm{LR}(l)$ for the event that $B(l)$ is crossed from left to right by an open path, and $\mathrm{O}(l)$ for the event that there is an open circuit of the annulus $A(l) = B(3l) \setminus B(l)$ containing the origin in its interior.

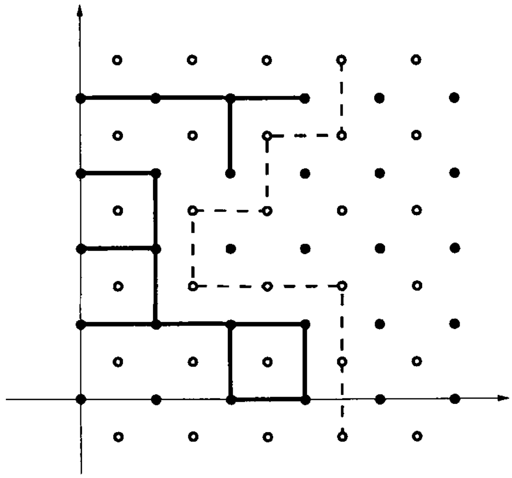


Fig. 9.2. If there is no open left-right crossing of $S(n)$, then there must be a closed top-bottom crossing in the dual.

Theorem 9.3 (RSW Lemma). *If* $P_p(\mathrm{LR}(l)) = \tau$ *then*

$$P_p\big(\mathrm{O}(l)\big) \geq \{\tau(1-\sqrt{1-\tau})^4\}^{12}.$$

When $p = \frac{1}{2}$, we have from self-duality that $P_{\frac{1}{2}}(\mathrm{LR}(l)) \geq \frac{1}{4}$ for $l \geq 1$, whence

$$P_{\frac{1}{2}}\big(\mathrm{O}(l)\big) \geq 2^{-24}\left(1-\frac{\sqrt{3}}{2}\right)^{48} \qquad \text{for } l \geq 1. \tag{9.4}$$

We refer the reader to [G] for a proof of the RSW lemma. In common with almost every published proof of the lemma ([331] is an exception, possibly amongst others), the proof given in [G] contains a minor error. Specifically, the event G below (9.80) on page 223 is not increasing, and therefore we may not simply use the FKG inequality at (9.81). Instead, let A_π be the event that the path π is open. Then, in the notation of [G],

$$\begin{aligned}
P_{\frac{1}{2}}\big(\mathrm{LR}(\tfrac{3}{2}l, l)\big) &\geq P_{\frac{1}{2}}\left(N^+ \cap \Big(\bigcup_{\pi \in \mathcal{T}^-} [A_\pi \cap M_\pi^-]\Big)\right) \\
&\geq P_{\frac{1}{2}}(N^+) P_{\frac{1}{2}}\Big(\bigcup_\pi [A_\pi \cap M_\pi^-]\Big) && \text{by FKG} \\
&\geq P_{\frac{1}{2}}(N^+) \sum_\pi P_{\frac{1}{2}}(L_\pi \cap M_\pi^-) \\
&\geq P_{\frac{1}{2}}(N^+)(1-\sqrt{1-\tau}) \sum_\pi P_{\frac{1}{2}}(L_\pi) && \text{by (9.84)} \\
&= P_{\frac{1}{2}}(N^+)(1-\sqrt{1-\tau}) P_{\frac{1}{2}}(L^-) \\
&\geq (1-\sqrt{1-\tau})^3 && \text{as in [G].}
\end{aligned}$$

There are several applications of the RSW lemma, of which we present one.

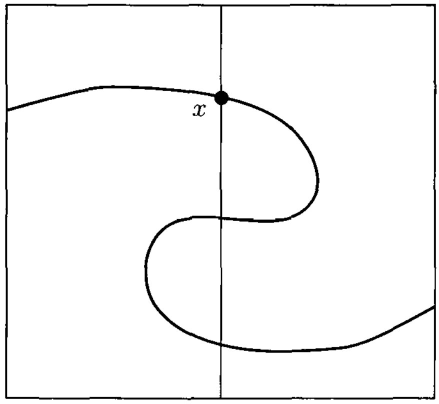


Fig. 9.3. If there is an open left-right crossing of the box, then there must exist some vertex x in the centre which is connected *disjointly* to the left and right sides.

Theorem 9.5. *There exist constants* A, α *satisfying* $0 < A, \alpha < \infty$ *such that*

$$\tag{9.6} \tfrac{1}{2}n^{-1/2} \le P_{\frac{1}{2}}\big(0 \leftrightarrow \partial B(n)\big) \le An^{-\alpha}.$$

Similar power-law estimates are known for other macroscopic quantities at and near the critical point $p_c = \frac{1}{2}$. In the absence of a proof that quantities have 'power-type' singularities near the critical point, it is reasonable to look for upper and lower bounds of the appropriate type. As a general rule, one such bound is usually canonical, and applies to all percolation models (viz. the inequality (7.36) that $\theta(p) - \theta(p_c) \ge a(p - p_c)$). The complementary bound is harder, and is generally unavailable at the moment when $d \ge 3$ (but d is not too large).

Proof. Let $R(n) = [0, 2n] \times [0, 2n - 1]$, and let $\mathrm{LR}(n)$ be the event that $R(n)$ is traversed from left to right by an open path. We have by self-duality that $P_{\frac{1}{2}}(\mathrm{LR}(n)) = \frac{1}{2}$. On the event $\mathrm{LR}(n)$, there exists a vertex x with $x_1 = n$ such that $x \leftrightarrow x + \partial B(n)$ by two *disjoint* open paths. See Figure 9.3. Therefore

$$\tfrac{1}{2} = P_{\frac{1}{2}}\big(\mathrm{LR}(n)\big) \le \sum_{k=0}^{2n-1} P_{\frac{1}{2}}(A_k \circ A_k) \le 2nP_{\frac{1}{2}}\big(0 \leftrightarrow \partial B(n)\big)^2$$

where $A_k = \{(n,k) \leftrightarrow (n,k) + \partial B(n)\}$, and we have used the BK inequality. This provides the lower bound in (9.6).

For the upper bound, we have from (9.4) that $P_{\frac{1}{2}}(\mathrm{O}(l)) \ge \xi$ for all l, where $\xi > 0$. Now, on the event $\{0 \leftrightarrow \partial B(n)\}$, there can be no closed dual circuit surrounding the origin and contained within $B(n)$. In particular, no dual annulus of the form $B(3^{r+1}) \backslash B(3^r) + (\frac{1}{2}, \frac{1}{2})$, for $0 \le r < \log_3 n - 1$, can contain such a closed circuit. Therefore

$$\tag{9.7} P_{\frac{1}{2}}\big(0 \leftrightarrow \partial B(n)\big) \le (1 - \xi)^{\log_3 n - 2}$$

as required for the upper bound in (9.6). $\square$

9.3 Conformal Invariance

We concentrate on bond percolation in two dimensions with $p = p_c = \frac{1}{2}$. With $S(n) = [0, n+1] \times [0, n]$, we have by self-duality that

$$P_{\frac{1}{2}}\big(S(n) \text{ traversed from left to right by open path}\big) = \tfrac{1}{2} \tag{9.8}$$

for all n. Certainly $\mathbb{L}^2$ must contain long open paths, but no infinite paths (since $\theta(\frac{1}{2}) = 0$). One of the features of (hypothetical) universality is that the chances of long-range connections (when $p = p_c$) should be independent of the choice of lattice structure. In particular, local deformations of space should, within limits, not affect such probabilities. One family of local changes arises by local rotations and dilations, and particularly by applying a conformally invariant mapping to $\mathbb{R}^2$. This suggests the possibility that long-range crossing probabilities are, in some sense to be explored, invariant under conformal maps of $\mathbb{R}^2$. (See [8] for an account of conformal maps.)

Such a hypothesis may be formulated, and investigated numerically. Such a programme has been followed by Langlands, Pouliot, and Saint-Aubin [231] and Aizenman [10], and their results support the hypothesis. In this summary, we refer to bond percolation on $\mathbb{L}^2$ only, although such conjectures may be formulated for any two-dimensional percolation model.

We begin with a concrete conjecture concerning crossing probabilities. Let $B(kl, l)$ be a $2kl$ by $2l$ rectangle, and let $\mathrm{LR}(kl, l)$ be the event that $B(kl, l)$ is traversed between its opposite sides of length $2l$ by an open path, as in Section 9.2. It is not difficult to show, using (9.8), that

$$P_{\frac{1}{2}}\big(\mathrm{LR}(l, l)\big) \to \tfrac{1}{2} \qquad \text{as } l \to \infty,$$

and it is reasonable to conjecture that the limit

$$\lambda_k = \lim_{l \to \infty} P_{\frac{1}{2}}\big(\mathrm{LR}(kl, l)\big) \tag{9.9}$$

exists for all $0 < k < \infty$. By self-duality, we have that $\lambda_k + \lambda_{k^{-1}} = 1$ if the λ_k exist. It is apparently difficult to establish the limit in (9.9).

In [231] we see a generalisation of this conjecture which is fundamental for a Monte Carlo approach to conformal invariance. Take a simple closed curve C in the plane, and arcs $\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \ldots, \beta_m$, as well as arcs $\gamma_1, \gamma_2, \ldots, \gamma_n, \delta_1, \ldots, \delta_n$, of C. For a dilation factor r, define

$$\pi_r(G) = P(r\alpha_i \leftrightarrow r\beta_i,\ r\gamma_i \nleftrightarrow r\delta_i, \text{ for all } i, \text{ in } rC) \tag{9.10}$$

where $P = P_{p_c}$ and G denotes the collection $(C; \alpha_i, \beta_i; \gamma_i, \delta_i)$.

Conjecture 9.11. *The following limit exists:*

$$\pi(G) = \lim_{r\to\infty} \pi_r(G).$$

Some convention is needed in order to make sense of (9.10), arising from the fact that rC lives in the plane $\mathbb{R}^2$ rather than on the lattice $\mathbb{L}^2$; this poses no major problem. Conjecture (9.9) is a special case of (9.11), with $C = B(k,1)$, and α_1, β_1 being the left and right sides of the box.

Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be a reasonably smooth function. The composite object $G = (C; \alpha_i, \beta_i; \gamma_i, \delta_i)$ has an image under ϕ, namely $\phi G = (\phi C; \phi\alpha_i, \phi\beta_i; \phi\gamma_i, \phi\delta_i)$, which itself corresponds to an event concerning the existence or non-existence of certain open paths. If we believe that crossing probabilities are not affected (as $r \to \infty$, in (9.10)) by local dilations and rotations, then it becomes natural to formulate a conjecture of invariance under conformal maps [10, 231].

Conjecture 9.12 (Conformal Invariance). *For all $G = (C; \alpha_i, \beta_i; \gamma_i, \delta_i)$, we have that $\pi(\phi G) = \pi(G)$ for any $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ which is bijective on C and conformal on its interior.*

Lengthy computer simulations, reported in [231], support this conjecture. Particularly stimulating evidence is provided by a formula known as Cardy's formula [87]. By following a sequence of transformations of models, and applying ideas of conformal field theory, Cardy was led to an explicit formula for crossing probabilities between two sub-intervals of a simple closed curve C.

Let C be a simple closed curve, and let z_1, z_2, z_3, z_4 be four points on C in clockwise order. There is a conformal map ϕ on the interior of C which maps to the unit disc, taking C to its circumference, and the points z_i to the points w_i. There are many such maps, but the cross-ratio of such maps,

$$u = \frac{(w_4 - w_3)(w_2 - w_1)}{(w_3 - w_1)(w_4 - w_2)}, \tag{9.13}$$

is a constant satisfying $0 \le u \le 1$ (we think of z_i and w_i as points in the complex plane). We may parametrise the w_i as follows: we may assume that

$$w_1 = e^{i\theta}, \quad w_2 = e^{-i\theta}, \quad w_3 = -e^{i\theta}, \quad w_4 = -e^{-i\theta}$$

for some θ satisfying $0 \le \theta \le \frac{\pi}{2}$. Note that $u = \sin^2\theta$. We take α to be the segment of C from z_1 to z_2, and β the segment from z_3 to z_4. Using the hypothesis of conformal invariance, we have that $\pi(G) = \pi(\phi G)$, where $G = (C; \alpha, \beta; \varnothing, \varnothing)$, implying that $\pi(G)$ may be expressed as some function $f(u)$, where u is given in (9.13). Cardy has derived a differential equation for f, namely

$$u(1-u)f''(u) + \tfrac{2}{3}(1-2u)f'(u) = 0, \tag{9.14}$$

together with the boundary conditions $f(0) = 0, f(1) = 1$. The solution is a hypergeometric function,

$$f(u) = \frac{3\Gamma(\frac{2}{3})}{\Gamma(\frac{1}{3})^2}\, u^{1/3}\, {}_2F_1(\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{4}{3}; u). \tag{9.15}$$

Recall that $u = \sin^2 \theta$. The derivation is somewhat speculative, but the predictions of the formula may be verified by Monte Carlo simulation (see Figure 3.2 of [231]).

The above 'calculation' is striking. Similar calculations may well be possible for more complicated crossing probabilities than the case treated above. See, for example, [10, 345].

In the above formulation, the principle of conformal invariance is expressed in terms of a collection $\{\pi(G)\}$ of limiting 'crossing probabilities'. It would be useful to have a representation of these $\pi(G)$ as probabilities associated with a specific random variable on a specific probability space. Aizenman [10] has made certain proposals about how this might be possible. In his formulation, we observe a bounded region $D_R = [0, R]^2$, and we shrink the lattice spacing a of bond percolation restricted to this domain. Let $p = p_c$, and let G_a be the graph of open connections of bond percolation with lattice spacing a on D_R. By describing G_a through the set of Jordan curves describing the realised paths, he has apparently obtained sufficient compactness to imply the existence of weak limits as $a \to 0$. Possibly there is a *unique* weak limit, and Aizenman has termed an object sampled according to this limit as the 'web'. The fundamental conjectures are therefore that there is a unique weak limit, and that this limit is conformally invariant.

The quantities $\pi(G)$ should then arise as crossing probabilities in 'web-measure'. This geometrical vision may be useful to physicists and mathematicians in understanding conformal invariance.

In one interesting 'continuum' percolation model, conformal invariance may actually be proved rigorously. Drop points $\{X_1, X_2, \dots\}$ in the plane $\mathbb{R}^2$ in the manner of a Poisson process with intensity λ. Now divide $\mathbb{R}^2$ into tiles $\{T(X_1), T(X_2), \dots\}$, where $T(X)$ is defined as the set of points in $\mathbb{R}^2$ which are no further from X than they are from any other point of the Poisson process (this is the 'Voronoi tessela-tion'). We designate each tile to be open with probability $\frac{1}{2}$ and closed otherwise. This continuum percolation model has a property of self-duality, and it inherits properties of conformal invariance from those of the underlying Poisson point process. See [11, 57].

10. RANDOM WALKS IN RANDOM LABYRINTHS

10.1 Random Walk on the Infinite Percolation Cluster

It is a classical result that symmetric random walk on $\mathbb{L}^d$ is recurrent when $d = 2$ but transient when $d \geq 3$ (see [170], pages 188, 266). Three-dimensional space is sufficiently large that a random walker may become lost, whereas two-dimensional space is not. The transience or recurrence of a random walk on a graph G is a crude measure of the 'degree of connectivity' of G, a more sophisticated measure being the transition probabilities themselves. In studying the geometry of the infinite open percolation cluster, we may ask whether or not a random walk on this cluster is recurrent.

Theorem 10.1. *Suppose $p > p_c$. Random walk on the (a.s. unique) infinite open cluster is recurrent when $d = 2$ and a.s. transient when $d \geq 3$.*

This theorem, proved in [164][6], follows by a consideration of the infinite open cluster viewed as an electrical network. The relationship between random walks and electrical networks is rather striking, and has proved useful in a number of contexts; see [118].

We denote the (a.s.) unique infinite open cluster by $I = I(\omega)$, whenever it exists. On the graph I, we construct a random walk as follows. First, we set $S_0 = x$ where x is a given vertex of I. Given $S_0, S_1, \dots, S_n$, we specify that S_{n+1} is chosen uniformly from the set of neighbours of S_n in I, this choice being independent of all earlier choices. We call ω a *transient configuration* if the random walk is transient, and a *recurrent configuration* otherwise. Since I is connected, the transience or recurrence of S does not depend on the choice of the starting point x.

The corresponding electrical network arises as follows. For $x \in I$, we denote by $B_n(x)$ the set of all vertices y of I such that $\delta(x, y) \leq n$, and we write $\partial B_n(x) = B_n(x) \backslash B_{n-1}(x)$. We turn $B_n(x)$ into a graph by adding all induced (open) edges of I. Next we turn this graph into an electrical network by replacing each edge by a unit resistor, and by 'shorting together' all vertices in $\partial B_n(x)$. Let $R_n(x)$ be the effective resistance of the network between x and the composite vertex $\partial B_n(x)$.

By an argument using monotonicity of effective resistance (as a function of the individual resistances), the increasing limit $R_\infty(x) = \lim_{n\to\infty} R_n(x)$ exists for all x. It is a consequence of the relationship between random walk and electrical networks that the random walk on I, beginning at x, is transient if and only if $R_\infty(x) < \infty$. Therefore Theorem 10.1 is a consequence of the following.

Theorem 10.2. *Let $p > p_c$, and let I be the (a.s.) unique infinite open cluster.*
(a) *If $d = 2$ then $R_\infty(x) = \infty$ for all $x \in I$.*
(b) *If $d \geq 3$ then $P_p(R_\infty(0) < \infty \mid 0 \in I) = 1$.*

Part (a) is obvious, as follows. The electrical resistance of a graph can only increase if any individual edge-resistance is increased. Since the network on I may be obtained from that on $\mathbb{L}^2$ by setting the resistances of closed edges to ∞, we have that $R_\infty(x)$ is no smaller than the resistance 'between x and ∞' of $\mathbb{L}^2$. The

[6] For a quite different and more recent approach, see [54].

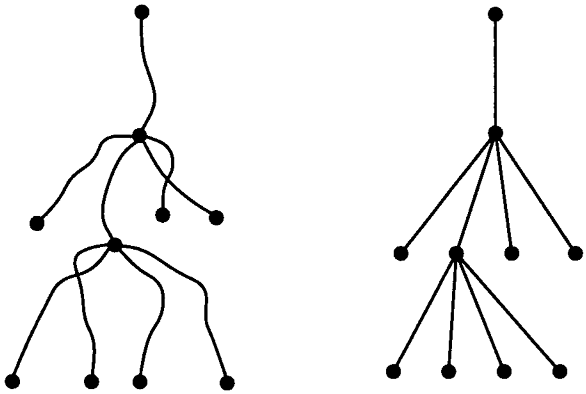


Fig. 10.1. The left picture depicts a tree-like subgraph of the lattice. The right picture is obtained by the removal of common points and the replacement of component paths by single edges. The resistance of such an edge emanating from the kth generation has order β^k.

latter resistance is infinite (since random walk is recurrent, or by direct estimation), implying that $R_\infty(x) = \infty$.

Part (b) is harder, and may be proved by showing that I contains a subgraph having *finite* resistance. We begin with a sketch of the proof. Consider first a tree T of 'down-degree' 4; see Figure 10.1. Assume that any edge joining the kth generation to the $(k+1)$th generation has electrical resistance β^k where $\beta > 1$. Using the series and parallel laws, the resistance of the tree, between the root and infinity, is $\sum_k (\beta/4)^k$; this is finite if $\beta < 4$. Now we do a little geometry. Let us try to imbed such a tree in the lattice $\mathbb{L}^3$, in such a way that the vertices of the tree are vertices of the lattice, and that the edges of the tree are paths of the lattice which are 'almost' disjoint. Since the resistance from the root to a point in the kth generation is

$$\sum_{r=0}^{k-1} \beta^r = \frac{\beta^k - 1}{\beta - 1},$$

it is reasonable to try to position the kth generation vertices on or near the surface $\partial B(\beta^{k-1})$. The number of kth generation vertices if 4^k, and the volume of $\partial B(\beta^{k-1})$ has order $\beta^{(k-1)(d-1)}$. The above construction can therefore only succeed when $4^k < \beta^{(k-1)(d-1)}$ for all large k, which is to say that $\beta > 4^{1/(d-1)}$.

This crude picture suggests the necessary inequalities

$$4^{1/(d-1)} < \beta < 4, \tag{10.3}$$

which can be satisfied if and only if $d \geq 3$.

Assume now that $d \geq 3$. Our target is to show that the infinite cluster I contains sufficiently many disjoint paths to enable a comparison of its effective resistance with that of the tree in Figure 10.1, and with some value of β satisfying (10.3). In presenting a full proof of this, we shall use the following two percolation estimates, which are consequences of the results of Chapter 7.

Lemma 10.4. *Assume that $p > p_c$.*
(a) *There exists a strictly positive constant $\gamma = \gamma(p)$ such that*

$$P_p(B(n) \leftrightarrow \infty) \geq 1 - e^{-\gamma n} \qquad \text{for all } n. \tag{10.5}$$

(b) *Let $\sigma > 1$, and let $A(n, \sigma)$ be the event that there exist two vertices inside $B(n)$ with the property that each is joined by an open path to $\partial B(\sigma n)$ but that there is no open path of $B(\sigma n)$ joining these two vertices. There exists a strictly positive constant $\delta = \delta(p)$ such that*

$$P_p(A(n, \sigma)) \leq e^{-\delta n(\sigma - 1)} \qquad \text{for all } n. \tag{10.6}$$

We restrict ourselves here to the case $d = 3$; the general case $d \geq 3$ is similar. The surface of $B(n)$ is the union of six faces, and we concentrate on the face

$$F(n) = \{x \in \mathbb{Z}^3 : x_1 = n, \ |x_2|, |x_3| \leq n\}.$$

We write $B_k = B(3^k)$ and $F_k = F(3^k)$. On F_k, we distinguish 4^k points, namely

$$x_k(i, j) = (id_k, jd_k), \quad -2^{k-1} < i, j \leq 2^{k-1}$$

where $d_k = \lfloor (4/3)^k \rfloor$. The $x_k(i, j)$ are distributed on F_k in the manner of a rectangular grid, and they form the 'centres of attraction' corresponding to the kth generation of the tree discussed above.

With each $x_k(i, j)$ we associate four points on F_{k+1}, namely those in the set

$$I_k(i, j) = \{x_{k+1}(r, s) : r = 2i - 1, 2i, \ s = 2j - 1, 2j\}.$$

These four points are called *children* of $x_k(i, j)$. The centroid of $I_k(i, j)$ is denoted $\bar{I}_k(i, j)$. We shall attempt to construct open paths from points near $x_k(i, j)$ to points near to each member of $I_k(i, j)$, and this will be achieved with high probability. In order to control the geometry of such paths, we shall build them within certain 'tubes' to be defined next.

Write $L(u, v)$ for the set of vertices lying within euclidean distance $\sqrt{3}$ of the line segment of $\mathbb{R}^3$ joining u to v. Let $a > 0$. Define the region

$$T_k(i, j) = A_k(i, j) \cup C_k(i, j)$$

where

$$A_k(i, j) = B(ak) + L(x_k(i, j), \bar{I}_k(i, j))$$
$$C_k(i, j) = B(ak) + \bigcup_{x \in I_k(i,j)} L(\bar{I}_k(i, j), x).$$

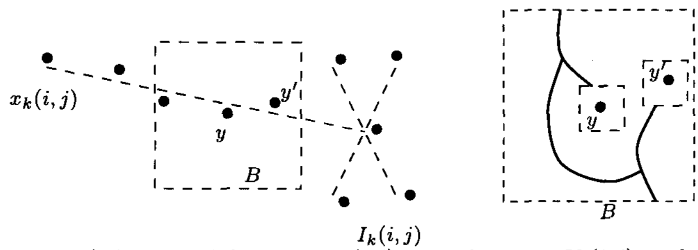


Fig. 10.2. A diagram of the region $T_k(i,j)$, with the points $Y_k(i,j)$ marked. The larger box is an enlargement of the box B on the left. In the larger box appear open paths of the sort required for the corresponding event E_u, where $y = y_u$. Note that the two smaller boxes within B are joined to the surface of B, and that any two such connections are joined to one another within B.

See Figure 10.2.

Within each $T_k(i,j)$ we construct a set of vertices as follows. In $A_k(i,j)$ we find vertices $y_1, y_2, \ldots, y_t$ such that the following holds. Firstly, there exists a constant ν such that $t \le \nu 3^k$ for all k. Secondly, each y_u lies in $A_k(i,j)$,

$$(10.7)\qquad y_u \in L\big(x_k(i,j), \overline{I}_k(i,j)\big), \quad \tfrac{1}{3}ak \le \delta(y_u, y_{u+1}) \le \tfrac{2}{3}ak$$

for $1 \le u < t$, and furthermore $y_1 = x_k(i,j)$, and $|y_t - \overline{I}_k(i,j)| \le 1$.

Likewise, for each $x \in I_k(i,j)$, we find a similar sequence $y_1(x), y_2(x), \ldots, y_v(x)$ satisfying (10.7) with $x_k(i,j)$ replaced by x, and with $y_1(x) = y_t$, $y_v(x) = x$, and $v = v(x) \le \nu 3^k$.

The set of all such y given above is denoted $Y_k(i,j)$. We now construct open paths using $Y_k(i,j)$ as a form of skeleton. Let $0 < 7b < a$. For $1 \le u < t$, let $E_u = E_u(k,i,j)$ be the event that

(a) there exist $z_1 \in y_u + B(bk)$ and $z_2 \in y_{u+1} + B(bk)$ such that $z_i \leftrightarrow y_u + \partial B(ak)$ for $i = 1, 2$, and

(b) any two points lying in $\{y_u, y_{u+1}\} + B(bk)$ which are joined to $y_u + \partial B(ak)$ are also joined to one another within $y_u + \partial B(ak)$.

We define similar events $E_{x,u} = E_{x,u}(k,i,j)$ for $x \in I_k(i,j)$ and $1 \le u < v = v(x)$, and finally let

$$\mathcal{E}_k(i,j) = \Big\{\bigcap_{1 \le u < t} E_u\Big\} \cap \Big\{\bigcap_{\substack{1 \le u < v(x)\\ x \in I_k(i,j)}} E_{x,u}\Big\}.$$

Let us estimate $P_p(\mathcal{E}_k(i,j))$. Using Lemma 10.4, we have that

$$(10.8)\qquad P_p(\overline{\mathcal{E}_k(i,j)}) \le 5\nu 3^k e^{-\gamma bk} + 5\nu 3^k e^{-\delta ak/6}.$$

We call $x_{m+l}(r,s)$ a *descendant* of $x_m(0,0)$ if it is a child of a child $\ldots$ of a child of $x_m(0,0)$. Write $\mathcal{K}_m$ for the set of all $(m+l, r, s)$ such that $x_{m+l}(r,s)$ is a descendant of $x_m(0,0)$. We have from (10.8) that

$$U_m = \sum_{(k,r,s) \in \mathcal{K}_m} P_p(\overline{\mathcal{E}_k(r,s)}) \le \sum_{k=m}^{\infty} 4^{k-m} 5\nu 3^k \big(e^{-\gamma bk} + e^{-\delta ak/6}\big).$$

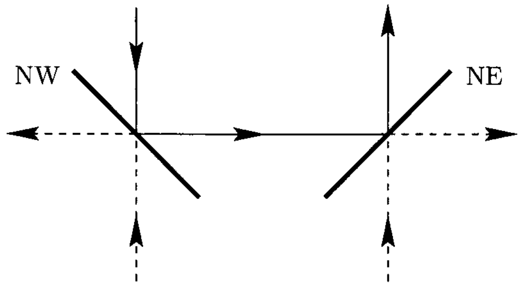


Fig. 10.3. NW and NE reflectors in action.

Now pick a, b such that $0 < 7b < a$ and $e^{-\gamma b}, e^{-\delta a/6} < \frac{1}{12}$, so that $U_m \to 0$ as $m \to \infty$. This implies that there exists a (random) value M of m such that $\mathcal{E}_k(r, s)$ occurs for all $(k, r, s) \in \mathcal{K}_M$.

Turning to the geometry implied in the definition of the $\mathcal{E}_k(r, s)$, we find that the infinite open cluster contains a topological copy of the tree in Figure 10.1, where the length of a path joining a kth generation vertex to one of its children is no greater than $Ck^3 3^k$ for some constant C. In particular, this length is smaller than $C'\beta^k$ for any β satisfying $3 < \beta < 4$ and for some $C' = C'(\beta)$. Choosing β and C' accordingly, and referring to the discussion around (10.3), we conclude that I contains a tree having *finite* resistance between its root and infinity. The second claim of Theorem 10.2 follows.

10.2 Random Walks in Two-Dimensional Labyrinths

A beautiful question dating back to Lorentz [245] and Ehrenfest [129] concerns the behaviour of a particle moving in $\mathbb{R}^d$ but scattered according to reflecting obstacles distributed about $\mathbb{R}^d$. There is a notorious lattice version of this question which is largely unsolved. Start with the two-dimensional square lattice $\mathbb{L}^2$. A reflector may be placed at any vertex in either of two ways: either it is a NW reflector (which deflects incoming rays heading northwards, resp. southwards, to the west, resp. east, and vice versa) or it is a NE reflector (defined similarly); see Figure 10.3. Think of a reflector as being a two-sided mirror placed at 45° to the axes, so that an incoming light ray is reflected along an axis perpendicular to its direction of arrival. Now, for each vertex x, with probability p we place a reflector at x, and otherwise we place nothing at x. This is done independently for different x. If a reflector is placed at x, then we specify that it is equally likely to be NW as NE.

We shine a torch northwards from the origin. The light is reflected by the mirrors, and we ask whether or not the light ray returns to the origin. Letting

$$\eta(p) = P_p(\text{the light ray returns to the origin}),$$

we would like to know for which values of p it is the case that $\eta(p) = 1$. It is reasonable to conjecture that η is non-decreasing in p. Certainly $\eta(0) = 0$, and it is 'well known' that $\eta(1) = 1$.

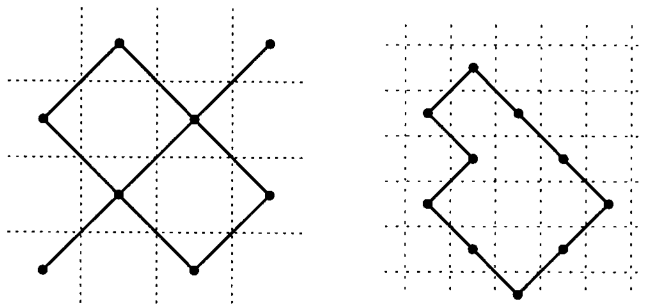


Fig. 10.4. (a) The heavy lines form the lattice $\mathbb{L}_A^2$, and the central point is the origin of $\mathbb{L}^2$. (b) An open circuit in $\mathbb{L}_A^2$ constitutes a barrier of mirrors through which no light may penetrate.

Theorem 10.9. *It is the case the* $\eta(1) = 1$.

Proof of Theorem 10.9. This proof is alluded to in [G] and included in [82]. From $\mathbb{L}^2$ we construct an ancillary lattice $\mathbb{L}_A^2$ as follows. Let

$$A = \left\{(m + \tfrac{1}{2}, n + \tfrac{1}{2}) : m + n \text{ is even}\right\}.$$

On A we define the adjacency relation $\sim$ by $(m + \frac{1}{2}, n + \frac{1}{2}) \sim (r + \frac{1}{2}, s + \frac{1}{2})$ if and only if $|m - r| = |n - s| = 1$, obtaining thereby a copy of $\mathbb{L}^2$ denoted as $\mathbb{L}_A^2$. See Figure 10.4.

We now use the above 'labyrinth' to define a bond percolation process on $\mathbb{L}_A^2$. We declare the edge of $\mathbb{L}_A^2$ joining $(m - \frac{1}{2}, n - \frac{1}{2})$ to $(m + \frac{1}{2}, n + \frac{1}{2})$ to be *open* if there is a NE mirror at (m, n); similarly we declare the edge joining $(m - \frac{1}{2}, n + \frac{1}{2})$ to $(m + \frac{1}{2}, n - \frac{1}{2})$ to be *open* if there is a NW mirror at (m, n). Edges which are not open are designated *closed*. This defines a percolation model in which north-easterly edges (resp. north-westerly edges) are open with probability $p_{\mathrm{NE}} = \frac{1}{2}$ (resp. $p_{\mathrm{NW}} = \frac{1}{2}$). Note that $p_{\mathrm{NE}} + p_{\mathrm{NW}} = 1$, which implies that the percolation model is critical (see [G, 202]).

Let N be the number of open circuits in $\mathbb{L}_A^2$ which contain the origin in their interiors. Using general results from percolation theory, we have that $\mathbb{P}(N \geq 1) = 1$, where $\mathbb{P}$ is the appropriate probability measure. (This follows from the fact that $\theta(\frac{1}{2}) = 0$; cf. Theorem 9.1, see also [G, 181, 202].) However, such an open circuit corresponds to a barrier of mirrors surrounding the origin, from which no light can escape (see Figure 10.4 again). Therefore $\eta(1) = 1$.

We note that the above proof is valid in the slightly more general setting in which NE mirrors are present with density p_{NE} and NW mirrors with density p_{NW} where $p_{\mathrm{NW}} + p_{\mathrm{NE}} = 1$ and $0 < p_{\mathrm{NW}} < 1$. This generalisation was noted in [82]. □

When $0 < p < 1$, the question of whether or not $\eta(p) = 1$ is wide open, despite many attempts to answer it[7]. It has been conjectured that $\eta(p) = 1$ for all $p >$

[7] I heard of this problem in a conversation with Hermann Rost and Frank Spitzer in Heidelberg in 1978. The proof that $\eta(1) = 1$ was known to me (and presumably to others) in 1978 also.

0, based on numerical simulations; see [111, 376]. Some progress has been made recently by Quas [321].

The above lattice version of the 'mirror model' appears to have been formulated first around 20 years ago. In a systematic approach to random environments of reflectors, Ruijgrok and Cohen [325] proposed a programme of study of 'mirror' and 'rotator' models. Since then, there have been reports of many Monte Carlo experiments, and several interesting conjectures have emerged (see [109, 110, 111, 344, 376]). Rigorous progress has been relatively slight; see [82, G, 321] for partial results.

The principal difficulty in the above model resides in the facts that the environment is random but that the trajectory of the light is (conditionally) deterministic. If we relax the latter determinism, then we arrive at model which is more tractable. In this new version, there are exactly three types of point, called *mirrors, crossings*, and *random walk (rw) points*. Let $p_{\mathrm{rw}}, p_+ \geq 0$ be such that $p_{\mathrm{rw}} + p_+ \leq 1$. We designate each vertex x to be

a random walk (rw) point, with probability p_{rw},
a crossing, with probability p_+,
a mirror, otherwise.

If a vertex is a mirror, then it is occupied by a NW reflector with probability $\frac{1}{2}$, and otherwise by a NE reflector. The environment of mirrors and rw points is denoted by $Z = (Z_x : x \in \mathbb{Z}^2)$ and is termed a 'labyrinth'; we write $\mathbb{P}$ for the probability measure associated with the labyrinth, so that $\mathbb{P}$ is product measure on the corresponding environment space.

The physical meaning of these terms is as follows. Suppose that some vertex x is occupied by a candle, which emits light rays along the four axes leaving x. When a ray is incident with a mirror, then it is reflected accordingly. When a ray encounters a crossing, then it continues undeflected. When a ray encounters a rw point, then it leaves this point in one of the four available directions, chosen at random in the manner of a random walk.

We formalise this physical explanation by defining a type of random walk $X = (X_0, X_1, \dots)$ on $\mathbb{L}^2$. Assume that $p_{\mathrm{rw}} > 0$, and sample a random labyrinth Z according to the measure $\mathbb{P}$. Let x be a rw point, and set $X_0 = x$. We choose a random neighbour X_1 of x, each of the four possibilities being equally likely. Having constructed $X_0, X_1, \dots, X_r$, we define X_{r+1} as follows. If X_r is a rw point, we let X_{r+1} be a randomly chosen neighbour of X_r (chosen independently of all earlier choices); if X_r is not a rw point, then we define X_{r+1} to be the next vertex illuminated by a ray of light which is incident with X_r travelling in the direction $X_r - X_{r-1}$. The consequent sequence X is called a 'random walk in a random labyrinth'. Let P_x^Z denote the law of X, conditional on Z, and starting at x. We say that the rw point x is *Z-recurrent* if there exists (P_x^Z-a.s.) an integer N such that $X_N = x$, and otherwise we say that x is *Z-transient*. We say that the labyrinth Z is *recurrent* if every rw point is Z-recurrent. It is easily seen, using the translation-invariance of $\mathbb{P}$ and the zero–one law, that the labyrinth is $\mathbb{P}$-a.s. recurrent if and only if

$$\mathbb{P}\big(0 \text{ is } Z\text{-recurrent} \,\big|\, 0 \text{ is a rw point}\big) = 1.$$

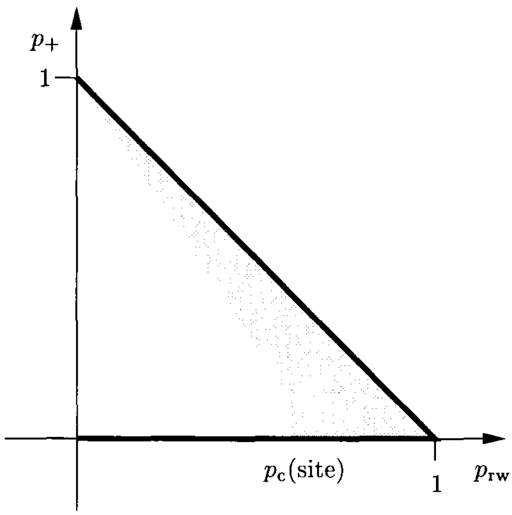


Fig. 10.5. The grey region and the heavy lines of the figure indicate the part of (p_{rw}, p_+) space for which non-localisation is proved. The labyrinth is a.s. localised when $p_{rw} = p_+ = 0$; see Theorem 10.9.

Theorem 10.10. *If $p_{rw} > 0$ then the labyrinth Z is $\mathbb{P}$-a.s. recurrent.*

This theorem, together with most other results in this chapter, appears in [166], and is proved by showing that a corresponding electrical network has infinite resistance. A brief proof appears at the end of this section.

Remembering that irreducible Markov chains on *finite* state spaces are necessarily recurrent, we turn our attention to a question of 'localisation'. Let x be a rw point in the random labyrinth Z, and let X be constructed as above. We say that x is *Z-localised* if X visits (P_x^Z-a.s.) only finitely many vertices; we call x *Z-non-localised* otherwise. We say that the random labyrinth Z is *localised* if all rw points are Z-localised, and we call it *non-localised* otherwise. Using the translation-invariance of Z and the zero–one law, we may see that Z is $\mathbb{P}$-a.s. localised if and only if

$$\mathbb{P}\big(0 \text{ is } Z\text{-localised} \,\big|\, 0 \text{ is a rw point}\big) = 1.$$

Theorem 10.11. *Let $p_{rw} > 0$. There exists a strictly positive constant $A = A(p_{rw})$ such that the following holds. The labyrinth Z is $\mathbb{P}$-a.s. non-localised if any of the following conditions hold:*

(a) $p_{rw} > p_c(\text{site})$, *the critical probability of site percolation on $\mathbb{L}^2$,*
(b) $p_+ = 0$,
(c) $p_{rw} + p_+ > 1 - A$.

We shall see in the proof of part (c) (see Theorem 10.17) that $A(p_{rw}) \to 0$ as $p_{rw} \downarrow 0$. This fact is reported in Figure 10.5, thereby correcting an error in the corresponding figure contained in [166].

Proof of Theorem 10.10. Assume $p_{rw} > 0$. We shall compare the labyrinth with a certain electrical network. By showing that the effective resistance of this network between 0 and ∞ is a.s. infinite, we shall deduce that Z is a.s. recurrent. For details of the relationship between Markov chains and electrical networks, see the book [118] and the papers [251, 277].

By the term *Z-path* we mean a path of the lattice (possibly with self-intersections) which may be followed by the light; i.e., at rw points it is unconstrained, while at reflectors and crossings it conforms to the appropriate rule. A formal definition will be presented in Section 10.3.

Let $e = \langle u, v \rangle$ be an edge of $\mathbb{L}^2$. We call e a *normal* edge if it lies in some Z-path $\pi(e)$ which is minimal with respect to the property that its two endvertices (and no others) are rw points, and furthermore that these two endvertices are distinct. If e is normal, we write $l(e)$ for the number of edges in $\pi(e)$; if e is not normal, we write $l(e) = 0$. We define $\rho(e) = 1/l(e)$, with the convention that $1/0 = \infty$.

Next, we construct an electrical network $E(\rho)$ on $\mathbb{L}^2$ by, for each edge e of $\mathbb{L}^2$, placing an electrical resistor of size $\rho(e)$ at e. Let $R = R(Z)$ be the effective resistance of this network between 0 and ∞ (which is to say that $R = \lim_{n\to\infty} R_n$, where R_n is the resistance between 0 and a composite vertex obtained by identifying all vertices in $\partial B(n)$).

Lemma 10.12. *We have that* $\mathbb{P}\big(R(Z) = \infty \,\big|\, 0 \text{ is a rw point}\big) = 1$.

Proof. We define the 'edge-boundary' $\Delta_e B(n)$ of $B(n)$ to be the set of edges $e = \langle x, y \rangle$ with $x \in \partial B(n)$ and $y \in \partial B(n+1)$. We claim that there exists a positive constant c and a random integer M such that

$$\rho(e) \geq \frac{c}{\log n} \quad \text{for all} \quad e \in \Delta_e B(n) \quad \text{and} \quad n \geq M. \tag{10.13}$$

To show this, we argue as follows. Assume that $e = \langle x, y \rangle$ is normal, and let λ_1 be the number of edges in the path $\pi(e)$ on one side of e (this side being chosen in an arbitrary way), and λ_2 for the number on the other side. Since each new vertex visited by the path is a rw point with probability p_{rw}, and since no vertex appears more than twice in $\pi(e)$, we have that

$$\mathbb{P}\big(l(e) > 2k,\ e \text{ is normal}\big) \leq 2\mathbb{P}\big(\lambda_1 \geq k,\ e \text{ is normal}\big) \leq 2(1 - p_{\mathrm{rw}})^{\frac{1}{2}(k-1)}.$$

Therefore, for $c > 0$ and all $n \geq 2$,

$$\begin{aligned}\mathbb{P}\left(\rho(e) < \frac{c}{\log n} \text{ for some } e \in \Delta_e B(n)\right) &\leq 4(2n+1)\mathbb{P}\left(l(e) > \frac{\log n}{c},\ e \text{ is normal}\right) \\ &\leq \beta n^{1-\alpha}\end{aligned}$$

where $\alpha = \alpha(c) = -(4c)^{-1}\log(1 - p_{\mathrm{rw}})$ and $\beta = \beta(c, p_{\mathrm{rw}}) < \infty$. We choose c such that $\alpha > \frac{5}{2}$, whence (10.13) follows by the Borel–Cantelli lemma.

The conclusion of the lemma is a fairly immediate consequence of (10.13), using the usual argument which follows. From the electrical network $E(\rho)$ we construct another network with no larger resistance. This we do by identifying all vertices contained in each $\partial B(n)$. In this new system there are $|\Delta_e B(n)|$ parallel connections between $\partial B(n)$ and $\partial B(n+1)$, each of which has (for $n \geq M$) a resistance at least $c/\log n$. The effective resistance from the origin to infinity is therefore at least

$$\sum_{n=M}^{\infty} \frac{c}{|\Delta_{\mathrm{e}} B(n)| \log n} = \sum_{n=M}^{\infty} \frac{c}{4(2n+1)\log n} = \infty,$$

and the proof of the lemma is complete. □

Returning to the proof of Theorem 10.10, suppose that 0 is a rw point, and consider a random walk X with $X_0 = 0$. Let C_0 be the set of rw points which may be reached by light originating at 0; C_0 is the state space of the embedded Markov chain obtained by sampling X at times when it visits rw points. This embedded chain constitutes an irreducible time-reversible Markov chain on C_0. There is a corresponding electrical network with nodes C_0, and with resistors of unit resistance joining every distinct pair u, v of such sites which are joined by some Z-path which visits no rw point other than its endpoints. This may be achieved by assigning to each corresponding edge e of $\mathbb{L}^2$ the resistance $\rho(e)$. Since the latter network may be obtained from $E(\rho)$ by deleting certain connections between paths, we have that the embedded Markov chain on C_0 is recurrent if $E(\rho)$ has infinite resistance. This latter fact was proved in Lemma 10.12. □

Proof of Theorem 10.11. Part (c) will be proved in the next section, as part of Theorem 10.17. We begin with part (a). If $p_{\mathrm{rw}} > p_{\mathrm{c}}(\text{site})$, then there exists a.s. a unique infinite cluster I of rw points having strictly positive density. Suppose $x \in I$. The walk X will (P_x^Z-a.s.) visit every vertex in I, whence the labyrinth is non-localised.

Next we prove (b), of which the proof is similar to that of Theorem 10.9. This time we construct two copies of $\mathbb{L}^2$ as follows. Let

$$A = \left\{(m+\tfrac{1}{2}, n+\tfrac{1}{2}) : m+n \text{ is even}\right\},\ B = \left\{(m+\tfrac{1}{2}, n+\tfrac{1}{2}) : m+n \text{ is odd}\right\}.$$

On $A \cup B$ we define the adjacency relation $(m+\frac{1}{2}, n+\frac{1}{2}) \sim (r+\frac{1}{2}, s+\frac{1}{2})$ if and only if $|m-r| = 1$ and $|n-s| = 1$, obtaining thereby two copies of $\mathbb{L}^2$ denoted respectively as $\mathbb{L}_A^2$ and $\mathbb{L}_B^2$. See Figure 10.6.

We now define bond percolation processes on $\mathbb{L}_A^2$ and $\mathbb{L}_B^2$. Assume $p_+ = 0$. We present the rules for $\mathbb{L}_A^2$ only; the rules for $\mathbb{L}_B^2$ are analogous. An edge of $\mathbb{L}_A^2$ joining $(m-\frac{1}{2}, n-\frac{1}{2})$ to $(m+\frac{1}{2}, n+\frac{1}{2})$ to declared to be *open* if there is a NE mirror at (m,n); similarly we declare the edge joining $(m-\frac{1}{2}, n+\frac{1}{2})$ to $(m+\frac{1}{2}, n-\frac{1}{2})$ to be *open* if there is a NW mirror at (m,n). Edges which are not open are called *closed*. This defines percolation models on $\mathbb{L}_A^2$ and $\mathbb{L}_B^2$ in which north-easterly edges (resp. north-westerly edges) are open with probability $p_{\mathrm{NE}} = \frac{1}{2}(1-p_{\mathrm{rw}})$ (resp. $p_{\mathrm{NW}} = \frac{1}{2}(1-p_{\mathrm{rw}})$). These processes are subcritical since $p_{\mathrm{NE}} + p_{\mathrm{NW}} = 1 - p_{\mathrm{rw}} < 1$. Therefore, there exists ($\mathbb{P}$-a.s.) no infinite open path in either $\mathbb{L}_A^2$ or $\mathbb{L}_B^2$, and we assume henceforth that no such infinite open path exists.

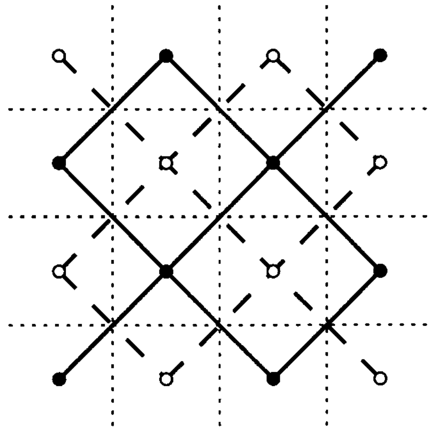


Fig. 10.6. The heavy lines are the edges of the lattice $\mathbb{L}_A^2$, and the dashed lines are the edges of the lattice $\mathbb{L}_B^2$.

Let $N(A)$ (resp. $N(B)$) be the number of open circuits in $\mathbb{L}_A^2$ (resp. $\mathbb{L}_B^2$) which contain the origin in their interiors. Since the above percolation processes are subcritical, there exists (by Theorem 6.10) a strictly positive constant $\alpha = \alpha(p_{\mathrm{NW}}, p_{\mathrm{NE}})$ such that

(10.14)

$$\mathbb{P}\Big(x \text{ lies in an open cluster of } \mathbb{L}_A^2 \text{ of diameter at least } n\Big) \le e^{-\alpha n} \quad \text{for all } n,$$

for any vertex x of $\mathbb{L}_A^2$. (By the *diameter* of a set C of vertices, we mean $\max\{|y-z| : y, z \in C\}$.) The same conclusion is valid for $\mathbb{L}_B^2$. We claim that

$$\mathbb{P}\big(0 \text{ is a rw point, and } N(A) = N(B) = 0\big) > 0, \tag{10.15}$$

and we prove this as follows. Let $\Lambda(k) = [-k, k]^2$, and let $N_k(A)$ (resp. $N_k(B)$) be the number of circuits contributing to $N(A)$ (resp. $N(B)$) which contain only points lying strictly outside $\Lambda(k)$. If $N_k(A) \ge 1$ then there exists some vertex $(m + \frac{1}{2}, \frac{1}{2})$ of $\mathbb{L}_A^2$, with $m \ge k$, which belongs to an open circuit of diameter exceeding m. Using (10.14),

$$\mathbb{P}\big(N_k(A) \ge 1\big) \le \sum_{m=k}^{\infty} e^{-\alpha m} < \tfrac{1}{3}$$

for sufficiently large k. We pick k accordingly, whence

$$\mathbb{P}\big(N_k(A) + N_k(B) \ge 1\big) \le \tfrac{2}{3}.$$

Now, if $N_k(A) = N_k(B) = 0$, and in addition all points of $\mathbb{L}^2$ inside $\Lambda(k)$ are rw points, then $N(A) = N(B) = 0$. These last events have strictly positive probabilities, and (10.15) follows.

Let J be the event that there exists a rw point $x = x(Z)$ which lies in the interior of no open circuit of either $\mathbb{L}_A^2$ or $\mathbb{L}_B^2$. Since J is invariant with respect to

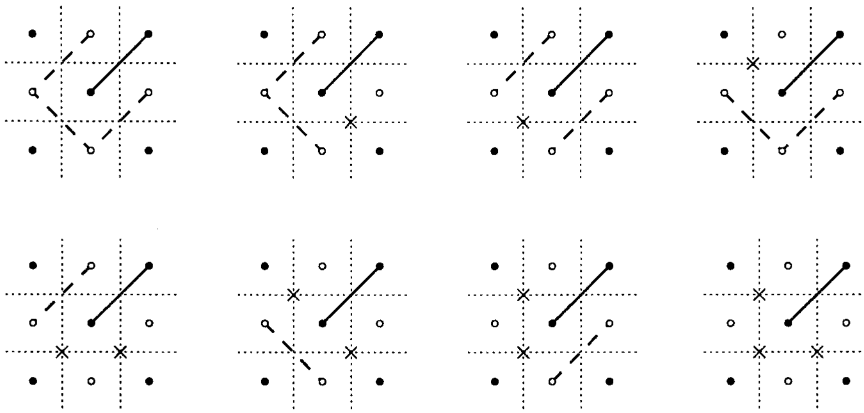


Fig. 10.7. The solid line in each picture is the edge $e = \langle u, v\rangle$, and the central vertex is u. If all three of the other edges of $\mathbb{L}_A^2$ incident with the vertex u are closed in $\mathbb{L}_A^2$, then there are eight possibilities for the corresponding edges of $\mathbb{L}_B^2$. The dashed lines indicate open edges of $\mathbb{L}_B^2$, and the crosses mark rw points of $\mathbb{L}^2$. In every picture, light incident with one side of the mirror at e will illuminate the other side also.

translations of $\mathbb{L}^2$, and since $\mathbb{P}$ is product measure, we have that $\mathbb{P}(J)$ equals either 0 or 1. Using (10.15), we deduce that $\mathbb{P}(J) = 1$. Therefore we may find a.s. some such vertex $x = x(Z)$. We claim that x is Z-non-localised, which will imply as claimed that the labyrinth if a.s. non-localised.

Let C_x be the set of rw points reachable by light originating at the rw point x. The set C_x may be generated in the following way. We allow light to leave x along the four axial directions. When a light ray hits a crossing or a mirror, it follows the associated rule; when a ray hits a rw point, it causes light to depart the point along each of the other three axial directions. Now C_x is the set of rw points thus reached. Following this physical picture, let F be the set of 'frontier mirrors', i.e., the set of mirrors only one side of which is illuminated. Assume that F is non-empty, say F contains a mirror at some point (m, n). Now this mirror must correspond to an open edge e in either $\mathbb{L}_A^2$ and $\mathbb{L}_B^2$ (see Figure 10.6 again), and we may assume without loss of generality that this open edge e is in $\mathbb{L}_A^2$. We write $e = \langle u, v\rangle$ where $u, v \in A$, and we assume that $v = u + (1, 1)$; an exactly similar argument holds otherwise. There are exactly three other edges of $\mathbb{L}_A^2$ which are incident to u (resp. v), and we claim that one of these is open. To see this, argue as follows. If none is open, then

$u + (-\frac{1}{2}, \frac{1}{2})$ either is a rw point or has a NE mirror,

$u + (-\frac{1}{2}, -\frac{1}{2})$ either is a rw point or has a NW mirror,

$u + (\frac{1}{2}, -\frac{1}{2})$ either is a rw point or has a NE mirror.

See Figure 10.7 for a diagram of the eight possible combinations. By inspection, each such combination contradicts the fact that $e = \langle u, v\rangle$ corresponds to a frontier mirror.

Therefore, u is incident to some other open edge f of $\mathbb{L}_A^2$, other than e. By a further consideration of each of $2^3 - 1$ possibilities, we may deduce that there exists such an edge f lying in F. Iterating the argument, we find that e lies in either an open circuit or an infinite open path of F lying in $\mathbb{L}_A^2$. Since there exists (by assumption) no infinite open path, this proves that f lies in an open circuit of F in $\mathbb{L}_A^2$.

By taking the union over all $e \in F$, we obtain that F is a union of open circuits of $\mathbb{L}_A^2$ and $\mathbb{L}_B^2$. Each such circuit has an interior and an exterior, and x lies (by assumption, above) in every exterior. There are various ways of deducing that x is Z-non-localised, and here is such a way.

Assume that x is Z-localised. Amongst the set of vertices $\{x + (n, 0) : n \geq 1\}$, let y be the rightmost vertex at which there lies a frontier mirror. By the above argument, y belongs to some open circuit G of F (belonging to either $\mathbb{L}_A^2$ or $\mathbb{L}_B^2$), whose exterior contains x. Since y is rightmost, we have that $y' = y + (-1, 0)$ is illuminated by light originating at x, and that light traverses the edge $\langle y', y \rangle$. Similarly, light does not traverse the edge $\langle y, y'' \rangle$, where $y'' = y + (1, 0)$. Therefore, the point $y + (\frac{1}{2}, 0)$ of $\mathbb{R}^2$ lies in the interior of G, which contradicts the fact that y is rightmost. This completes the proof for part (b). $\square$

10.3 General Labyrinths

There are many possible types of reflector, especially in three and more dimensions. Consider $\mathbb{Z}^d$ where $d \geq 2$. Let $I = \{u_1, u_2, \dots, u_d\}$ be the set of positive unit vectors, and let $I^\pm = \{-1, +1\} \times I$; members of $I^\pm$ are written as $\pm u_j$. We make the following definition. A *reflector* is a map $\rho : I^\pm \to I^\pm$ satisfying $\rho(-\rho(u)) = -u$ for all $u \in I^\pm$. We denote by $\mathcal{R}$ the set of reflectors. The 'identity reflector' is called a *crossing* (this is the identity map on $I^\pm$), and denoted by $+$.

The physical interpretation of a reflector is as follows. If light impinges on a reflector ρ, moving in a direction u ($\in I^\pm$) then it is required to depart the reflector in the direction $\rho(u)$. The condition $\rho(-\rho(u)) = -u$ arises from the reversibility of reflections.

Using elementary combinatorics, one may calculate that the number of distinct reflectors in d dimensions is

$$\sum_{s=0}^{d} \frac{(2d)!}{(2s)!\, 2^{d-s}(d-s)!}.$$

A random labyrinth is constructed as follows. Let p_{rw} and p_+ be non-negative reals satisfying $p_{\mathrm{rw}} + p_+ \leq 1$. Let $Z = (Z_x : x \in \mathbb{Z}^d)$ be independent random variables taking values in $\mathcal{R} \cup \{\varnothing\}$, with common mass function

$$\mathbb{P}(Z_0 = \alpha) = \begin{cases} p_{\mathrm{rw}} & \text{if } \alpha = \varnothing \\ p_+ & \text{if } \alpha = + \\ (1 - p_{\mathrm{rw}} - p_+)\pi(\rho) & \text{if } \alpha = \rho \in \mathcal{R}\backslash\{+\}, \end{cases}$$

where π is a prescribed probability mass function on $\mathcal{R}\backslash\{+\}$. We call a point x a *crossing* if $Z_x = +$, and a *random walk (rw) point* if $Z_x = \varnothing$.

A $\mathbb{L}^d$-path is defined to be a sequence $x_0, e_0, x_1, e_1, \dots$ of alternating vertices x_i and distinct edges e_j such that $e_j = \langle x_j, x_{j+1} \rangle$ for all j. If the path has a final vertex x_n, then it is said to have length n and to join x_0 to x_n. If it is infinite, then it is said to join x_0 to ∞. A $\mathbb{L}^d$-path may visit vertices more than once, but we insist that its edges be distinct.

We define a Z-path to be a $\mathbb{L}^d$-path $x_0, e_0, x_1, e_1, \ldots$ with the property that, for all j,

$$x_{j+1} - x_j = Z_{x_j}(x_j - x_{j-1}) \quad \text{whenever } Z_{x_j} \neq \varnothing,$$

which is to say that the path conforms to all reflectors.

Let N be the set of rw points. We define an equivalence relation $\leftrightarrow$ on N by $x \leftrightarrow y$ if and only if there exists a Z-path with endpoints x and y. We denote by C_x the equivalence class of $(N, \leftrightarrow)$ containing the rw point x, and by $\mathcal{C}$ the set of equivalence classes of $(N, \leftrightarrow)$. The following lemma will be useful; a sketch proof is deferred to the end of the section.

Lemma 10.16. *Let $d \geq 2$ and $p_{\mathrm{rw}} > 0$. The number M of equivalence classes of $(N, \leftrightarrow)$ having infinite cardinality satisfies*

$$\text{either} \quad \mathbb{P}(M = 0) = 1 \quad \text{or} \quad \mathbb{P}(M = 1) = 1.$$

Next we define a random walk in the random labyrinth Z. Let x be a rw point. A walker, starting at x, flips a fair coin (in the manner of a symmetric random walker) whenever it arrives at a rw point in order to determine its next move. When it meets a reflector, it moves according to the reflector (i.e., if it strikes the reflector ρ in the direction u, then it departs in the direction $\rho(u)$). Writing P_x^Z for the law of the walk, we say that the point x is *Z-recurrent* if $P_x^Z(X_N = x \text{ for some } N \geq 1) = 1$, and *$Z$-transient* otherwise. As before, we say that Z is *transient* if there exists a rw point x which is Z-transient, and *recurrent* otherwise.

Note that, if the random walker starts at the rw point x, then the sequence of rw points visited constitutes an irreducible Markov chain on the equivalence class C_x. Therefore, the rw point x is Z-localised if and only if $|C_x| < \infty$. As before, we say that Z is *localised* if all rw points are Z-localised, and *non-localised* otherwise.

Theorem 10.17. *Let $p_{\mathrm{rw}} > 0$. There exists a strictly positive constant $A = A(p_{\mathrm{rw}}, d)$ such that the following holds.*

(a) *Assume that $d \geq 2$. If $1 - p_{\mathrm{rw}} - p_+ < A$, then the labyrinth Z is $\mathbb{P}$-a.s. non-localised.*

(b) *Assume that $d \geq 3$. If $1 - p_{\mathrm{rw}} - p_+ < A$, then Z is $\mathbb{P}$-a.s. transient.*

As observed after 10.11, we have that $A = A(p_{\mathrm{rw}}, d) \to 0$ as $p_{\mathrm{rw}} \downarrow 0$.

Using methods presented in [115, 116, 219], one may obtain an invariance principle for a random walk in a random labyrinth, under the condition that $1 - p_{\mathrm{rw}} - p_+$ is sufficiently small. Such a principle is valid for a walk which starts in the (a.s) unique infinite equivalence class of Z. The details will appear in [74].

The following proof of Theorem 10.17 differs from that presented in [166][8]. It is slightly more complicated, but gives possibly a better numerical value for the constant A.

Proof. The idea is to relate the labyrinth to a certain percolation process, as follows. We begin with the usual lattice $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$, and from this we construct the 'line lattice' (or 'covering lattice') $\mathcal{L}$ as follows. The vertex set of $\mathcal{L}$ is the edge-set $\mathbb{E}^d$

[8]There is a small error in the proof of Theorem 7 of [166], but this may easily be corrected.

of $\mathbb{L}^d$, and two distinct vertices e_1, e_2 ($\in \mathbb{E}^d$) of $\mathcal{L}$ are called *adjacent* in $\mathcal{L}$ if and only if they have a common vertex of $\mathbb{L}^d$. If this holds, we write $\langle e_1, e_2 \rangle$ for the corresponding edge of $\mathcal{L}$, and denote by $\mathbb{F}$ the set of all such edges. We shall work with the graph $\mathcal{L} = (\mathbb{E}^d, \mathbb{F})$, and shall construct a bond percolation process on $\mathcal{L}$. We may identify $\mathbb{E}^d$ with the set of midpoints of members of $\mathbb{E}$; this embedding is useful in visualising $\mathcal{L}$.

Let $\langle e_1, e_2 \rangle \in \mathbb{E}^d$. If the edges e_1 and e_2 of $\mathbb{L}^d$ are perpendicular, we colour $\langle e_1, e_2 \rangle$ *amber*, and if they are parallel *blue*. Let $0 \le \alpha, \beta \le 1$. We declare an edge $\langle e_1, e_2 \rangle$ of $\mathbb{F}$ to be *open* with probability α (if amber) or β (if blue). This we do for each $\langle e, f \rangle \in \mathbb{F}$ independently of all other members of $\mathbb{F}$. Write $P_{\alpha,\beta}$ for the corresponding probability measure, and let $\theta(\alpha, \beta)$ be the probability that a given vertex e ($\in \mathbb{E}^d$) of $\mathcal{L}$ is in an infinite open cluster of the ensuing percolation process on $\mathcal{L}$. It is easily seen that $\theta(\alpha, \beta)$ is independent of the choice of e.

Lemma 10.18. *Let $d \ge 2$ and $0 < \alpha \le 1$. then*

$$\beta_c(\alpha, d) = \sup\{\beta : \theta(\alpha, \beta) = 0\}$$

satisfies $\beta_c(\alpha, d) < 1$.

Note that, when $d = 2$ and $\beta = 0$, the process is isomorphic to bond percolation on $\mathbb{L}^2$ with edge-parameter α. Therefore $\theta(\alpha, 0) > 0$ and $\beta_c(\alpha, 2) = 0$ when $\alpha > \frac{1}{2}$.

Proof. Since $\beta_c(\alpha, d)$ is non-increasing in d, it suffices to prove the conclusion when $d = 2$. Henceforth assume that $d = 2$.

Here is a sketch proof. Let $L \ge 1$ and let A_L be the event that every vertex of $\mathcal{L}$ lying within the box $B(L + \frac{1}{2}) = [-L - \frac{1}{2}, L + \frac{1}{2}]^2$ (of $\mathbb{R}^2$) is joined to every other vertex lying within $B(L + \frac{1}{2})$ by open paths of $\mathcal{L}$ lying inside $B(L + \frac{1}{2})$ which do not use boundary edges. For a given α satisfying $0 < \alpha < 1$, there exist L and β' such that

$$P_{\alpha,\beta'}(A_L) \ge p_c(\text{site}),$$

where $p_c(\text{site})$ is the critical probability of site percolation on $\mathbb{L}^2$. Now tile $\mathbb{Z}^2$ with copies of $B(L + \frac{1}{2})$, overlapping at the sides. It follows from the obvious relationship with site percolation that, with positive probability, the origin lies in an infinite cluster. $\square$

We now construct a labyrinth on $\mathbb{Z}^d$ from each realisation $\omega \in \{0,1\}^{\mathbb{F}}$ of the percolation process (where we write $\omega(f) = 1$ if and only if the edge f is open). That is, with each point $x \in \mathbb{Z}^d$ we shall associate a member ρ_x of $\mathcal{R} \cup \{\varnothing\}$, in such a way that ρ_x depends only on the edges $\langle e_1, e_2 \rangle$ of $\mathbb{F}$ for which e_1 and e_2 are distinct edges of $\mathbb{L}^d$ having common vertex x. It will follow that the collection $\{\rho_x : x \in \mathbb{Z}^d\}$ is a family of independent and identically distributed objects.

Since we shall define the ρ_x according to a translation-invariant rule, it will suffice to present only the definition of the reflector ρ_0 at the origin. Let $\mathbb{E}_0$ be the set of edges of $\mathbb{L}^d$ which are incident to the origin. There is a natural one–one correspondence between $\mathbb{E}_0$ and $I^{\pm}$, namely, the edge $\langle 0, u \rangle$ corresponds to the unit vector $u \in I^{\pm}$. Let $\rho \in \mathcal{R}$. Using the above correspondence, we may associate with

ρ a set of configurations in $\Omega = \{0,1\}^{\mathbb{F}}$, as follows. Let $\Omega(\rho, 0)$ be the subset of Ω containing all configurations ω satisfying

$$\omega(\langle 0, u_1\rangle, \langle 0, u_2\rangle) = 1 \quad \text{if and only if} \quad \rho(-u_1) = u_2$$

for all distinct pairs $u_1, u_2 \in I^{\pm}$.

It is not difficult to see that

$$\Omega(\rho_1, 0) \cap \Omega(\rho_2, 0) = \varnothing \quad \text{if} \quad \rho_1, \rho_2 \in \mathcal{R},\ \rho_1 \neq \rho_2.$$

Let $\omega \in \Omega$ be a percolation configuration on $\mathbb{F}$. We define the reflector $\rho_0 = \rho_0(\omega)$ at the origin by

$$\rho_0 = \begin{cases} \rho & \text{if } \omega \in \Omega(\rho, 0) \\ \varnothing & \text{if } \omega \notin \bigcup_{\rho \in \mathcal{R}} \Omega(\rho, 0). \end{cases} \tag{10.19}$$

If $\rho_0 = \rho \in \mathcal{R}$, then the behaviour of a light beam striking the origin behaves as in the corresponding percolation picture, in the following sense. Suppose light is incident in the direction u_1. There exists at most one direction u_2 ($\neq u_1$) such that $\omega(\langle 0, -u_1\rangle, \langle 0, u_2\rangle) = 1$. If such a u_2 exists, then the light is reflected in this direction. If no such u_2 exists, then it is reflected back on itself, i.e., in the direction $-u_1$.

For $\omega \in \{0,1\}^{\mathbb{F}}$, the above construction results in a random labyrinth $L(\omega)$. If the percolation process contains an infinite open cluster, then the corresponding labyrinth contains (a.s.) an infinite equivalence class.

Turning to probabilities, it is easy to see that, for $\rho \in \mathcal{R}$,

$$\pi(\rho; \alpha, \beta) = P_{\alpha,\beta}(\rho_0 = \rho)$$

satisfies $\pi(\rho; \alpha, \beta) > 0$ if $0 < \alpha, \beta < 1$, and furthermore

$$\pi(+; \alpha, \beta) = \beta^d (1-\alpha)^{\binom{2d}{2} - d}.$$

Also,

$$\pi(\varnothing; \alpha, \beta) = P_{\alpha,\beta}(\rho_0 = \varnothing) = 1 - \sum_{\rho \in \mathcal{R}} \pi(\rho; \alpha, \beta).$$

Let p_{rw}, p_+ satisfy $p_{\mathrm{rw}}, p_+ > 0$, $p_{\mathrm{rw}} + p_+ \leq 1$. We pick α, β such that $0 < \alpha, \beta < 1$, $\beta > \beta_{\mathrm{c}}(\alpha, 2)$ and

$$\pi(+; \alpha, \beta) \geq 1 - p_{\mathrm{rw}} \ (\geq p_+). \tag{10.20}$$

(That this may be done is a consequence of the fact that $\beta_{\mathrm{c}}(\alpha, 2) < 1$ for all $\alpha > 0$; cf. Lemma 10.18.)

With this choice of α, β, let

$$A = \min\left\{ \frac{\pi(\rho; \alpha, \beta)}{\pi(\rho)} : \rho \neq +,\ \rho \in \mathcal{R} \right\}$$

with the convention that $1/0 = \infty$. (Thus defined, A depends on π as well as on p_{rw}. If we set $A = \min\{\pi(\rho;\alpha,\beta) : \rho \neq +,\ \rho \in \mathcal{R}\}$, we obtain a (smaller) constant which is independent of π, and we may work with this definition instead.) Then

$$\pi(\rho;\alpha,\beta) \geq A\pi(\rho) \quad \text{for all } \rho \neq +,$$

and in particular

$$\pi(\rho;\alpha,\beta) \geq (1 - p_{\text{rw}} - p_+)\pi(\rho) \quad \text{if} \quad \rho \neq + \tag{10.21}$$

so long as p_+ satisfies $1 - p_{\text{rw}} - p_+ < A$. Note that $A = A(\alpha,\beta) > 0$.

We have from the fact that $\beta > \beta_c(\alpha, 2)$ that the percolation process ω a.s. contains an infinite open cluster. It follows that there exists a.s. a rw point in $L(\omega)$ which is $L(\omega)$-non-localised. The labyrinth Z of the theorem may be obtained (in distribution) from L as follows. Having sampled $L(\omega)$, we replace any crossing (resp. reflector ρ ($\neq +$)) by a rw point with probability $\pi(+;\alpha,\beta) - p_+$ (resp. $\pi(\rho;\alpha,\beta) - (1-p_{\text{rw}}-p_+)\pi(\rho)$); cf. (10.20) and (10.21). The ensuing labyrinth $L'(\omega)$ has the same probability distribution as Z. Furthermore, if $L(\omega)$ is non-localised, then so is $L'(\omega)$.

The first part of Theorem 10.17 has therefore been proved. Assume henceforth that $d \geq 3$, and consider part (b).

Now consider a labyrinth defined by p_{rw}, p_+, $\pi(\cdot)$. Let e be an edge of $\mathbb{Z}^d$. Either e lies in a unique path joining two rw points (but no other rw point) of some length $l(e)$, or it does not (in which case we set $l(e) = 0$). Now, the random walk in this labyrinth induces an embedded Markov chain on the set of rw points. This chain corresponds to an electrical network obtained by placing an electrical resistor at each edge e having resistance $l(e)^{-1}$. We now make two comparisons, the effect of each of which is to increase all effective resistances of the network. At the first stage, we replace all finite edge-resistances $l(e)^{-1}$ by unit resistances. This cannot decrease any effective resistance. At the next stage we replace each rw point by

$$\begin{array}{rll} + & \text{with probability} & \pi(+;\alpha,\beta) - p_+ \\ \rho\ (\neq +) & \text{with probability} & \pi(\rho;\alpha,\beta) - (1 - p_{\text{rw}} - p_+)\pi(\rho), \end{array}$$

in accordance with (10.20) and (10.21) (and where α, β are chosen so that (10.20), (10.21) hold, and furthermore $\beta_c(\alpha, 2) < \beta < 1$). Such replacements can only increase effective resistance.

In this way we obtain a comparison between the resistance of the network arising from the above labyrinth and that of the labyrinth $L(\omega)$ defined around (10.19). Indeed, it suffices to prove that the effective resistance between 0 and ∞ (in the infinite equivalence class) of the labyrinth $L(\omega)$ is a.s. finite. By examining the geometry, we claim that this resistance is no greater (up to a multiplicative constant) than the resistance between the origin and infinity of the corresponding infinite open percolation cluster of ω. By the next lemma, the last resistance is a.s. finite, whence the original walk is a.s. transient (when confined to the almost surely unique infinite equivalence class).

Lemma 10.22. *Let $d \geq 3$, $0 < \alpha < 1$, and $\beta_c(\alpha, 2) < \beta < 1$. Let R be the effective resistance between the origin and the points at infinity, in the above bond percolation process ω on $\mathcal{L}$. Then*

$$P_{\alpha,\beta}\big(R < \infty \,\big|\, 0 \text{ belongs to the infinite open cluster}\big) = 1.$$

Presumably the same conclusion is valid under the weaker hypothesis that $\beta > \beta_c(\alpha, d)$.

Sketch Proof. Rather than present all the details, here are some notes. The main techniques used in [164] arise from [165], and principally one uses the exponential decay noted in Lemma 10.4. That such decay is valid whenever $\beta > \beta_c(\alpha, d)$ uses the machinery of [165]. This machinery may be developed in the present setting (in [165] it is developed only for the hypercubic lattice $\mathbb{L}^d$). Alternatively, 'slab arguments' show (a) and (b) of Lemma 10.4 for sufficiently large β; certainly the condition $\beta > \beta_c(\alpha, 2)$ suffices for the conclusion. □

Comments on the Proof of Lemma 10.16. This resembles closely the proof of the uniqueness of the infinite percolation cluster (Theorem 7.1). We do not give the details. The notion of 'trifurcation' is replaced by that of an 'encounter zone'. Let $R \geq 1$ and $B = B(R)$. A translate $x + B$ is called an *encounter zone* if

(a) all points in $x + B$ are rw points, and

(b) in the labyrinth $\mathbb{Z}^d \setminus \{x + B\}$, there are three or more infinite equivalence classes which are part of the same equivalence class of $\mathbb{L}^d$.

Note that different encounter zones may overlap. See [74] for more details. □

11. FRACTAL PERCOLATION

11.1. Random Fractals

Many so called 'fractals' are generated by iterative schemes, of which the classical middle-third Cantor construction is a canonical example. When the scheme incorporates a randomised step, then the ensuing set may be termed a 'random fractal'. Such sets may be studied in some generality (see [131, 153, 183, 313]), and properties of fractal dimension may be established. The following simple example is directed at a 'percolative' property, namely the possible existence in the random fractal of long paths.

We begin with the unit square $C_0 = [0,1]^2$. At the first stage, we divide C_0 into nine (topologically closed) subsquares of side-length $\frac{1}{3}$ (in the natural way), and we declare each of the subsquares to be *open* with probability p (independently of any other subsquare). Write C_1 for the union of the open subsquares thus obtained. We now iterate this construction on each subsquare in C_1, obtaining a collection of open (sub)subsquares of side-length $\frac{1}{9}$. After k steps we have obtained a union C_k of open squares of side-length $(\frac{1}{3})^k$. The limit set

$$C = \lim_{k\to\infty} C_k = \bigcap_{k\geq 1} C_k \tag{11.1}$$

is a random set whose metrical properties we wish to study. See Figure 11.1.

Constructions of the above type were introduced by Mandelbrot [254] and initially studied by Chayes, Chayes, and Durrett [92]. Recent papers include [114, 134, 302]. Many generalisations of the above present themselves.

(a) Instead of working to base 3, we may work to base M where $M \geq 2$.
(b) Replace *two* dimensions by d dimensions where $d \geq 2$.
(c) Generalise the use of a square.

In what follows, (a) and (b) are generally feasible, while (c) poses a different circle of problems.

It is easily seen that the number X_k of squares present in C_k is a branching process with family-size generating function $G(x) = (1-p+px)^9$. Its extinction probability η is a root of the equation $\eta = G(\eta)$, and is such that

$$P_p(\text{extinction})\begin{cases} = 1 & \text{if } p \leq \frac{1}{9}, \\ < 1 & \text{if } p > \frac{1}{9}. \end{cases}$$

Therefore

$$P_p(C = \varnothing) = 1 \quad \text{if and only if} \quad p \leq \tfrac{1}{9}. \tag{11.2}$$

When $p > \frac{1}{9}$, then C (when non-extinct) is large but ramified.

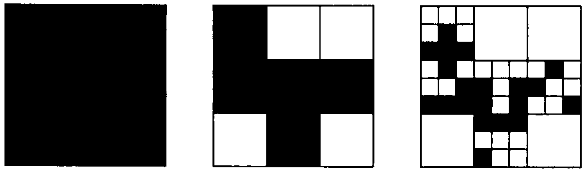


Fig. 11.1. Three stages in the construction of the 'random Cantor set' C. At each stage, a square is replaced by a 3×3 grid of smaller squares, each of which is retained with probability p.

Theorem 11.3. *Let $p > \frac{1}{9}$. The Hausdorff dimension of C, conditioned on the event $\{C \neq \varnothing\}$, equals a.s. $\log(9p)/\log 3$.*

Rather than prove this in detail, we motivate the answer. The set C is covered by X_k squares of side-length $(\frac{1}{3})^k$. Therefore the δ-dimensional box measure $H_\delta(C)$ satisfies

$$H_\delta(C) \leq X_k 3^{-k\delta}.$$

Conditional on $\{C \neq \varnothing\}$, the random variables X_k satisfy

$$\frac{\log X_k}{k} \to \log \mu \quad \text{as} \quad k \to \infty, \quad \text{a.s.}$$

where $\mu = 9p$ is the mean family-size of the branching process. Therefore

$$H_\delta(C) \leq (9p)^{k(1+o(1))} 3^{-k\delta} \quad \text{a.s.}$$

which tends to 0 as $k \to \infty$ if

$$\delta > \frac{\log(9p)}{\log 3}.$$

It follows that the box dimension of C is (a.s.) no larger than $\log(9p)/\log 3$. Experts may easily show that this bound for the dimension of C is (a.s.) exact on the event that $C \neq \varnothing$ (see [131, 183, 313]).

Indeed the exact Hausdorff measure function of C may be ascertained (see [153]), and is found to be $h(t) = t^d(\log|\log t|)^{1-\frac{1}{2}d}$ where d is the Hausdorff dimension of C.

11.2 Percolation

Can C contain long paths? More concretely, can C contain a crossing from left to right of the original unit square C_0 (which is to say that C contains a connected subset which has non-trivial intersection with the left and right sides of the unit square)? Let LR denote the event that such a crossing exists in C, and define the percolation probability

$$\theta(p) = P_p(\text{LR}). \tag{11.4}$$

In [92], it was proved that there is a non-trivial critical probability

$$p_c = \sup\{p : \theta(p) = 0\}.$$

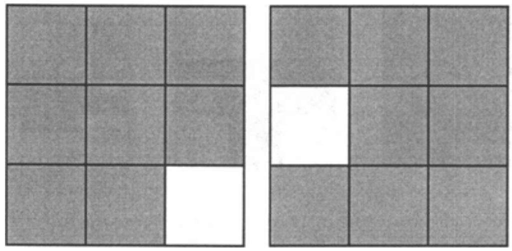


Fig. 11.2. The key fact of the construction is the following. Whenever two larger squares abut, and each has the property that at least 8 of its subsquares are retained, then their union contains a crossing from the left side to the right side.

Theorem 11.5. *We have that* $0 < p_c < 1$, *and furthermore* $\theta(p_c) > 0$.

Partial Proof. This proof is taken from [92] with help from [114]. Clearly $p_c \geq \frac{1}{9}$, and we shall prove next that

$$p_c \leq \tfrac{8}{9}\left(\tfrac{64}{63}\right)^7 \simeq 0.99248\ldots.$$

Write $\mathcal{C} = (C_0, C_1, \ldots)$. We call $\mathcal{C}$ 1-*good* if $|C_1| \geq 8$. More generally, we call $\mathcal{C}$ $(k+1)$-*good* if at least 8 of the squares in C_1 are k-good. The following fact is crucial for the argument: if $\mathcal{C}$ is k-good then C_k contains a left-right crossing of the unit square (see Figure 11.2). Therefore (using the fact that the limit of a decreasing sequence of compact connected sets is connected, and a bit more[9])

$$P_p(\mathcal{C} \text{ is } k\text{-good}) \leq P_p(C_k \text{ crosses } C_0) \downarrow \theta(p) \quad \text{as } k \to \infty, \tag{11.6}$$

whence it suffices to find a value of p for which $\pi_k = \pi_k(p) = P_p(\mathcal{C} \text{ is } k\text{-good})$ satisfies

$$\pi_k(p) \to \pi(p) > 0 \quad \text{as} \quad k \to \infty. \tag{11.7}$$

We define $\pi_0 = 1$. By an easy calculation,

$$\pi_1 = 9p^8(1-p) + p^9 = F_p(\pi_0) \tag{11.8}$$

where

$$F_p(x) = p^8 x^8 (9 - 8px). \tag{11.9}$$

More generally,

$$\pi_{k+1} = F_p(\pi_k).$$

[9]There are some topological details which are necessary for the limit in (11.6). Look at the set $\mathcal{S}_k$ of maximal connected components of C_k which intersect the left and right sides of C_0. These are closed connected sets. We call such a component a *child* of a member of $\mathcal{S}_{k-1}$ if it is a subset of that member. On the event $\{C \text{ crosses } C_0\}$, the ensuing family tree has finite vertex degrees and contains an infinite path $S_1, S_2, \ldots$ of compact connected sets. The intersection $S_\infty = \lim_{k\to\infty} S_k$ is non-empty and connected. By a similar argument, S_∞ has non-trivial intersections with the left and right sides of C_0. It follows that $\{C_k \text{ crosses } C_0\} \downarrow \{C \text{ crosses } C_0\}$, as required in (11.6). Part of this argument was suggested by Alan Stacey.

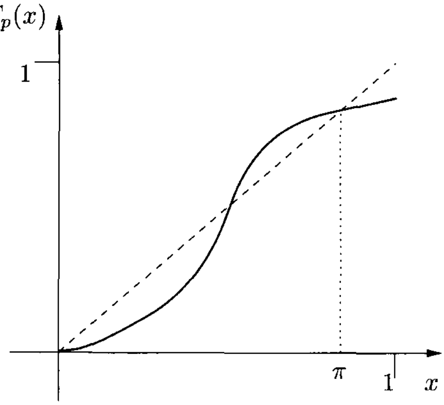


Fig. 11.3. A sketch of the function F_p for p close to 1, with the largest fixed point π marked.

About the function F_p we note that $F_p(0) = 0$, $F_p(1) < 1$, and

$$F_p'(x) = 72p^8x^7(1-px) \geq 0 \quad \text{on} \quad [0,1].$$

See Figure 11.3 for a sketch of F_p.

It follows that $\pi_k \downarrow \pi$ as $k \to \infty$ where π is the largest fixed point of F_p in the interval $[0,1]$.

It is elementary that $F_{p_0}(x_0) = x_0$, where

$$x_0 = \tfrac{9}{8}\left(\tfrac{63}{64}\right)^8, \quad p_0 = \tfrac{8}{9}\left(\tfrac{64}{63}\right)^7.$$

It follows that $\pi(p_0) \geq x_0$, yielding $\theta(p_0) > 0$. Therefore $p_c \leq p_0$, as required in (11.5).

The proof that $\theta(p_c) > 0$ is more delicate; see [114]. □

Consider now the more general setting in which the random step involves replacing a typical square of side-length M^{-k} by a $M \times M$ grid of subsquares of common side-length $M^{-(k+1)}$ (the above concerns the case $M = 3$). For general M, a version of the above argument yields that the corresponding critical probability $p_c(M)$ satisfies $p_c(M) \geq M^{-2}$ and also

$$p_c(M) < 1 \quad \text{if } M \geq 3. \tag{11.10}$$

When $M = 2$, we need a special argument in order to obtain that $p_c(2) < 1$, and this may be achieved by using the following coupling of the cases $M = 2$ and $M = 4$ (see [92, 114]). Divide C_0 into a 4×4 grid and do as follows. At the first stage, with probability p we retain all four squares in the top left corner; we do similarly for the three batches of four squares in each of the other three corners of C_0. Now for the second stage: examine each subsquare of side-length $\frac{1}{4}$ so far retained, and delete such a subsquare with probability p (different subsquares being treated independently). Note that the probability measure at the first stage dominates (stochastically) product measure with intensity π so long as $(1-\pi)^4 \geq$

$1-p$. Choose π to satisfy equality here. The composite construction outlined above dominates (stochastically) a single step of a 4×4 random fractal with parameter $p\pi = p\big(1-(1-p)^{\frac{1}{4}}\big)$, which implies that

$$p_c(2)\big(1-(1-p_c(2))^{\frac{1}{4}}\big) \le p_c(4)$$

and therefore $p_c(2) < 1$ by (11.10).

11.3 A Morphology

Random fractals have many phases, of which the existence of left-right crossings characterises only one. A weaker property than the existence of crossings is that the projection of C onto the x-axis is the whole interval $[0,1]$. Projections of random fractals are of independent interest (see, for example, the 'digital sundial' theorem of [132]). Dekking and Meester [114] have cast such properties within a more general morphology.

We write C for a random fractal in $[0,1]^2$ (such as that presented in Section 11.1). The projection of C is denoted as

$$\pi C = \{x \in \mathbb{R} : (x,y) \in C \text{ for some } y\},$$

and λ denotes Lebesgue measure. We say that C lies in one of the following phases if it has the stated property. A set is said to *percolate* if it contains a left-right crossing of $[0,1]^2$; *dimension* is denoted by 'dim'.

I. $C = \varnothing$ a.s.
II. $P(C \ne \varnothing) > 0$, $\dim(\pi C) = \dim C$ a.s.
III. $\dim(\pi C) < \dim C$ a.s. on $\{C \ne \varnothing\}$, but $\lambda(\pi C) = 0$ a.s.
IV. $0 < \lambda(\pi C) < 1$ a.s. on $\{C \ne \varnothing\}$.
V. $P\big(\lambda(\pi C) = 1\big) > 0$ but C does not percolate a.s.
VI. $P(C \text{ percolates}) > 0$.

In many cases of interest, there is a parameter p, and the ensuing fractal moves through the phases, from I to VI, as p increases from 0 to 1. There may be critical values $p_{M,N}$ at which the model moves from Phase M to Phase N. In a variety of cases, the critical values $p_{\mathrm{I,II}}$, $p_{\mathrm{II,III}}$, $p_{\mathrm{III,IV}}$ can be determined exactly, whereas $p_{\mathrm{IV,V}}$ and $p_{\mathrm{V,VI}}$ can be much harder to find.

Here is a reasonably large family of random fractals. As before, they are constructed by dividing a square into 9 equal subsquares. In this more general system, we are provided with a probability measure μ, and we replace a square by the union of a random collection of subsquares sampled according to μ. This process is iterated on all relevant scales. Certain parameters are especially relevant. Let σ_l be the number of subsquares retained from the lth column, and let $m_l = E(\sigma_l)$ be its mean.

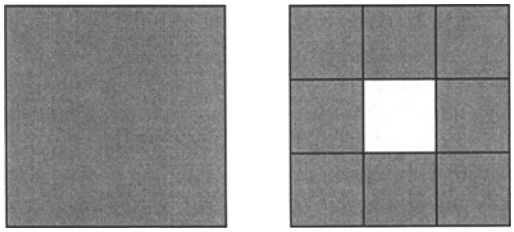


Fig. 11.4. In the construction of the Sierpinski carpet, the middle square is always deleted.

Theorem 11.11. *We have that*
(a) $C = \varnothing$ *if and only if* $\sum_{l=1}^{3} m_l \le 1$ *(unless some* σ_l *is a.s. equal to* 1*),*
(b) $\dim(\pi C) = \dim(C)$ *a.s. if and only if*

$$\sum_{l=1}^{3} m_l \log m_l \le 0,$$

(c) $\lambda(\pi C) = 0$ *a.s. if and only if*

$$\sum_{l=1}^{3} \log m_l \le 0.$$

For the proofs, see [113, 134]. Consequently, one may check the Phases I, II, III by a knowledge of the m_l only.

Next we apply Theorem 11.11 to the random Cantor set of Section 11.1, to obtain that, for this model, $p_{\mathrm{I,II}} = \frac{1}{9}$, and we depart Phase II as p increases through the value $\frac{1}{3}$. The system is never in Phase III (by Theorem 11.1(c)) or in Phase IV (by Theorem 1 of [134]). It turns out that $p_{\mathrm{II,V}} = \frac{1}{3}$ and $\frac{1}{2} < p_{\mathrm{V,VI}} < 1$.

For the 'random Sierpinski carpet' (RSC) the picture is rather different. This model is constructed as the above but with one crucial difference: at each iteration, the central square is removed with probability one, and the others with probability $1 - p$ (see Figure 11.4). Applying Theorem 11.11 we find that

$$p_{\mathrm{I,II}} = \tfrac{1}{8}, \quad p_{\mathrm{II,III}} = 54^{-\frac{1}{4}}, \quad p_{\mathrm{III,IV}} = 18^{-\frac{1}{3}},$$

and it happens that

$$\tfrac{1}{2} < p_{\mathrm{IV,V}} \le 0.8085, \quad 0.812 \le p_{\mathrm{V,VI}} \le 0.991.$$

See [114] for more details.

We close this section with a conjecture which has received some attention. Writing p_c (resp. $p_c(\mathrm{RSC})$) for the critical point of the random fractal of Section 11.1 (resp. the random Sierpinski carpet), it is evident that $p_c \le p_c(\mathrm{RSC})$. Prove or disprove the strict inequality $p_c < p_c(\mathrm{RSC})$.

11.4 Relationship to Brownian Motion

Peres [312] has discovered a link between fractal percolation and Brownian Motion, via a notion called 'intersection-equivalence'. For a region $U \subseteq \mathbb{R}^d$, we call two random sets B and C *intersection-equivalent in* U if

$$P(B \cap \Lambda \neq \varnothing) \asymp P(C \cap \Lambda \neq \varnothing) \text{ for all closed } \Lambda \subseteq U \tag{11.12}$$

(i.e., there exist positive finite constants c_1, c_2, possibly depending on U, such that

$$c_1 \leq \frac{P(B \cap \Lambda \neq \varnothing)}{P(C \cap \Lambda \neq \varnothing)} \leq c_2$$

for all closed $\Lambda \subseteq U$).

We apply this definition for two particular random sets. First, write B for the range of Brownian Motion in $\mathbb{R}^d$, starting at a point chosen uniformly at random in the unit cube. Also, for $d \geq 3$, let C be a random Cantor set constructed by binary splitting (rather than the ternary splitting of Section 11.1) and with parameter $p = 2^{2-d}$.

Theorem 11.13. *Suppose that $d \geq 3$. The random sets B and C are intersection-equivalent.*

A similar result is valid when $d = 2$, but with a suitable redefinition of the random set C. This is achieved by taking different values of p at the different stages of the construction, namely $p = k/(k+1)$ at the kth stage.

This correspondence is not only beautiful and surprising, but also useful. It provides a fairly straightforward route to certain results concerning intersections of random walks and Brownian Motions, for example. Conversely, using the rotation-invariance of Brownian Motion, one may obtain results concerning projections of the random Cantor set in other directions than onto an axis (thereby complementing results of [113], in the case of the special parameter-value given above).

The proof of Theorem 11.13 is analytical, and proceeds by utilising

- classical potential theory for Brownian Motion,
- the relationship between capacity and percolation for trees ([249]), and
- the relationship between capacity on trees and capacity on an associated Euclidean space ([55, 309]).

It is an attractive target to understand Theorem 11.13 via a coupling of the two random sets.

12. ISING AND POTTS MODELS

12.1 Ising Model for Ferromagnets

In a famous experiment, a piece of iron is exposed to a magnetic field. The field increases from zero to a maximum, and then diminishes to zero. If the temperature is sufficiently low, the iron retains some 'residual magnetisation', otherwise it does not. The critical temperature for this phenomenon is often called the Curie point. In a famous scientific paper [194], Ising proposed a mathematical model which may be phrased in the following way, using the modern idiom.

Let Λ be a box of $\mathbb{Z}^d$, say $\Lambda = [-n, n]^d$. Each vertex in Λ is allocated a random spin according to a Gibbsian probability measure as follows. Since spins come in two basic types, we take as sample space the set $\Sigma_\Lambda = \{-1, +1\}^\Lambda$, and we consider the probability measure π_Λ which allocates a probability to a spin vector $\sigma \in \Sigma_\Lambda$ given by

$$\pi_\Lambda(\sigma) = \frac{1}{Z_\Lambda} \exp\{-\beta H_\Lambda(\sigma)\}, \qquad \sigma \in \Sigma_\Lambda, \tag{12.1}$$

where $\beta = T^{-1}$ (the reciprocal of temperature, on a certain scale) and the Hamiltonian $H_\Lambda : \Sigma_\Lambda \to \mathbb{R}$ is given by

$$H_\Lambda(\sigma) = -\sum_{e=\langle i,j\rangle} J_e \sigma_i \sigma_j - h \sum_i \sigma_i \tag{12.2}$$

for constants (J_e) and h (called the 'external field') which parameterise the process. The sums in (12.2) are over all edges and vertices of Λ, respectively.

The measure (12.1) is said to arise from 'free boundary conditions', since the boundary spins have no special role. It turns out to be interesting to allow other types of boundary conditions. For any assignment $\gamma : \partial\Lambda \to \{-1, +1\}$ there is a corresponding probability measure π_Λ^γ obtained by restricting the vector σ to the set of vectors which agree with γ on $\partial\Lambda$. In this way we may obtain measures π_Λ^+, π_Λ^-, and π_Λ^{f} (with free boundary conditions) on appropriate subsets of Σ_Λ.

For simplicity, we assume here that $J_e = J > 0$ for all edges e. In this 'ferromagnetic' case, measures of the form (12.1) prefer to see configurations σ in which neighbouring vertices have like spins. The antiferromagnetic case $J < 0$ can be somewhat tricky.

Inspecting (12.1)–(12.2) with $J > 0$, we see that spins tend to align with the sign of any external field h.

The following questions are basic.

(a) What weak limits $\lim_{\Lambda\to\mathbb{Z}^d} \pi_\Lambda^\gamma$ exist for possible boundary conditions γ? (This requires redefining π_Λ^γ as a probability measure associated with the sample space $\Sigma = \{-1, +1\}^{\mathbb{Z}^d}$.)

(b) Under what conditions on J, h, d is there a unique limit measure?

(c) How may limit measures be characterised?

(d) What are their properties; for example, at what rate do their correlations decay over large distances?

(e) Is there a phase transition?

It turns out that there is a unique limit if either $d = 1$ or $h \neq 0$. There is non-uniqueness when $d \geq 2$, $h = 0$, and β is sufficiently large (i.e., $\beta > T_c^{-1}$ where T_c is the Curie point).

A great deal is known about the Ising model; see, for example, [9, 130, 135, 150, 236] and many other sources. We choose here to follow a random-cluster analysis, the details of which will follow.

The Ising model on $\mathbb{L}^2$ permits one of the famous exact calculations of statistical physics, following Onsager [300].

12.2 Potts Models

Whereas the Ising model permits two possible spin-values at each vertex, the Potts model permits a general number $q \in \{2, 3, \dots\}$. The model was introduced by Potts [318] following an earlier paper of Ashkin and Teller [46].

Let q be an integer, at least 2, and take as sample space $\Sigma_\Lambda = \{1, 2, \dots, q\}^\Lambda$ where Λ is given as before. This time we set

$$\pi_\Lambda(\sigma) = \frac{1}{Z_\Lambda} \exp\{-\beta H_\Lambda(\sigma)\}, \qquad \sigma \in \Sigma_\Lambda, \tag{12.3}$$

where

$$H_\Lambda(\sigma) = -J \sum_{e=\langle i,j \rangle} \delta_{\sigma_i, \sigma_j} \tag{12.4}$$

and $\delta_{u,v}$ is the Kronecker delta

$$\delta_{u,v} = \begin{cases} 1 & \text{if } u = v, \\ 0 & \text{otherwise.} \end{cases}$$

External field is absent from this formulation, but can be introduced if required by the addition to (12.4) of the term $-h \sum_i \delta_{\sigma_i, 1}$, which favours an arbitrarily chosen spin-value, being here the value 1.

The labelling $1, 2, \dots, q$ of the spin-values is of course arbitrary. The case $q = 2$ is identical to the Ising model (without external field and with an amended value of J), since

$$\sigma_i \sigma_j = 2\delta_{\sigma_i, \sigma_j} - 1 \qquad \text{for } \sigma_i, \sigma_j \in \{-1, +1\}.$$

12.3 Random-Cluster Models

It was Fortuin and Kasteleyn who discovered that Potts models may be recast as 'random-cluster models'. In doing so, they described a class of models, including percolation, which merits attention in their own right, and through whose analysis we discover fundamental facts concerning Ising and Potts models. See [159] for a recent account of the relevant history and bibliography.

The neatest construction of random-cluster models from Potts models is that reported in [128]. Let $G = (V, E)$ be a finite graph, and define the sample spaces

$$\Sigma = \{1, 2, \ldots, q\}^V, \quad \Omega = \{0, 1\}^E,$$

where q is a positive integer. We now define a probability mass function μ on $\Sigma \times \Omega$ by

$$\mu(\sigma, \omega) \propto \prod_{e \in E} \Big\{(1-p)\delta_{\omega(e),0} + p\delta_{\omega(e),1}\delta_e(\sigma)\Big\} \tag{12.5}$$

where $0 \le p \le 1$, and

$$\delta_e(\sigma) = \delta_{\sigma_i, \sigma_j} \qquad \text{if } e = \langle i, j \rangle \in E. \tag{12.6}$$

Elementary calculations reveal the following facts.

(a) *Marginal on* Σ. The marginal measure

$$\mu(\sigma, \cdot) = \sum_{\omega \in \Omega} \mu(\sigma, \omega)$$

is given by

$$\mu(\sigma, \cdot) \propto \exp\Big\{\beta J \sum_e \delta_e(\sigma)\Big\}$$

where $p = 1 - e^{-\beta J}$. This is the Potts measure (12.3). Note that $\beta J \ge 0$.

(b) *Marginal on* Ω. Similarly

$$\mu(\cdot, \omega) = \sum_{\sigma \in \Sigma} \mu(\sigma, \omega) \propto \Big\{\prod_e p^{\omega(e)}(1-p)^{1-\omega(e)}\Big\} q^{k(\omega)}$$

where $k(\omega)$ is the number of connected components (or 'clusters') of the graph with vertex set V and edge set $\eta(\omega) = \{e \in E : \omega(e) = 1\}$.

(c) *The conditional measures*. Given ω, the conditional measure on Σ is obtained by putting (uniformly) random spins on entire clusters of ω (of which there are $k(\omega)$), which are constant on given clusters, and independent between clusters. Given σ, the conditional measure on Ω is obtained by setting $\omega(e) = 0$ if $\delta_e(\sigma) = 0$, and otherwise $\omega(e) = 1$ with probability p (independently of other edges).

In conclusion, the measure μ is a coupling of a Potts measure $\pi_{\beta,J}$ on V, together with a 'random-cluster measure'

$$\phi_{p,q}(\omega) \propto \left\{ \prod_{e\in E} p^{\omega(e)}(1-p)^{1-\omega(e)} \right\} q^{k(\omega)}, \quad \omega \in \Omega. \tag{12.7}$$

The parameters of these measures correspond to one another by the relation $p = 1 - e^{-\beta J}$. Since $0 \le p \le 1$, this is only possible if $\beta J \ge 0$.

Why is this interesting? The 'two-point correlation function' of the Potts measure $\pi_{\beta,J}$ on $G = (V, E)$ is defined to be the function $\tau_{\beta,J}$ given by

$$\tau_{\beta,J}(i,j) = \pi_{\beta,J}(\sigma_i = \sigma_j) - \frac{1}{q}, \qquad i,j \in V.$$

The 'two-point connectivity function' of the random-cluster measure ϕ is $\phi_{p,q}(i \leftrightarrow j)$, i.e., the probability that i and j are in the same cluster of a configuration sampled according to ϕ. It turns out that these 'two-point functions' are (except for a constant factor) the same.

Theorem 12.8. *If $q \in \{2,3,\dots\}$ and $p = 1 - e^{-\beta J}$ satisfies $0 \le p \le 1$, then*

$$\tau_{\beta,J}(i,j) = (1-q^{-1})\phi_{p,q}(i \leftrightarrow j).$$

Proof. We have that

$$\begin{aligned}\tau_{\beta,J}(i,j) &= \sum_{\sigma,\omega} \left\{ 1_{\{\sigma_i=\sigma_j\}}(\sigma) - q^{-1} \right\} \mu(\sigma,\omega) \\ &= \sum_{\omega} \phi_{p,q}(\omega) \sum_{\sigma} \mu(\sigma \mid \omega) \left\{ 1_{\{\sigma_i=\sigma_j\}}(\sigma) - q^{-1} \right\} \\ &= \sum_{\omega} \phi_{p,q}(\omega) \left\{ (1-q^{-1}) 1_{\{i\leftrightarrow j\}}(\omega) + 0 \cdot 1_{\{i\nleftrightarrow j\}}(\omega) \right\} \\ &= (1-q^{-1})\phi_{p,q}(i \leftrightarrow j).\end{aligned}$$

□

This fundamental correspondence implies that properties of Potts *correlation* can be mapped to properties of random-cluster *connection*. Since Pottsian phase transition can be formulated in terms of correlation functions, this implies that information about percolative phase transition for random-cluster models is useful for studying Pottsian transitions. In doing so, we study the 'stochastic geometry' of random-cluster models.

The random-cluster measure (12.7) was constructed under the assumption that $q \in \{2,3,\dots\}$, but (12.7) makes sense for any positive real q. We have therefore obtained a rich family of measures which includes percolation ($q = 1$) as well as the Ising ($q = 2$) and Potts measures.

13. RANDOM-CLUSTER MODELS

13.1 Basic Properties

First we summarise some useful properties of random-cluster measures. Let $G = (V, E)$ be a finite graph, and write $\Omega_E = \{0,1\}^E$. The random-cluster measure on Ω_E, with parameters p, q satisfying $0 \le p \le 1$ and $q > 0$, is given by

$$\phi_{p,q}(\omega) = \frac{1}{Z}\left\{\prod_{e \in E} p^{\omega(e)}(1-p)^{1-\omega(e)}\right\} q^{k(\omega)}, \quad \omega \in \Omega_E$$

where $Z = Z_{G,p,q}$ is a normalising constant, and $k(\omega)$ is the number of connected components of the graph $(V, \eta(\omega))$, where $\eta(\omega) = \{e : \omega(e) = 1\}$ is the set of 'open' edges.

Theorem 13.1. *The measure $\phi_{p,q}$ satisfies the FKG inequality if $q \ge 1$.*

Proof. If $p = 0, 1$, the conclusion is obvious. Assume $0 < p < 1$, and check the condition (5.2), which amounts to the assertion that

$$k(\omega \vee \omega') + k(\omega \wedge \omega') \ge k(\omega) + k(\omega') \quad \text{for } \omega, \omega' \in \Omega_E.$$

This we leave as a graph-theoretic exercise. □

Theorem 13.2 (Comparison Inequalities). *We have that*

$$\phi_{p',q'} \le \phi_{p,q} \quad \text{if} \quad p' \le p,\ q' \ge q,\ q' \ge 1, \tag{13.3}$$

$$\phi_{p',q'} \ge \phi_{p,q} \quad \text{if} \quad \frac{p'}{q'(1-p')} \ge \frac{p}{q(1-p)},\ q' \ge q,\ q' \ge 1. \tag{13.4}$$

Proof. Use Holley's Inequality (Theorem 5.5) after checking condition (5.6). □

In the next theorem, the role of the graph G is emphasised in the use of the notation $\phi_{G,p,q}$. The graph $G \backslash e$ (resp. $G.e$) is obtained from G by deleting (resp. contracting) the edge e.

Theorem 13.5 (Tower Property). *Let $e \in E$.*

(a) *Given $\omega(e) = 0$, the conditional measure obtained from $\phi_{G,p,q}$ is $\phi_{G\backslash e,p,q}$.*
(b) *Given $\omega(e) = 1$, the conditional measure obtained from $\phi_{G,p,q}$ is $\phi_{G.e,p,q}$.*

Proof. This is an elementary calculation of conditional probabilities. □

More details of these facts may be found in [27, 160, 163]. Another comparison inequality may be found in [162].

13.2 Weak Limits and Phase Transitions

Let $d \geq 2$, and $\Omega = \{0,1\}^{\mathbb{E}^d}$. The appropriate σ-field of Ω is the σ-field $\mathcal{F}$ generated by the finite-dimensional sets. For $\omega \in \Omega$ and $e \in \mathbb{E}^d$, the edge e is called *open* if $\omega(e) = 1$ and *closed* otherwise.

Let Λ be a finite box in $\mathbb{Z}^d$. For $b \in \{0,1\}$ define

$$\Omega_\Lambda^b = \{\omega \in \Omega : \omega(e) = b \text{ for } e \notin \mathbb{E}_\Lambda\},$$

where $\mathbb{E}_A$ is the set of edges of $\mathbb{L}^d$ joining pairs of vertices belonging to A. On Ω_Λ^b we define a random-cluster measure $\phi_{\Lambda,p,q}^b$ as follows. Let $0 \leq p \leq 1$ and $q > 0$. Let

$$\phi_{\Lambda,p,q}^b(\omega) = \frac{1}{Z_{\Lambda,p,q}^b}\left\{\prod_{e\in\mathbb{E}_\Lambda} p^{\omega(e)}(1-p)^{1-\omega(e)}\right\}q^{k(\omega,\Lambda)} \tag{13.6}$$

where $k(\omega,\Lambda)$ is the number of clusters of $(\mathbb{Z}^d, \eta(\omega))$ which intersect Λ (here, as before, $\eta(\omega) = \{e \in \mathbb{E}^d : \omega(e) = 1\}$ is the set of open edges). The boundary condition $b = 0$ (resp. $b = 1$) is sometimes termed 'free' (resp. 'wired').

Theorem 13.7. *The weak limits*

$$\phi_{p,q}^b = \lim_{\Lambda\to\mathbb{Z}^d} \phi_{\Lambda,p,q}^b, \qquad b = 0,1,$$

exist if $q \geq 1$.

Proof. Let A be an increasing cylinder event (i.e., an increasing finite-dimensional event). If $\Lambda \subseteq \Lambda'$ and Λ includes the 'base' of A, then

$$\phi_{\Lambda,p,q}^1(A) = \phi_{\Lambda',p,q}^1(A \mid \text{all edges in } \mathbb{E}_{\Lambda'\setminus\Lambda} \text{ are open}) \geq \phi_{\Lambda',p,q}^1(A),$$

where we have used the tower property and the FKG inequality. Therefore the limit $\lim_{\Lambda\to\mathbb{Z}^d}\phi_{\Lambda,p,q}^1(A)$ exists by monotonicity. Since $\mathcal{F}$ is generated by such events A, the weak limit $\phi_{p,q}^1$ exists. A similar argument is valid in the case $b = 0$. □

The measures $\phi_{p,q}^0$ and $\phi_{p,q}^1$ are called 'random-cluster measures' on $\mathbb{L}^d$ with parameters p and q. Another route to a definition of such measures uses a type of Dobrushin–Lanford–Ruelle (DLR) formalism rather than weak limits (see [163])[10]. There is a set of 'DLR measures' ϕ satisfying $\phi_{p,q}^0 \leq \phi \leq \phi_{p,q}^1$, whence there is a unique such measure if and only if $\phi_{p,q}^0 = \phi_{p,q}^1$.

Henceforth we assume that $q \geq 1$. Turning to the question of phase transition, and remembering percolation, we define the *percolation probabilities*

$$\theta^b(p,q) = \phi_{p,q}^b(0 \leftrightarrow \infty), \quad b = 0,1, \tag{13.8}$$

i.e., the probability that 0 belongs to an infinite open cluster. The corresponding *critical* probabilities are given by

$$p_c^b(q) = \sup\{p : \theta^b(p,q) = 0\}, \quad b = 0,1.$$

Faced possibly with two (or more) distinct critical probabilities, we present the following result, abstracted from [17, 159, 160, 163].

[10] Let $\phi_{\Lambda,p,q}^\xi$ be the random-cluster measure on Λ having boundary conditions inherited from the configuration ξ off Λ. It is proved in [163] that any limit point ϕ of the family of probability measures $\{\phi_{\Lambda,p,q}^\xi : \Lambda \subseteq \mathbb{Z}^d,\ \xi \in \Omega\}$ is a DLR measure whenever ϕ has the property that the number I of infinite open clusters satisfies $\phi(I \leq 1) = 1$. It is an open problem to decide exactly which weak limits are DLR measures (if not all).

Theorem 13.9. *Assume that $d \geq 2$ and $q \geq 1$. There exists a countable subset $\mathcal{P} = \mathcal{P}_{q,d}$ of $[0,1]$, possibly empty, such that $\phi^0_{p,q} = \phi^1_{p,q}$ if either $\theta^1(p,q) = 0$ or $p \notin \mathcal{P}$.*

Consequently, $\theta^0(p,q) = \theta^1(p,q)$ if p does not belong to the countable set $\mathcal{P}_{q,d}$, whence $p^0_c(q) = p^1_c(q)$. Henceforth we refer to the critical value as $p_c(q)$. It is believed that $\mathcal{P}_{q,d} = \varnothing$ for small q (depending on the value of d), and that $\mathcal{P}_{q,d} = \{p_c(q)\}$ for large q; see the next section.

Next we prove the non-triviality of $p_c(q)$ for $q \geq 1$ (see [17]).

Theorem 13.10. *If $d \geq 2$ and $q \geq 1$ then $0 < p_c(q) < 1$.*

Proof. We compare the case of general q with the case $q = 1$ (percolation). Using the comparison inequalities (Theorem 13.2), we find that

$$p_c(1) \leq p_c(q) \leq \frac{q p_c(1)}{1 + (q-1)p_c(1)}, \qquad q \geq 1, \tag{13.11}$$

where $p_c(1)$ is the critical probability of bond percolation on $\mathbb{L}^d$. Cf. Theorem 3.2. □

We note that $p_c(q)$ is monotone non-decreasing in q, by use of the comparison inequalities. Actually it is strictly monotone and Lipschitz continuous (see [162]).

Finally we return to the Potts model, and we review the correspondence of phase transitions. The relevant 'order parameter' of the Potts model is given by

$$M(\beta J, q) = \lim_{\Lambda \to \mathbb{Z}^d} \left\{ \pi^1_{\Lambda,\beta,J}\big(\sigma(0) = 1\big) - q^{-1} \right\},$$

where $\pi^1_{\Lambda,\beta,J}$ is a Potts measure on Λ 'with boundary condition 1'. We may think of $M(\beta J, q)$ as a measure of the degree to which the boundary condition '1' is noticed at the origin. By an application of Theorem 12.8 to a suitable graph obtained from Λ, we have that

$$\pi^1_{\Lambda,\beta,J}\big(\sigma(0) = 1\big) - q^{-1} = (1 - q^{-1})\phi^1_{\Lambda,p,q}(0 \leftrightarrow \partial\Lambda)$$

where $p = 1 - e^{-\beta J}$. Therefore

$$M(\beta J, q) = (1 - q^{-1}) \lim_{\Lambda \to \mathbb{Z}^d} \phi^1_{\Lambda,p,q}(0 \leftrightarrow \partial\Lambda).$$

By an interchange of limits (which may be justified, see [17, 163]), we have that[11]

$$\lim_{\Lambda \to \mathbb{Z}^d} \phi^1_{\Lambda,p,q}(0 \leftrightarrow \partial\Lambda) = \theta^1(p,q),$$

whence $M(\beta J, q)$ and $\theta^1(p,q)$ differ only by the factor $(1 - q^{-1})$.

[11]We note that the corresponding limit for the free measure, $\lim_{\Lambda \to \mathbb{Z}^d} \phi^0_{\Lambda,p,q}(0 \leftrightarrow \partial\Lambda) = \theta^0(p,q)$, has not been proved in its full generality; see [163, 316].

13.3 First and Second Order Transitions

Let $q \geq 1$ and $0 \leq p \leq 1$. As before, $\phi^b_{p,q}$ is the random-cluster measure on $\mathbb{L}^d$ constructed according to the boundary condition $b \in \{0, 1\}$. The corresponding percolation probability is $\theta^b(p,q) = \phi^b_{p,q}(0 \leftrightarrow \infty)$. There is a phase transition at the point $p_c = p_c(q)$. Much of the interest in Potts models (and therefore random-cluster models) has been directed at a dichotomy in the type of phase transition, which depends apparently on whether q is small or large. The following picture is credible but proved only in part.

(a) *Small q, say $1 \leq q < Q(d)$.* It is believed that $\phi^0_{p,q} = \phi^1_{p,q}$ for all p, and that $\theta^b(p_c(q), q) = 0$ for $b = 0, 1$. This will imply (see [163]) that there is a unique random-cluster measure, and that each $\theta^b(\cdot, q)$ is continuous at the critical point. Such a transition is sometimes termed 'second order'. The two-point connectivity function

$$\tau^b_{p,q}(x,y) = \phi^b_{p,q}(x \leftrightarrow y) \tag{13.12}$$

satisfies

$$-\frac{1}{n} \log \tau^b_{p,q}(0, ne_1) \to \sigma(p,q) \qquad \text{as } n \to \infty \tag{13.13}$$

where $\sigma(p,q) > 0$ if and only if $p < p_c(q)$. In particular $\sigma(p_c(q), q) = 0$.

(b) *Large q, say $q > Q(d)$.* We have that $\phi^0_{p,q} = \phi^1_{p,q}$ if and only if $p \neq p_c(q)$. When $p = p_c(q)$, then $\phi^0_{p,q}$ and $\phi^1_{p,q}$ are the unique translation-invariant random-cluster measures on $\mathbb{L}^d$. Furthermore $\theta^0(p_c(q), q) = 0$ and $\theta^1(p_c(q), q) > 0$, which implies that $\theta^1(\cdot, q)$ is discontinuous at the critical point. Such a transition is sometimes termed 'first order'. The limit function σ, given by (13.13) with $b = 0$, satisfies $\sigma(p_c(q), q) > 0$, which is to say that the measure $\phi^0_{p,q}$ has exponentially decaying connectivities even *at* the critical point. Trivially $\sigma(p,q) = 0$ when $p > p_c(q)$, and this discontinuity at $p_c(q)$ is termed the 'mass gap'.

It is further believed that $Q(d)$ is non-increasing in d with

$$Q(d) = \begin{cases} 4 & \text{if } d = 2 \\ 2 & \text{if } d \geq 6. \end{cases} \tag{13.14}$$

Some progress has been made towards verifying the main features of this picture. When $d = 2$, special properties of two-dimensional space (particularly, a duality property) may be utilised (see Section 13.5). As for general values of d, we have partial information when $q = 1$, $q = 2$, or q is sufficiently large. There is no full proof of a 'sharp cut-off' in the value of q, i.e., the existence of a critical value $Q(d)$ for q (even when $d = 2$, but see [191]).

Specifically, the following is known.

(c) When $q = 1$, it is elementary that there is a unique random-cluster measure, namely product measure. Also, $\theta(p_c(1), 1) = 0$ if $d = 2$ or $d \geq 19$ (and perhaps for other d also). There is no mass gap, but $\sigma(p,q) > 0$ for $p < p_c(1)$. See [G].

(d) When $q = 2$, we have information via technology developed for the Ising model. For example, $\theta^1(p_c(2), 2) = 0$ if $d \neq 3$. Also, $\sigma(p, 2) > 0$ if $p < p_c(2)$. See [14].

(e) When q is sufficiently large, the Pirogov–Sinai theory of contours may be applied to obtain the picture described in (b) above. See [223, 225, 226, 255].

Further information about the above arguments is presented in Section 13.5 for the special case of two dimensions.

13.4 Exponential Decay in the Subcritical Phase

The key theorem for understanding the subcritical phase of percolation states that long-range connections have exponentially decaying probabilities (Theorem 6.10). Such a result is believed to hold for all random-cluster models with $q \geq 1$, but no full proof has been found. The result is known only when $q = 1$, $q = 2$, or q is sufficiently large, and the three sets of arguments for these cases are somewhat different from one another. As for results valid for all q (≥ 1), the best that is currently known is that the connectivity function decays *exponentially* whenever it decays at a sufficient *polynomial* rate. We describe this result in this section; see [168] for more details.

As a preliminary we introduce another definition of a critical point. Let

$$Y(p,q) = \limsup_{n\to\infty}\left\{n^{d-1}\phi^0_{p,q}\big(0 \leftrightarrow \partial B(n)\big)\right\} \tag{13.15}$$

and

$$p_g(q) = \sup\{p : Y(p,q) < \infty\}. \tag{13.16}$$

Evidently $p_g(q) \leq p_c(q)$, and it is believed that equality is valid here. Next, we define the kth iterate of (natural) logarithm by

$$\lambda_1(n) = \log n, \quad \lambda_k(n) = \log^+\{\lambda_{k-1}(n)\} \quad \text{for } k \geq 2$$

where $\log^+ x = \max\{1, \log x\}$.

We present the next theorem in two parts, and shall give a full proof of part (a) only; for part (b), see [168].

Theorem 13.17. *Let $0 < p < 1$ and $q \geq 1$, and assume that $p < p_g(q)$.*

(a) *If $k \geq 1$, there exists $\alpha = \alpha(p,q,k)$ satisfying $\alpha > 0$ such that*

$$\phi^0_{p,q}\big(0 \leftrightarrow \partial B(n)\big) \leq \exp\{-\alpha n/\lambda_k(n)\} \quad \textit{for all large } n. \tag{13.18}$$

(b) *If (13.18) holds, then there exists $\beta = \beta(p,q)$ satisfying $\beta > 0$ such that*

$$\phi^0_{p,q}\big(0 \leftrightarrow \partial B(n)\big) \leq e^{-\beta n} \quad \textit{for all large } n.$$

The spirit of the theorem is close to that of Hammersley [174] and Simon–Lieb [235, 332], who derived exponential estimates when $q = 1, 2$, subject to a hypothesis of finite susceptibility (i.e., that $\sum_x \phi^0_{p,q}(0 \leftrightarrow x) < \infty$). The latter hypothesis is slightly stronger than the assumption of Theorem 13.17 when $d = 2$.

Underlying any theorem of this type is an inequality. In this case we use two, of which the first is a consequence of the following version of Russo's formula, taken from [75].

Theorem 13.19. *Let $0 < p < 1$, $q > 0$, and let ψ_p be the corresponding random-cluster measure on a finite graph $G = (V, E)$. Then*

$$\frac{d}{dp}\psi_p(A) = \frac{1}{p(1-p)}\left\{\psi_p(N1_A) - \psi_p(N)\psi_p(A)\right\}$$

for any event A, where $N = N(\omega)$ is the number of open edges of a configuration ω.

Here, ψ_p is used both as probability measure and expectation operator.

Proof. We express $\psi_p(A)$ as

$$\psi_p(A) = \frac{\sum_\omega 1_A(\omega)\pi_p(\omega)}{\sum_\omega \pi_p(\omega)}$$

where $\pi_p(\omega) = p^{N(\omega)}(1-p)^{|E|-N(\omega)}q^{k(\omega)}$. Now differentiate throughout with respect to p, and gather the terms to obtain the required formula. □

Lemma 13.20. *Let $0 < p < 1$ and $q \geq 1$. For any non-empty increasing event A,*

$$\frac{d}{dp}\left\{\log\psi_p(A)\right\} \geq \frac{\psi_p(F_A)}{p(1-p)}$$

where

$$F_A(\omega) = \inf\left\{\sum_e \big(\omega'(e) - \omega(e)\big) : \omega' \geq \omega,\ \omega' \in A\right\}.$$

Proof. It may be checked that $F_A 1_A = 0$, and that $N + F_A$ is increasing. Therefore, by the FKG inequality,

$$\begin{aligned}\psi_p(N1_A) - \psi_p(N)\psi_p(A) &= \psi_p\big((N + F_A)1_A\big) - \psi_p(N)\psi_p(A)\\ &\geq \psi_p(F_A)\psi_p(A).\end{aligned}$$

Now use Theorem 13.19. □

The quantity F_A is central to the proof of Theorem 13.17. In the proof, we shall make use of the following fact. If A is increasing and $A \subseteq B_1 \cap B_2 \cap \cdots \cap B_m$, where the B_i are cylinder events defined on disjoint sets of edges, then

$$F_A \geq \sum_{i=1}^{m} F_{B_i}. \tag{13.21}$$

Lemma 13.22. *Let $q \geq 1$ and $0 < r < s < 1$. There exists a function $c = c(r, s, q)$, satisfying $1 < c < \infty$, such that*

$$\psi_r(F_A \leq k) \leq c^k \psi_s(A) \qquad \textit{for all } k \geq 0$$

and for all increasing events A.

Proof. We sketch this, which is similar to the so called 'sprinkling lemma' of [15]; see also [G, 168].

Let $r < s$. The measures ψ_r and ψ_s may be coupled together in a natural way. That is, there exists a probability measure μ on $\Omega_E^2 = \{0,1\}^E \times \{0,1\}^E$ such that:

(a) the first marginal of μ is ψ_r,

(b) the second marginal of μ is ψ_s,

(c) μ puts measure 1 on the set of configurations $(\pi, \omega) \in \Omega_E^2$ such that $\pi \leq \omega$.

Furthermore μ may be found such that the following holds. There exists a positive number $\beta = \beta(r, s, q)$ such that, for any fixed $\xi \in \Omega_E$ and subset B of edges (possibly depending on ξ), we have that

$$\frac{\mu\big(\{(\pi,\omega) : \omega(e) = 1 \text{ for } e \in B,\ \pi = \xi\}\big)}{\mu\big(\{(\pi,\omega) : \pi = \xi\}\big)} \geq \beta^{|B|}. \tag{13.23}$$

That is to say, conditional on the first component of a pair (π, ω) sampled according to μ, the measure of the second component dominates a non-trivial product measure.

Now suppose that ξ ($\in \Omega_E$) is such that $F_A(\xi) \leq k$, and find a set $B = B(\xi)$ of edges, such that $|B| \leq k$, and with the property that $\xi^B \in A$, where ξ^B is the configuration obtained from ξ by declaring all edges in B to be open. By (13.23),

$$\begin{aligned}\psi_s(A) &\geq \sum_{\xi : F_A(\xi) \leq k} \mu\big(\{(\pi,\omega) : \omega(e) = 1 \text{ for } e \in B,\ \pi = \xi\}\big)\\ &\geq \beta^k \psi_r(F_A \leq k)\end{aligned}$$

as required. □

Proof of Theorem 13.17. (a) Write $A_n = \{0 \leftrightarrow \partial B(n)\}$ and $\psi_p = \phi^0_{B(m),p,q}$ where $p < p_c = p_c(q)$. We apply Lemma 13.20 (in an integrated form), and pass to the limit as $m \to \infty$, to obtain that the measures $\phi_{p,q} = \phi^0_{p,q}$ satisfy

$$\phi_{r,q}(A_n) \leq \phi_{s,q}(A_n) \exp\big\{-4(s-r)\phi_{s,q}(F_n)\big\}, \quad \text{if } r \leq s, \tag{13.24}$$

where $F_n = F_{A_n}$ (we have used Theorem 13.2(a) here, together with the fact that F_n is a decreasing random variable).

Similarly, by summing the corresponding inequality of Lemma 13.22 over k, and letting $m \to \infty$, we find that

$$\phi_{r,q}(F_n) \geq \frac{-\log \phi_{s,q}(A_n)}{\log c} - \frac{c}{c-1} \quad \text{if } r < s. \tag{13.25}$$

We shall use (13.24) and (13.25) in an iterative scheme. At the first stage, assume $r < s < t < p_g = p_g(q)$. Find $c_1(t)$ such that

$$\phi_{p,q}(A_n) \le \frac{c_1(t)}{n^{d-1}} \qquad \text{for all } n. \tag{13.26}$$

By (13.25),

$$\phi_{s,q}(F_n) \ge \frac{(d-1)\log n}{\log c} + O(1),$$

which we insert into (13.24) to obtain

$$\phi_{r,q}(A_n) \le \frac{c_2(r)}{n^{d-1+\Delta_2(r)}} \qquad \text{for all } n, \tag{13.27}$$

and for some constants $c_2(r)$, $\Delta_2(r)$ (> 0). This is an improvement over (13.26).

At the next stages we shall need to work slightly harder. Fix a positive integer m, and let $R_i = im$ for $0 \le i \le K$ where $K = \lfloor n/m \rfloor$. Let $L_i = \{\partial B(R_i) \leftrightarrow \partial B(R_{i+1})\}$ and $H_i = F_{L_i}$. By (13.21),

$$F_n \ge \sum_{i=0}^{K-1} H_i. \tag{13.28}$$

Now there exists a constant η ($< \infty$) such that

$$\phi_{p,q}(L_i) \le |\partial B(R_i)| \phi_{p,q}(A_m) \le \eta n^{d-1} \phi_{p,q}(A_m) \tag{13.29}$$

for $0 \le i \le K-1$.

Let $r < s < p_g$, and let $c_2 = c_2(s)$, $\Delta_2 = \Delta_2(s)$ as in (13.27). From (13.28)–(13.29),

$$\begin{aligned}\phi_{s,q}(F_n) &\ge \sum_{i=0}^{K-1} \phi_{s,q}(\overline{L_i}) \ge K\big(1 - \eta n^{d-1}\phi_{p,q}(A_m)\big)\\ &\ge K\left(1 - \eta n^{d-1}\frac{c_2}{m^{d-1+\Delta_2}}\right).\end{aligned}$$

We now choose m by

$$m = \left\{(2\eta c_2)n^{d-1}\right\}^{1/(d-1+\Delta_2)}$$

(actually, an integer close to this value) to find that

$$\phi_{s,q}(F_n) \ge \tfrac{1}{2}K \ge Dn^{\Delta_3}$$

for some $D > 0$, $0 < \Delta_3 < 1$. Substitute into (13.24) to obtain

$$\phi_{r,q}(A_n) \le \exp\{-c_3 n^{\Delta_3}\} \tag{13.30}$$

for some positive $c_3 = c_3(r)$, $\Delta_3 = \Delta_3(r)$. This improves (13.27) substantially.

We repeat the last step, using (13.30) in place of (13.27), to obtain

$$\phi_{r,q}(A_n) \le \exp\left\{-\frac{c_4 n}{(\log n)^{\Delta_4}}\right\} \qquad \text{if } r < p_{\mathrm{g}} \tag{13.31}$$

for some $c_4 = c_4(r) > 0$ and $1 < \Delta_4 = \Delta_4(r) < \infty$.

At the next stage, we use (13.28)–(13.29) more carefully. This time, set $m = (\log n)^2$, and let $r < s < t < p_{\mathrm{g}}$. By (13.29) and (13.31),

$$\phi_{t,q}(L_i) \le \eta n^{d-1} \exp\left\{-\frac{c_4 m}{(\log m)^{\Delta_4}}\right\}$$

which, via Lemma 13.22, implies as in (13.25) that

$$\phi_{s,q}(H_i) \ge \frac{D(\log n)^2}{(\log\log n)^{\Delta_4}}$$

for some $D > 0$. By (13.28), and the fact that $K = \lfloor n/m \rfloor$,

$$\phi_{s,q}(F_n) \ge \frac{D' n}{(\log\log n)^{\Delta_4}}$$

for some $D' > 0$. By (13.24),

$$\phi_{r,q}(A_n) \le \exp\left\{-\frac{c_5 n}{(\log\log n)^{\Delta_4}}\right\}.$$

Since $\Delta_4 > 1$, this implies the claim of the theorem with $k = 1$. The claim for general k requires $k-1$ further iterations of the argument.

(b) We omit the proof of this part. The fundamental argument is taken from [139], and the details are presented in [168]. □

13.5 The Case of Two Dimensions

In this section we consider the case of random-cluster measures on the square lattice $\mathbb{L}^2$. Such measures have a property of self-duality which generalises that of bond percolation. We begin by describing this duality.

Let $G = (V, E)$ be a plane graph with planar dual $G^{\mathrm{d}} = (V^{\mathrm{d}}, E^{\mathrm{d}})$. Any configuration $\omega \in \Omega_E$ gives rise to a dual configuration $\omega^{\mathrm{d}} \in \Omega_{E^{\mathrm{d}}}$ defined as follows. If e ($\in E$) is crossed by the dual edge e^{d} ($\in E^{\mathrm{d}}$), we define $\omega^{\mathrm{d}}(e^{\mathrm{d}}) = 1 - \omega(e)$. As usual, $\eta(\omega)$ denotes the set $\{e : \omega(e) = 1\}$ of edges which are open in ω. By drawing a picture, one may be convinced that every face of $(V, \eta(\omega))$ contains a unique component of $(V^{\mathrm{d}}, \eta(\omega^{\mathrm{d}}))$, and therefore the number $f(\omega)$ of faces (including the infinite face) of $(V, \eta(\omega))$ satisfies $f(\omega) = k(\omega^{\mathrm{d}})$. See Figure 13.1. (Note that this definition

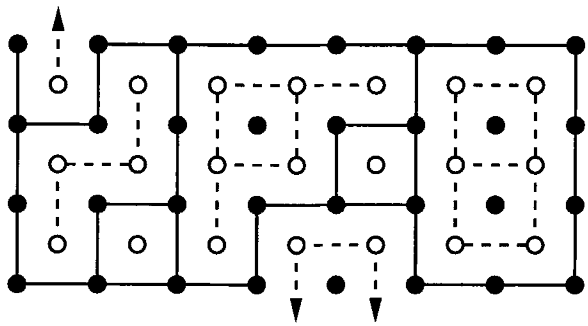


Fig. 13.1. A primal configuration ω (with solid lines) and its dual configuration ω^{d} (with dashed lines). The arrows join the given vertices of the dual to a dual vertex in the infinite face. Note that each face of the primal graph (including the infinite face) contains a unique component of the dual graph.

of the dual configuration differs slightly from that used earlier for two-dimensional percolation.)

The random-cluster measure on G is given by

$$\phi_{G,p,q}(\omega) \propto \left(\frac{p}{1-p}\right)^{|\eta(\omega)|} q^{k(\omega)}.$$

Using Euler's formula,

$$k(\omega) = |V| - |\eta(\omega)| + f(\omega) - 1,$$

and the facts that $f(\omega) = k(\omega^{\mathrm{d}})$ and $|\eta(\omega)| + |\eta(\omega^{\mathrm{d}})| = |E|$, we have that

$$\phi_{G,p,q}(\omega) \propto \left(\frac{q(1-p)}{p}\right)^{|\eta(\omega^{\mathrm{d}})|} q^{k(\omega^{\mathrm{d}})},$$

which is to say that

$$\phi_{G,p,q}(\omega) = \phi_{G^{\mathrm{d}},p^{\mathrm{d}},q}(\omega^{\mathrm{d}}) \qquad \text{for } \omega \in \Omega_E, \tag{13.32}$$

where the dual parameter p^{d} is given according to

$$\frac{p^{\mathrm{d}}}{1-p^{\mathrm{d}}} = \frac{q(1-p)}{p}. \tag{13.33}$$

The unique fixed point of the mapping $p \mapsto p^{\mathrm{d}}$ is easily seen to be given by $p = \kappa_q$ where

$$\kappa_q = \frac{\sqrt{q}}{1+\sqrt{q}}.$$

If we keep track of the constants of proportionality in the above calculation, we find that the partition function

$$Z_{G,p,q} = \sum_{\omega \in \Omega_E} p^{|\eta(\omega)|}(1-p)^{|E \setminus \eta(\omega)|} q^{k(\omega)}$$

satisfies the duality relation

$$Z_{G,p,q} = q^{|V|-1}\left(\frac{1-p}{p^{\mathrm{d}}}\right)^{|E|} Z_{G^{\mathrm{d}},p^{\mathrm{d}},q} \tag{13.34}$$

which, when $p = p^{\mathrm{d}} = \kappa_q$, becomes

$$Z_{G,\kappa_q,q} = q^{|V|-1-\frac{1}{2}|E|} Z_{G^{\mathrm{d}},\kappa_q,q}. \tag{13.35}$$

We shall find a use for this later.

Turning to the square lattice, let $\Lambda_n = [0,n]^2$, whose dual graph Λ_n^{d} may be obtained from $[-1,n]^2 + (\frac{1}{2},\frac{1}{2})$ by identifying all boundary vertices. This implies by (13.32) that

$$\phi^0_{\Lambda_n,p,q}(\omega) = \phi^1_{\Lambda_n^{\mathrm{d}},p^{\mathrm{d}},q}(\omega^{\mathrm{d}}) \tag{13.36}$$

for configurations ω on Λ_n (and with a small 'fix' on the boundary of Λ_n^{d}). Letting $n \to \infty$, we obtain that

$$\phi^0_{p,q}(A) = \phi^1_{p^{\mathrm{d}},q}(A^{\mathrm{d}}) \tag{13.37}$$

for all cylinder events A, where $A^{\mathrm{d}} = \{\omega^{\mathrm{d}} : \omega \in A\}$.

As a consequence of this duality, we may obtain as in the proof of Theorem 9.1 that

$$\theta^0(\kappa_q, q) = 0 \tag{13.38}$$

(see [163, 346]), whence the critical value of the square lattice satisfies

$$p_{\mathrm{c}}(q) \geq \frac{\sqrt{q}}{1+\sqrt{q}} \qquad \text{for } q \geq 1. \tag{13.39}$$

It is widely believed that

$$p_{\mathrm{c}}(q) = \frac{\sqrt{q}}{1+\sqrt{q}} \quad \text{for } q \geq 1.$$

This is known to hold when $q = 1$ (percolation), when $q = 2$ (Ising model), and for sufficiently large values of q. Following the route of the proof of Theorem 9.1, it suffices to show that

$$\phi^0_{p,q}\big(0 \leftrightarrow \partial B(n)\big) \leq e^{-n\psi(p,q)} \qquad \text{for all } n,$$

and for some $\psi(p,q)$ satisfying $\psi(p,q) > 0$ when $p < p_c(q)$. (Actually, rather less than exponential decay is required; it would be enough to have decay at rate n^{-1}.) This was proved by the work of [14, 235, 332] when $q = 2$. When q is large, this and more is known. Let μ be the connective constant of $\mathbb{L}^2$, and let $Q = \{\frac{1}{2}(\mu + \sqrt{\mu^2 - 4})\}^4$. We have that $2.620 < \mu < 2.696$ (see [335]), whence $21.61 < Q < 25.72$. We set

$$\psi(q) = \frac{1}{24} \log \left\{ \frac{(1+\sqrt{q})^4}{q\mu^4} \right\},$$

noting that $\psi(q) > 0$ if and only if $q > Q$.

Theorem 13.40. *If $d = 2$ and $q > Q$ then the following hold.*
 (a) *The critical point is given by $p_c(q) = \sqrt{q}/(1+\sqrt{q})$.*
 (b) *We have that $\theta^1(p_c(q), q) > 0$.*
 (c) *For any $\psi < \psi(q)$,*

$$\phi^0_{p_c(q),q}\big(0 \leftrightarrow \partial B(n)\big) \le e^{-n\psi} \qquad \textit{for all large } n.$$

We stress that these conclusions may be obtained for general d (≥ 2) when q is sufficiently large ($q > Q = Q(d)$), as may be shown using so called Pirogov–Sinai theory (see [225]). In the case $d = 2$ presented here, the above duality provides a beautiful and simple proof. This proof is an adaptation and extension of that of [226].

Proof. Let $B = B(n) = [-n,n]^2$ as usual, and let $B^{\mathrm{d}} = [-n, n-1]^2 + (\frac{1}{2}, \frac{1}{2})$ be those vertices of the dual of $B(n)$ which lie inside $B(n)$ (i.e., we omit the vertex in the infinite face of B). We shall work with 'wired' boundary conditions on B, and we let ω be a configuration on the edges of B. A circuit Γ of B^{d} is called an *outer circuit* of a configuration ω if the following properties hold:
 (a) all edges of Γ are open in the dual configuration ω^{d}, which is to say that they traverse closed edges of B,
 (b) the origin of $\mathbb{L}^2$ is in the interior of Γ,
 (c) every vertex of B lying in the exterior of Γ, but within distance of $1/\sqrt{2}$ of some vertex of Γ, belongs to the same component of ω.

See Figure 13.2 for an illustration of the meaning of 'outer circuit'.

Each circuit Γ of B^{d} partitions the set $\mathbb{E}_B$ of edges of B into three sets, being

$$\begin{aligned} E &= \{\text{edges of } B \text{ exterior to } \Gamma\}, \\ I &= \{\text{edges of } B \text{ interior to } \Gamma\}, \\ \Gamma' &= \{\text{edges of } B \text{ crossing } \Gamma\}. \end{aligned}$$

The edges I form a connected subgraph of B.

Our target is to obtain an upper bound for the probability that a given Γ is an outer circuit. This we shall do by examining certain partition functions. Since no open component of ω contains points lying in both the exterior and interior of

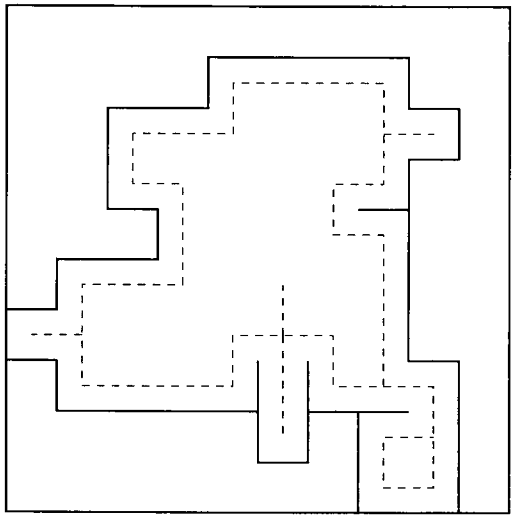


Fig. 13.2. The dashed lines include an outer circuit Γ of the dual B^{d}.

an outer circuit, the event $\mathrm{OC}(\Gamma) = \{\Gamma \text{ is an outer circuit}\}$ satisfies, for any dual circuit Γ having 0 in its interior,

$$\phi^1_{B,p,q}(\mathrm{OC}(\Gamma)) = \frac{1}{Z^1_{B,p,q}} \sum_{\omega} 1_{\mathrm{OC}(\Gamma)}(\omega)\pi_p(\omega) = \frac{1}{Z^1_{B,p,q}} (1-p)^{|\Gamma|} Z^1_E(\Gamma) Z_I \tag{13.41}$$

where $\pi_p(\omega) = p^{N(\omega)}(1-p)^{|E|-N(\omega)}q^{k(\omega)}$, $Z^1_E(\Gamma)$ is the sum of $\pi_p(\omega')$ over all $\omega' \in \{0,1\}^E$ with '1' boundary conditions on ∂B and consistent with Γ being an outer circuit (i.e., property (c) above), and Z_I is the sum of $\pi_p(\omega'')$ over all $\omega'' \in \{0,1\}^I$.

Next we use duality. Let I^{d} be the set of dual edges which cross the primal edges I, and let m be the number of vertices of B inside Γ. By (13.34),

$$Z_I = q^{m-1}\left(\frac{1-p}{p^{\mathrm{d}}}\right)^{|I|} Z^1_{I^{\mathrm{d}},p^{\mathrm{d}},q} \tag{13.42}$$

where p^{d} satisfies (13.33), and where $Z^1_{I^{\mathrm{d}},p^{\mathrm{d}},q}$ is the partition function for dual configurations, having wired boundary conditions, on the set V^{d} of vertices incident to I^{d} (i.e., all vertices of V^{d} on its boundary are identified, as indicated in Figure 13.3).

We note two general facts about partition functions. First, for any graph G, $Z_{G,p,q} \geq 1$ if $q \geq 1$. Secondly, $Z_{\cdot,p,q}$ has a property of supermultiplicativity when $q \geq 1$, which implies in particular that

$$Z^1_{B,p,q} \geq Z^1_E(\Gamma) Z^1_{I\cup\Gamma',p,q}$$

for any circuit Γ of B^{d}. (This is where we use property (c) above.)

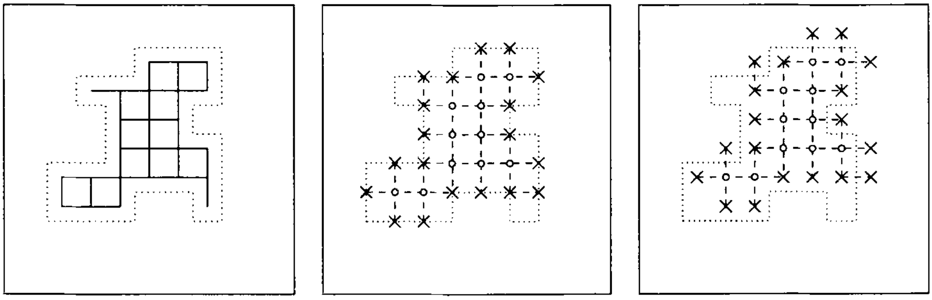


Fig. 13.3. The interior edges I of Γ are marked in the leftmost picture, and the dual I^{d} in the centre picture (the vertices marked with a cross are identified as a single vertex). The shifted set $I^* = I^{\mathrm{d}} + (\frac{1}{2}, \frac{1}{2})$ is drawn in the rightmost picture. Note that $I^* \subseteq I \cup \Gamma'$.

Let $I^* = I^{\mathrm{d}} + (\frac{1}{2}, \frac{1}{2})$, where I^{d} is thought of as a subset of $\mathbb{R}^2$. Note from Figure 13.3 that $I^* \subseteq I \cup \Gamma'$. Using the two general facts above, we have that

$$Z^1_{B,p,q} \geq Z^1_E(\Gamma) Z^1_{I^*,p,q} = Z^1_E(\Gamma) Z^1_{I^{\mathrm{d}},p,q}. \tag{13.43}$$

Now assume that $p = \sqrt{q}/(1+\sqrt{q})$, so that $p = p^{\mathrm{d}}$. Then, by (13.41)–(13.43) and (13.35),

$$\begin{aligned} \phi^1_{B,p,q}(\mathrm{OC}(\Gamma)) &= (1-p)^{|\Gamma|} \frac{Z^1_E(\Gamma) Z_I}{Z^1_{B,p,q}} \\ &= (1-p)^{|\Gamma|} q^{m-1-\frac{1}{2}|I|} \frac{Z^1_E(\Gamma) Z^1_{I^{\mathrm{d}},p,q}}{Z^1_{B,p,q}} \\ &\leq (1-p)^{|\Gamma|} q^{m-1-\frac{1}{2}|I|}. \end{aligned} \tag{13.44}$$

Since each vertex of B (inside Γ) has degree 4, we have that

$$4m = 2|I| + |\Gamma|,$$

whence

$$\phi^1_{B,p,q}(\mathrm{OC}(\Gamma)) \leq (1-p)^{|\Gamma|} q^{\frac{1}{4}|\Gamma|-1} = \frac{1}{q} \left(\frac{q}{(1+\sqrt{q})^4} \right)^{|\Gamma|/4}. \tag{13.45}$$

The number of dual circuits of B having length l and containing the origin in their interior is no greater than la_l, where a_l is the number of self-avoiding walks of $\mathbb{L}^2$ beginning at the origin and having length l. Therefore

$$\sum_\Gamma \phi^1_{B,p,q}(\mathrm{OC}(\Gamma)) \leq \sum_{l=4}^{\infty} \frac{1}{q} \left(\frac{q}{(1+\sqrt{q})^4} \right)^{l/4} la_l.$$

Now $l^{-1} \log a_l \to \mu$ as $l \to \infty$, where μ is the connective constant of $\mathbb{L}^2$. Suppose now that $q > Q$, so that $q\mu^4 < (1+\sqrt{q})^4$. It follows that there exists $A(q)$ ($< \infty$) such that

$$\sum_{\Gamma} \phi^1_{B,p,q}\big(\mathrm{OC}(\Gamma)\big) < A(q) \quad \text{for all } n.$$

If $A(q) < 1$ (which holds for sufficiently large q), then

$$\begin{aligned}\phi^1_{B,p,q}(0 \leftrightarrow \partial B) &= \phi^1_{B,p,q}\big(\mathrm{OC}(\Gamma) \text{ occurs for no } \Gamma\big)\\ &\geq 1 - A(q) > 0.\end{aligned}$$

(We have used the assumption of wired boundary conditions here.) This implies, by taking the limit $n \to \infty$, that $\theta^1(p,q) > 0$ when $p = \sqrt{q}/(1+\sqrt{q})$. Using (13.39), this implies parts (a) and (b) of the theorem, when q is sufficiently large.

For general $q > Q$, we have only that $A(q) < \infty$. In this case, we find N ($< n$) such that

$$\sum_{\Gamma \text{ outside } B(N)} \phi^1_{B,p,q}\big(\mathrm{OC}(\Gamma)\big) < \tfrac{1}{2}$$

where Γ is said to be outside $B(N)$ if it contains $B(N)$ in its interior. This implies that $\phi^1_{B,p,q}\big(B(N) \leftrightarrow \partial B\big) \geq \frac{1}{2}$. Let $n \to \infty$, and deduce that $\phi^1_{p,q}\big(B(N) \leftrightarrow \infty\big) \geq \frac{1}{2}$, implying that $\theta^1(p,q) > 0$ as required.

Turning to part (c)[12], let $p = p^{\mathrm{d}} = \sqrt{q}/(1+\sqrt{q})$ and $n \leq r$. Let A_n be the (cylinder) event that the point $(\frac{1}{2}, \frac{1}{2})$ lies in the interior of an open circuit of length at least n, this circuit having the property that its interior is contained in the interior of *no* open circuit having length strictly less than n. We have from (13.36) and (13.45) that

$$(13.46) \qquad \phi^0_{B(r),p,q}(A_n) \leq \sum_{m=n}^{\infty} \frac{m a_m}{q} \left(\frac{q}{(1+\sqrt{q})^4}\right)^{m/4} \qquad \text{for all large } r.$$

[Here, we use the observation that, if A_n occurs in $B(r)$, then there exists a maximal open circuit Γ of $B(r)$ containing $(\frac{1}{2}, \frac{1}{2})$. In the dual of $B(r)$, Γ constitutes an outer circuit.]

We write LR_n for the event that there is an open crossing of the rectangle $R_n = [0,n] \times [0,2n]$ from its left to its right side, and we set $\lambda_n = \phi^0_{p,q}(\mathrm{LR}_n)$. We may find a point x on the left side of R_n and a point y on the right side such that

$$\phi^0_{p,q}(x \leftrightarrow y \text{ in } R_n) \geq \frac{\lambda_n}{(2n+1)^2}.$$

By placing six of these rectangles side by side (as in Figure 13.4), we find by the FKG inequality that

$$(13.47) \qquad \phi^0_{p,q}\Big(x \leftrightarrow x + (6n, 0) \text{ in } [0,6n] \times [0,2n]\Big) \geq \left(\frac{\lambda_n}{(2n+1)^2}\right)^6.$$

[12]The present proof was completed following a contribution by Ken Alexander, see [37] for related material.

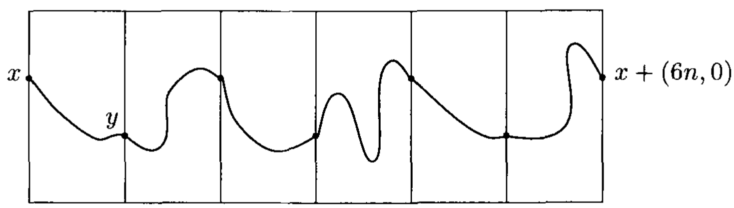


Fig. 13.4. Six copies of a rectangle having width n and height $2n$ may be put together to make a rectangle with size $6n$ by $2n$. If each is crossed by an open path joining the images of x and y, then the larger rectangle is crossed between its shorter sides.

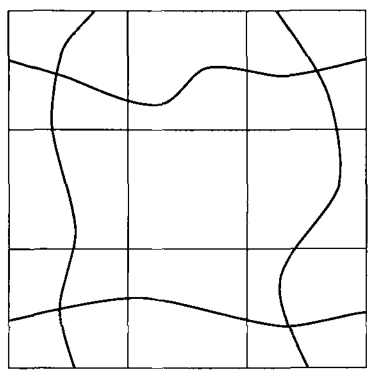


Fig. 13.5. If each of four rectangles having dimensions $6n$ by $2n$ is crossed by an open path between its shorter sides, then the annulus contains an open circuit having the origin in its interior.

We now use four copies of the rectangle $[0, 6n] \times [0, 2n]$ to construct an annulus around the origin (see Figure 13.5). If each of these copies contains an open crossing, then the annulus contains a circuit. Using the FKG inequality again, we deduce that

$$\phi^0_{p,q}(A_{4n}) \geq \left(\frac{\lambda_n}{(2n+1)^2}\right)^{24}. \tag{13.48}$$

Finally, if $0 \leftrightarrow \partial B(n)$, then one of the four rectangles $[0, n] \times [-n, n]$, $[-n, n] \times [0, n]$, $[-n, 0] \times [-n, n]$, $[-n, n] \times [-n, 0]$ is traversed by an open path betwen its two longer sides. This implies that

$$\phi^0_{p,q}\big(0 \leftrightarrow \partial B(n)\big) \leq 4\lambda_n. \tag{13.49}$$

Combining (13.46)–(13.49), we obtain that

$$\begin{aligned}\phi^0_{p,q}\big(0 \leftrightarrow \partial B(n)\big) &\leq 4\Big\{(2n+1)^2 \phi^0_{p,q}(A_{4n})\Big\}^{1/24} \\ &\leq 4\left\{(2n+1)^2 \sum_{m=4n}^{\infty} \frac{m a_m}{q}\left(\frac{q}{(1+\sqrt{q})^4}\right)^{m/4}\right\}^{1/24}.\end{aligned}$$

As before, $m^{-1} \log a_m \to \mu$ as $m \to \infty$, whence part (c) follows. $\square$

References

1. Abraham, D. B., Fontes, L., Newman, C. M., and Piza, M. S. T. (1995). Surface deconstruction and roughening in the multi-ziggurat model of wetting. *Physical Review E* **52**, R1257–R1260.
2. Abraham, D. B. and Newman, C. M. (1988). Wetting in a three dimensional system: an exact solution. *Physical Review Letters* **61**, 1969–1972.
3. Abraham, D. B. and Newman, C. M. (1989). Surfaces and Peierls contours: 3-d wetting and 2-d Ising percolation. *Communications in Mathematical Physics* **125**, 181–200.
4. Abraham, D. B. and Newman, C. M. (1990). Recent exact results on wetting. *Wetting Phenomena* (J. De Coninck and F. Dunlop, eds.), Lecture Notes in Physics, vol. 354, Springer, Berlin, pp. 13–21.
5. Abraham, D. B. and Newman, C. M. (1991). Remarks on a random surface. *Stochastic Orders and Decision under Risk* (K. Mosler and M. S. Scarsini, eds.), IMS Lecture Notes, Monograph Series, vol. 19, pp. 1–6.
6. Abraham, D. B. and Newman, C. M. (1991). The wetting transition in a random surface model. *Journal of Statistical Physics* **63**, 1097–1111.
7. Aharony, A. and Stauffer, D. (1991). *Introduction to Percolation Theory* (Second edition). Taylor and Francis.
8. Ahlfors, L. V. (1966). *Complex Analysis.* McGraw-Hill, New York.
9. Aizenman, M. (1982). Geometric analysis of ϕ^4 fields and Ising models. *Communications in Mathematical Physics* **86**, 1–48.
10. Aizenman, M. (1995). The geometry of critical percolation and conformal invariance. *Proceedings STATPHYS 1995 (Xianmen)* (Hao Bai-lin, ed.), World Scientific.
11. Aizenman, M. (1996). The critical percolation web: construction and conjectured conformal invariance properties, in preparation.
12. Aizenman, M. (1997). On the number of incipient spanning clusters. *Nuclear Physics B* **485**, 551–582.
13. Aizenman, M. and Barsky, D. J. (1987). Sharpness of the phase transition in percolation models. *Communications in Mathematical Physics* **108**, 489–526.
14. Aizenman, M., Barsky, D. J., and Fernández, R. (1987). The phase transition in a general class of Ising-type models is sharp. *Journal of Statistical Physics* **47**, 343–374.
15. Aizenman, M., Chayes, J. T., Chayes, L., Fröhlich, J, and Russo, L. (1983). On a sharp transition from area law to perimeter law in a system of random surfaces. *Communications in Mathematical Physics* **92**, 19–69.
16. Aizenman, M., Chayes, J. T., Chayes, L., and Newman, C. M. (1987). The phase boundary in dilute and random Ising and Potts ferromagnets. *Journal of Physics A: Mathematical and General* **20**, L313.
17. Aizenman, M., Chayes, J. T., Chayes, L., and Newman, C. M. (1988). Discontinuity of the magnetization in one-dimensional $1/|x-y|^2$ Ising and Potts models. *Journal of Statistical Physics* **50**, 1–40.
18. Aizenman, M. and Fernández, R. (1986). On the critical behavior of the magnetization in high-dimensional Ising models. *Journal of Statistical Physics* **44**, 393–454.

19. Aizenman, M. and Fernández, R. (1988). Critical exponents for long-range interactions. *Letters in Mathematical Physics* **16**, 39.
20. Aizenman, M. and Grimmett, G. R. (1991). Strict monotonicity for critical points in percolation and ferromagnetic models. *Journal of Statistical Physics* **63**, 817–835.
21. Aizenman, M., Kesten, H., and Newman, C. M. (1987). Uniqueness of the infinite cluster and continuity of connectivity functions for short- and long-range percolation. *Communications in Mathematical Physics* **111**, 505–532.
22. Aizenman, M., Kesten, H., and Newman, C. M. (1987). Uniqueness of the infinite cluster and related results in percolation. *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.), IMA Volumes in *Mathematics and its Applications*, vol. 8, Springer-Verlag, Berlin–Heidelberg–New York, pp. 13–20.
23. Aizenman, M., Klein, A., and Newman, C. M. (1993). Percolation methods for disordered quantum Ising models. *Phase Transitions: Mathematics, Physics, Biology ...* (R. Kotecky, ed.), World Scientific, Singapore, pp. 1–26.
24. Aizenman, M. and Lebowitz, J. (1988). Metastability in bootstrap percolation. *Journal of Physics A: Mathematical and General* **21**, 3801–3813.
25. Aizenman, M., Lebowitz, J., and Bricmont, J. (1987). Percolation of the minority spins in high-dimensional Ising models. *Journal of Statistical Physics* **49**, 859–865.
26. Aizenman, M. and Nachtergaele, B. (1994). Geometric aspects of quantum spin states. *Communications in Mathematical Physics* **164**, 17–63.
27. Aizenman, M. and Newman, C. M. (1984). Tree graph inequalities and critical behavior in percolation models. *Journal of Statistical Physics* **36**, 107–143.
28. Aldous, D. and Steele, J. M. (1992). Asymptotics for Euclidean minimal spanning trees on random points. *Probability Theory and Related Fields* **92**, 247–258.
29. Alexander, K. S. (1990). Lower bounds on the connectivity function in all directions for Bernoulli percolation in two and three dimensions. *Annals of Probability* **18**, 1547–1562.
30. Alexander, K. S. (1992). Stability of the Wulff construction and fluctuations in shape for large finite clusters in two-dimensional percolation. *Probability Theory and Related Fields* **91**, 507–532.
31. Alexander, K. S. (1993). Finite clusters in high-density continuous percolation: compression and sphericality. *Probability Theory and Related Fields* **97**, 35–63.
32. Alexander, K. S. (1993). A note on some rates of convergence in first-passage percolation. *Annals of Applied Probability* **3**, 81–90.
33. Alexander, K. S. (1995). Simultaneous uniqueness of infinite graphs in stationary random labeled graphs. *Communications in Mathematical Physics* **168**, 39–55.
34. Alexander, K. S. (1995). Percolation and minimal spanning forests in infinite graphs. *Annals of Probability* **23**, 87–104.
35. Alexander, K. S. (1996). Approximation of subadditive functions and convergence rates in limiting-shape results. *Annals of Probability* (to appear).

36. Alexander, K. S. (1996). The RSW theorem for continuum percolation and the CLT for Euclidean minimal spanning trees. *Annals of Applied Probability* **6**, 466–494.
37. Alexander, K. S. (1996). On weak mixing in lattice models (to appear).
38. Alexander, K. S., Chayes, J. T., and Chayes, L. (1990). The Wulff construction and asymptotics of the finite cluster distribution for two-dimensional Bernoulli percolation. *Communications in Mathematical Physics* **131**, 1–50.
39. Alexander, K. S. and Molchanov, S. A. (1994). Percolation of level sets for two-dimensional random fields with lattice symmetry. *Journal of Statistical Physics* **77**, 627–643.
40. Andjel, E. (1993). Characteristic exponents for two-dimensional bootstrap percolation. *Annals of Probability* **21**, 926–935.
41. Andjel, E. and Schinazi, R. (1996). A complete convergence theorem for an epidemic model. *Journal of Applied Probability* **33**, 741–748.
42. Antal, P. and Pisztora, A. (1996). On the chemical distance for supercritical Bernoulli percolation. *Annals of Probability* **24**, 1036–1048.
43. Appel, M. J. and Wierman, J. C. (1987). On the absence of AB percolation in bipartite graphs. *Journal of Physics A: Mathematical and General* **20**, 2527–2531.
44. Appel, M. J. and Wierman, J. C. (1987). Infinite AB percolation clusters exist on the triangular lattice. *Journal of Physics A: Mathematical and General* **20**, 2533–2537.
45. Appel, M. J. and Wierman, J. C. (1993). AB percolation on bond decorated graphs. *Journal of Applied Probability* **30**, 153–166.
46. Ashkin, J. and Teller, E. (1943). Statistics of two-dimensional lattices with four components. *The Physical Review* **64**, 178–184.
47. Barlow, M. T., Pemantle, R., and Perkins, E. (1997). Diffusion-limited aggregation on a homogenous tree. *Probability Theory and Related Fields* **107**, 1–60.
48. Barsky, D. J. and Aizenman, M. (1991). Percolation critical exponents under the triangle condition. *Annals of Probability* **19**, 1520–1536.
49. Barsky, D. J., Grimmett, G. R., and Newman, C. M. (1991). Dynamic renormalization and continuity of the percolation transition in orthants. *Spatial Stochastic Processes* (K. S. Alexander and J. C. Watkins, eds.), Birkhäuser, Boston, pp. 37–55.
50. Barsky, D. J., Grimmett, G. R., and Newman, C. M. (1991). Percolation in half spaces: equality of critical probabilities and continuity of the percolation probability. *Probability Theory and Related Fields* **90**, 111–148.
51. Benjamini, I. (1996). Percolation on groups (to appear).
52. Benjamini, I. and Kesten, H. (1995). Percolation of arbitrary words in $\{0,1\}^{\mathbb{N}}$. *Annals of Probability* **23**, 1024–1060.
53. Benjamini, I., Pemantle, R., and Peres, Y. (1995). Martin capacity for Markov chains. *Annals of Probability* **23**, 1332–1346.
54. Benjamini, I., Pemantle, R., and Peres, Y. (1997). Unpredictable paths and percolation (to appear).

55. Benjamini, I. and Peres, Y. (1992). Random walks on a tree and capacity in the interval. *Annales de l'Institut Henri Poincaré* **28**, 557–592.
56. Benjamini, I. and Peres, Y. (1994). Tree-indexed random walks on groups and first passage percolation. *Probability Theory and Related Fields* **98**, 91–112.
57. Benjamini, I. and Schramm, O. (1996). Conformal invariance of Voronoi percolation (to appear).
58. Berg, J. van den (1985). Disjoint occurrences of events: results and conjectures. *Particle Systems, Random Media and Large Deviations* (R. T. Durrett, ed.), Contemporary Mathematics no. 41, American Mathematical Society, Providence, R. I., pp. 357–361.
59. Berg, J. van den (1993). A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising anti-ferromagnet. *Communications in Mathematical Physics* **152**, 161–166.
60. Berg, J. van den (1996). A constructive mixing condition for 2-D Gibbs measures with random interactions (to appear).
61. Berg, J. van den (1996). A note on disjoint-occurrence inequalities for marked Poisson point processes. *Journal of Applied Probability* **33**, 420–426.
62. Berg, J. van den (1997). Some reflections on disjoint occurrences of events (to appear).
63. Berg, J. van den and Ermakov, A. (1996). A new lower bound for the critical probability of site percolation on the square lattice. *Random Structures and Algorithms*, 199–212.
64. Berg, J. van den and Fiebig, U. (1987). On a combinatorial conjecture concerning disjoint occurrences of events. *Annals of Probability* **15**, 354–374.
65. Berg, J. van den, Grimmett, G. R., and Schinazi, R. B. (1997). Dependent random graphs and spatial epidemics (to appear).
66. Berg, J. van den and Keane, M. (1985). On the continuity of the percolation probability function. *Particle Systems, Random Media and Large Deviations* (R. T. Durrett, ed.), Contemporary Mathematics Series, vol. 26, AMS, Providence, R. I., pp. 61–65.
67. Berg, J. van den and Kesten, H. (1985). Inequalities with applications to percolation and reliability. *Journal of Applied Probability* **22**, 556–569.
68. Berg, J. van den and Kesten, H. (1993). Inequalities for the time constant in first-passage percolation. *Advances in Applied Probability* **3**, 56–80.
69. Berg, J. van den and Maes, C. (1994). Disagreement percolation in the study of Markov fields. *Annals of Probability* **22**, 749–763.
70. Berg, J. van den and Meester, R. (1991). Stability properties of a flow process in graphs. *Random Structures and Algorithms* **2**, 335–341.
71. Berg, J. van den and Steif, J. (1994). Percolation and the hard-core lattice gas model. *Stochastic Processes and their Applications* **49**, 179–197.
72. Bezuidenhout, C. E. and Grimmett, G. R. (1990). The critical contact process dies out. *Annals of Probability* **18**, 1462–1482.
73. Bezuidenhout, C. E. and Grimmett, G. R. (1991). Exponential decay for subcritical contact and percolation processes. *Annals of Probability* **19**, 984–1009.

74. Bezuidenhout, C. E. and Grimmett, G. R. (1996). A central limit theorem for random walks in random labyrinths (to appear).
75. Bezuidenhout, C. E., Grimmett, G. R., and Kesten, H. (1993). Strict inequality for critical values of Potts models and random-cluster processes. *Communications in Mathematical Physics* **158**, 1–16.
76. Bezuidenhout, C. E., Grimmett, G. R., and Löffler, A. (1996). Minimal spanning tree analysis of site percolation clusters (to appear).
77. Bollobás, B., Grimmett, G. R., and Janson, S. (1996). The random-cluster process on the complete graph. *Probability Theory and Related Fields* **104**, 283–317.
78. Borgs, C. and Chayes, J. T. (1996). The covariance matrix of the Potts model: a random cluster analysis. *Journal of Statistical Physics* **82**, 1235–1297.
79. Borgs, C., Chayes, J. T., Kesten, H., and Spencer, J. (1996). The birth of the infinite cluster (to appear).
80. Bramson, M., Durrett, R. T., and Schonmann, R. H. (1991). The contact process in a random environment. *Annals of Probability* **19**, 960–983.
81. Broadbent, S. R. and Hammersley, J. M. (1957). Percolation processes I. Crystals and mazes. *Proceedings of the Cambridge Philosophical Society* **53**, 629–641.
82. Bunimovitch, L. A. and Troubetzkoy, S. E. (1992). Recurrence properties of Lorentz lattice gas cellular automata. *Journal of Statistical Physics* **67**, 289–302.
83. Burton, R. M. and Keane, M. (1989). Density and uniqueness in percolation. *Communications in Mathematical Physics* **121**, 501–505.
84. Burton, R. M. and Meester, R. (1993). Long range percolation in stationary point processes. *Random Structures and Algorithms* **4**, 177–190.
85. Burton, R. M. and Pemantle, R. (1993). Local characteristics, entropy and limit theorems for uniform spanning trees and domino tilings via transfer-impedances. *Annals of Probability* **21**, 1329–1371.
86. Campanino, M., Chayes, J. T., and Chayes, L. (1991). Gaussian fluctuations of connectivities in the subcritical regime of percolation. *Probability Theory and Related Fields* **88**, 269–341.
87. Cardy, J. (1992). Critical percolation in finite geometries. *Journal of Physics A: Mathematical and General* **25**, L201.
88. Chayes, J. T. and Chayes, L. (1987). The mean field bound for the order parameter of Bernoulli percolation,. *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.), Springer-Verlag, New York, pp. 49–71.
89. Chayes, J. T. and Chayes, L. (1987). On the upper critical dimension of Bernoulli percolation,. *Communications in Mathematical Physics* **113**, 27–48.
90. Chayes, J. T. and Chayes, L. (1989). The large-N limit of the threshold values in Mandelbrot's fractal percolation process. *Journal of Physics A: Mathematical and General* **22**, L501–L506.
91. Chayes, J. T., Chayes, L., and Durrett, R. T. (1986). Critical behavior of the two-dimensional first passage time. *Journal of Statistical Physics* **45**, 933–948.

92. Chayes, J. T., Chayes, L., and Durrett, R. T. (1988). Connectivity properties of Mandelbrot's percolation process. *Probability Theory and Related Fields* **77**, 307–324.
93. Chayes, J. T., Chayes, L., Fisher, D. S., and Spencer, T. (1989). Correlation length bounds for disordered Ising ferromagnets. *Communications in Mathematical Physics* **120**, 501–523.
94. Chayes, J. T., Chayes, L., Grannan, E., and Swindle, G. (1991). Phase transitions in Mandelbrot's percolation process in 3 dimensions. *Probability Theory and Related Fields* **90**, 291–300.
95. Chayes, J. T., Chayes, L., Grimmett, G. R., Kesten, H., and Schonmann, R. H. (1989). The correlation length for the high density phase of Bernoulli percolation. *Annals of Probability* **17**, 1277–1302.
96. Chayes, J. T., Chayes, L., and Kotecky, R. (1995). The analysis of the Widom–Rowlinson model by stochastic geometric methods. *Communications in Mathematical Physics* **172**, 551–569.
97. Chayes, J. T., Chayes, L., and Newman, C M. (1987). Bernoulli percolation above threshold: an invasion percolation analysis. *Annals of Probability* **15**, 1272–1287.
98. Chayes, J. T., Chayes, L., and Schonmann, R. H. (1987). Exponential decay of connectivities in the two dimensional Ising model. *Journal of Statistical Physics* **49**, 433–445.
99. Chayes, L. (1991). On the critical behavior of the first passage time in $d = 3$. *Helvetica Physica Acta* **64**, 1055–1069.
100. Chayes, L. (1993). The density of Peierls contours in $d = 2$ and the height of the wedding cake. *Journal of Physics A: Mathematical and General* **26**, L481–L488.
101. Chayes, L. (1995). On the absence of directed fractal percolation. *Journal of Physics A: Mathematical and General* **28**, L295–L301.
102. Chayes, L. (1995). Aspects of the fractal percolation process. *Fractal Geometry and Stochastics* (C. Bandt, S. Graf and M. Zähle, eds.), Birkhäuser, Boston, pp. 113–143.
103. Chayes, L. (1996). On the length of the shortest crossing in the super-critical phase of Mandelbrot's percolation process. *Stochastic Processes and their Applications* **61**, 25–43.
104. Chayes, L. (1996). Percolation and ferromagnetism on $\mathbb{Z}^2$: the q-state Potts cases (to appear).
105. Chayes, L., Kotecky, R., and Shlosman, S. B. (1995). Aggregation and intermediate phases in dilute spin-systems. *Communications in Mathematical Physics* **171**, 203–232.
106. Chayes, L. and Winfield, C. (1993). The density of interfaces: a new first passage problem. *Journal of Applied Probability* **30**, 851–862.
107. Chow, Y. S. and Teicher, H. (1978). *Probability Theory.* Springer, Berlin.
108. Clifford, P. (1990). Markov random fields in statistics. *Disorder in Physical Systems* (G. R. Grimmett and D. J. A. Welsh, eds.), Oxford University Press, Oxford, pp. 19–32.
109. Cohen, E. G. D. (1991). New types of diffusions in lattice gas cellular automata. *Microscopic Simulations of Complex Hydrodynamical Phenomena*

(M. Mareschal and B. L. Holian, eds.), Plenum Press, New York, pp. 137–152.

110. Cohen, E. G. D. and Wang, F. (1995). New results for diffusion in Lorentz lattice gas cellular automata. *Journal of Statistical Physics* **81**, 445–466.
111. Cohen, E. G. D. and Wang, F. (1995). Novel phenomena in Lorentz lattice gases. *Physica A* **219**, 56–87.
112. Cox, J. T., Gandolfi, A., Griffin, P., and Kesten, H. (1993). Greedy lattice animals I: Upper bounds. *Advances in Applied Probability* **3**, 1151–1169.
113. Dekking, F. M. and Grimmett, G. R. (1988). Superbranching processes and projections of random Cantor sets. *Probability Theory and Related Fields* **78**, 335–355.
114. Dekking, F. M. and Meester, R. W. J. (1990). On the structure of Mandelbrot's percolation process and other random Cantor sets. *Journal of Statistical Physics* **58**, 1109–1126.
115. DeMasi, A., Ferrari, P. A., Goldstein, S., and Wick, W. D. (1985). Invariance principle for reversible Markov processes with application to diffusion in the percolation regime. *Particle Systems, Random Media and Large Deviations* (R. T. Durrett, ed.), Contemporary Mathematics no. 41, American Mathematical Society, Providence, R. I., pp. 71–85.
116. DeMasi, A., Ferrari, P. A., Goldstein, S., and Wick, W. D. (1989). An invariance principle for reversible Markov processes. Applications to random motions in random environments. *Journal of Statistical Physics* **55**, 787–855.
117. Deuschel, J.-D. and Pisztora, A. (1996). Surface order large deviations for high-density percolation. *Probability Theory and Related Fields* **104**, 467–482.
118. Doyle, P. G. and Snell, E. L. (1984). *Random Walks and Electric Networks.* Carus Mathematical Monograph no. 22. AMA, Washington, D. C.
119. Durrett, R. T. (1988). *Lecture Notes on Particle Systems and Percolation.* Wadsworth and Brooks/Cole, Pacific Grove, California.
120. Durrett, R. T. (1991). The contact process, 1974–1989. *Mathematics of Random Media* Blacksburg, VA, Lectures in Applied Mathematics, vol. 27, AMS, pp. 1–18.
121. Durrett, R. T. (1995). Ten lectures on particle systems. *Ecole d'Eté de Probabilités de Saint Flour* XXIII-1993 (P. Bernard, ed.), Lecture Notes in Mathematics no. 1608, Springer, Berlin, pp. 97–201.
122. Durrett, R. T. and Kesten, H. (1990). The critical parameter for connectedness of some random graphs. *A Tribute to Paul Erdős* (A. Baker, B. Bollobás, and A. Hajnal, eds.), Cambridge University Press, Cambridge, pp. 161–176.
123. Durrett, R. T. and Neuhauser, C. (1991). Epidemics with recovery. *Advances in Applied Probability* **1**, 189–206.
124. Durrett, R. T. and Schonmann, R. H. (1988). The contact process on a finite set II. *Annals of Probability* **16**, 1570–1583.
125. Durrett, R. T. and Schonmann, R. H. (1988). Large deviations for the contact process and two dimensional percolation. *Probability Theory and Related Fields* **77**, 583–603.

126. Durrett, R. T., Schonmann, R. H., and Tanaka, N. I. (1989). The contact process on a finite set III. The critical case. *Annals of Probability* **17**, 1303–1321.
127. Durrett, R. T., Schonmann, R. H., and Tanaka, N. I. (1989). Correlation lengths for oriented percolation. *Journal of Statistical Physics* **55**, 965–979.
128. Edwards, R. G. and Sokal, A. D. (1988). Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *The Physical Review* D **38**, 2009–2012.
129. Ehrenfest, P. (1959). *Collected Scientific Papers* (M. J. Klein, ed.). North-Holland, Amsterdam.
130. Enter, A. C. D. van and Hollander, F. den (1993). Interacting particle systems and Gibbs measures, Lecture notes for Master Class 1992–1993, Mathematical Research Institute, The Netherlands.
131. Falconer, K. J. (1986). Random fractals. *Mathematical Proceedings of the Cambridge Philosophical Society* **100**, 559–582.
132. Falconer, K. J. (1986). Sets with prescribed projections and Nikodym sets. *Proceedings of the London Mathematical Society* **53**, 48–64.
133. Falconer, K. J. (1989). Projections of random Cantor sets. *Journal of Theoretical Probability* **2**, 65–70.
134. Falconer, K. J. and Grimmett, G. R. (1992). On the geometry of random Cantor sets and fractal percolation. *Journal of Theoretical Probability* **5**, 465–485; Erratum (1994) **7**, 209–210.
135. Fernández, R., Fröhlich, J., and Sokal, A. D. (1992). *Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory.* Springer-Verlag, Berlin.
136. Fill, J. and Pemantle, R. (1992). Oriented percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *Annals of Applied Probability* **3**, 593–629.
137. Fiocco, M. and Zwet, W. R. van (1996). Consistent estimation for the supercritical contact process (to appear).
138. Fisher, M. E. (1961). Critical probabilities for cluster size and percolation problems. *Journal of Mathematical Physics* **2**, 620–627.
139. Fontes, L. and Newman, C. M. (1993). First passage percolation for random colorings of $\mathbb{Z}^d$. *Annals of Applied Probability* **3**, 746–762; Erratum **4**, 254.
140. Fortuin, C. M. (1971). *On the random-cluster model.* Doctoral thesis, University of Leiden.
141. Fortuin, C. M. (1972). On the random cluster model. II. The percolation model. *Physica* **58**, 393–418.
142. Fortuin, C. M. (1972). On the random cluster model. III. The simple random-cluster process. *Physica* **59**, 545–570.
143. Fortuin, C. M. and Kasteleyn, P. W. (1972). On the random cluster model. I. Introduction and relation to other models. *Physica* **57**, 536–564.
144. Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics* **22**, 89–103.

145. Friedgut, E. and Kalai, G. (1996). Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society* **124**, 2993–3002.
146. Gandolfi, A., Grimmett, G. R., and Russo, L. (1988). On the uniqueness of the infinite open cluster in the percolation model. *Communications in Mathematical Physics* **114**, 549–552.
147. Gandolfi, A., Keane, M., and Newman, C. M. (1992). Uniqueness of the infinite component in a random graph with applications to percolation and spin glasses. *Probability Theory and Related Fields* **92**, 511–527.
148. Gandolfi, A. and Kesten, H. (1994). Greedy lattice animals II: Linear growth. *Advances in Applied Probability* **4**, 76–107.
149. Gandolfi, A., Newman, C. M., and Stein, D. L. (1993). Exotic states in long range spin glasses. *Communications in Mathematical Physics* **157**, 371–387.
150. Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions.* Walter de Gruyter, Berlin.
151. Gordon, D. M. (1991). Percolation in high dimensions. *Bulletin of the London Mathematical Society* **44**, 373–384.
152. Graf, S. (1987). Statistically self-similar fractals. *Probability Theory and Related Fields* **74**, 357–392.
153. Graf, S., Mauldin, R. D., and Williams, S. C. (1987). Exact Hausdorff dimension in random recursive constructions. *Memoirs of the American Mathematical Society* **381**.
154. Griffeath, D. (1979). *Additive and Cancellative Interacting Particle Systems.* Lecture Notes in Mathematics, vol. 724. Springer, Berlin.
155. Grimmett, G. R. (1978). The rank functions of large random lattices. *Journal of London Mathematical Society* **18**, 567–575.
156. Grimmett, G. R. (1989). *Percolation.* Springer-Verlag, Berlin.
157. Grimmett, G. R. (1993). Differential inequalities for Potts and random-cluster processes. *Cellular Automata and Cooperative Systems* (N. Boccara et al., eds.), Kluwer, Dordrecht, pp. 227–236.
158. Grimmett, G. R. (1994). Potts models and random-cluster processes with many-body interactions. *Journal of Statistical Physics* **75**, 67–121.
159. Grimmett, G. R. (1994). The random-cluster model. *Probability, Statistics and Optimisation* (F. P. Kelly, ed.), John Wiley & Sons, Chichester, pp. 49–63.
160. Grimmett, G. R. (1994). Percolative problems. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 69–86.
161. Grimmett, G. R. (1994). *Probability and Phase Transition.* editor. Kluwer, Dordrecht.
162. Grimmett, G. R. (1995). Comparison and disjoint-occurrence inequalities for random-cluster models. *Journal of Statistical Physics* **78**, 1311–1324.
163. Grimmett, G. R. (1995). The stochastic random-cluster process and the uniqueness of random-cluster measures. *Annals of Probability* **23**, 1461–1510.
164. Grimmett, G. R., Kesten, H., and Zhang, Y. (1993). Random walk on the infinite cluster of the percolation model. *Probability Theory and Related Fields* **96**, 33–44.

165. Grimmett, G. R. and Marstrand, J. M. (1990). The supercritical phase of percolation is well behaved. *Proceedings of the Royal Society (London), Series A* **430**, 439–457.
166. Grimmett, G. R., Menshikov, M. V., and Volkov, S. E. (1996). Random walks in random labyrinths. *Markov Processes and Related Fields* **2**, 69–86.
167. Grimmett, G. R. and Newman, C. M. (1990). Percolation in ∞ + 1 dimensions. *Disorder in Physical Systems* (G. R. Grimmett and D. J. A. Welsh, eds.), Clarendon Press, Oxford, pp. 167–190.
168. Grimmett, G. R. and Piza, M. S. T. (1996). Decay of correlations in subcritical Potts and random-cluster models (to appear).
169. Grimmett, G. R. and Stacey, A. M. (1996). Inequalities for critical probabilities of site and bond percolation (to appear).
170. Grimmett, G. R. and Stirzaker, D. R. (1992). *Probability and Random Processes.* Oxford University Press, Oxford.
171. Häggström, O. (1996). The random-cluster model on a homogeneous tree. *Probability Theory and Related Fields* **104**, 231–253.
172. Häggström, O. and Pemantle, R. (1995). First-passage percolation and a model for competing spatial growth (to appear).
173. Häggström, O., Peres, Y., and Steif, J. (1995). Dynamical percolation (to appear).
174. Hammersley, J. M. (1957). Percolation processes. Lower bounds for the critical probability. *Annals of Mathematical Statistics* **28**, 790–795.
175. Hammersley, J. M. (1959). Bornes supérieures de la probabilité critique dans un processus de filtration. *Le Calcul de Probabilités et ses Applications*, CNRS, Paris, pp. 17–37.
176. Hara, T. and Slade, G. (1989). The mean-field critical behaviour of percolation in high dimensions. *Proceedings of the IXth International Congress on Mathematical Physics* (B. Simon, A. Truman, I. M. Davies, eds.), Adam Hilger, Bristol, pp. 450–453.
177. Hara, T. and Slade, G. (1989). The triangle condition for percolation. *Bulletin of the American Mathematical Society* **21**, 269–273.
178. Hara, T. and Slade, G. (1990). Mean-field critical behaviour for percolation in high dimensions. *Communications in Mathematical Physics* **128**, 333–391.
179. Hara, T. and Slade, G. (1994). Mean-field behaviour and the lace expansion. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 87–122.
180. Hara, T. and Slade, G. (1995). The self-avoiding walk and percolation critical points in high dimensions. *Combinatorics, Probability, Computing* **4**, 197–215.
181. Harris, T. E. (1960). A lower bound for the critical probability in a certain percolation process. *Proceedings of the Cambridge Philosophical Society* **56**, 13–20.
182. Hauge, E. H. and Cohen, E. G. D. (1967). Normal and abnormal effects in Ehrenfest wind-tree model. *Physics Letters* **25A**, 78–79.
183. Hawkes, J. (1981). Trees generated by a simple branching process. *Journal of the London Mathematical Society* **24**, 373–384.

184. Higuchi, Y. (1989). A remark on the percolation for the 2D Ising model. *Osaka Journal of Mathematics* **26**, 207–224.
185. Higuchi, Y. (1991). Level set representation for the Gibbs state of the ferromagnetic Ising model. *Probability Theory and Related Fields* **90**, 203–221.
186. Higuchi, Y. (1992). *Percolation* (in Japanese). Yuseisha/Seiunsha.
187. Higuchi, Y. (1993). A sharp transition for the two-dimensional Ising percolation. *Probability Theory and Related Fields* **97**, 489–514.
188. Higuchi, Y. (1993). Coexistence of infinite *-clusters II: Ising percolation in two dimensions. *Probability Theory and Related Fields* **97**, 1–34.
189. Higuchi, Y. (1996). Ising model percolation. *Sugaku Expositions*, vol. 47, pp. 111–124.
190. Higuchi, Y. (1996). Mixing condition and percolation for Ising models (to appear).
191. Hintermann, D., Kunz, H., and Wu, F. Y. (1978). Exact results for the Potts model in two dimensions. *Journal of Statistical Physics* **19**, 623–632.
192. Holley, R. (1974). Remarks on the FKG inequalities. *Communications in Mathematical Physics* **36**, 227–231.
193. Imbrie, J. and Newman, C. M. (1988). An intermediate phase with slow decay of correlations in one dimensional $1/|x-y|^2$ percolation, Ising and Potts models. *Communications in Mathematical Physics* **118**, 303–336.
194. Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik* **31**, 253–258.
195. Jacobs, D. J. and Thorpe, M. F. (1995). Generic rigidity percolation: the pebble game. *Physical Review Letters* **75**, 4051–4054.
196. Jacobs, D. J. and Thorpe, M. F. (1996). Rigidity percolation in two dimensions. The pebble game. *Physical Review E* **53**, 3682–3693.
197. Jagers, P. (1975). *Branching Processes with Biological Applications.* John Wiley, Chichester.
198. Kantor, T. and Hassold, G. N. (1988). Topological entanglements in the percolation problem. *Physical Review Letters* **60**, 1457–1460.
199. Kasteleyn, P. W. and Fortuin, C. M. (1969). Phase transitions in lattice systems with random local properties. *Journal of the Physical Society of Japan* **26,** 11–14, Supplement.
200. Kesten, H. (1980). The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Communications in Mathematical Physics* **74**, 41–59.
201. Kesten, H. (1980). On the time constant and path length of first-passage percolation. *Advances in Applied Probability* **12**, 848–863.
202. Kesten, H. (1982). *Percolation Theory for Mathematicians.* Birkhäuser, Boston.
203. Kesten, H. (1986). Aspects of first-passage percolation. *Ecole d'Eté de Probabilités de Saint Flour* XIV-1984 (P. L. Hennequin, ed.), Lecture Notes in Mathematics no. 1180, Springer, Berlin, pp. 125–264.
204. Kesten, H. (1987). Scaling relations for 2D-percolation. *Communications in Mathematical Physics* **109**, 109–156.

205. Kesten, H. (1987). A scaling relation at criticality for 2D-percolation. *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.), Springer, Berlin, pp. 203–212.
206. Kesten, H. (1987). Percolation theory and first-passage percolation. *Annals of Probability* **15**, 1231–1271.
207. Kesten, H. (1988). Recent results in rigorous percolation theory. *Astérisque* **157–158**.
208. Kesten, H. (1988). Correlation length and critical probabilities of slabs for percolation, preprint.
209. Kesten, H. (1990). Asymptotics in high dimensions for percolation. *Disorder in Physical Systems* (G. R. Grimmett and D. J. A. Welsh, eds.), Clarendon Press, Oxford.
210. Kesten, H. (1991). Asymptotics in high dimensions for the Fortuin–Kasteleyn random cluster mode. *Spatial Stochastic Processes* (K. S. Alexander and J. C. Watkins, eds.), Birkhäuser, Boston, pp. 57–85.
211. Kesten, H. (1992). Connectivity of certain graphs on halfspaces, quarter spaces, ... *Probability Theory* (L. H. Y. Chen, K. P. Choi, K. Hu, and J.-H. Lou, eds.), Walter de Gruyter, Berlin, pp. 91–104.
212. Kesten, H. (1992). First- and last-passage percolation. *Proceedings of Statistics Conference in Rio de Janeiro, Brasil.*
213. Kesten, H. and Schonmann, R. H. (1990). Behavior in large dimensions of the Potts and Heisenberg models. *Reviews in Mathematical Physics* **1**, 147–182.
214. Kesten, H. and Zhang, Y. (1987). Strict inequalities for some critical exponents in two-dimensional percolation. *Journal of Statistical Physics* **46**, 1031–1055.
215. Kesten, H. and Zhang, Y. (1990). The probability of a large finite cluster in supercritical Bernoulli percolation. *Annals of Probability* **18**, 537–555.
216. Kesten, H. and Zhang, Y. (1993). The tortuosity of occupied crossings of a box in critical percolation. *Journal of Statistical Physics* **70**, 599–611.
217. Kesten, H. and Zhang, Y. (1997). A central limit theorem for 'critical' first passage percolation in two dimensions. *Probability Theory and Related Fields* **107**, 137–160.
218. Kihara, T., Midzuno, Y., and Shizume, J. (1954). Statistics of two-dimensional lattices with many components. *Journal of the Physical Society of Japan* **9**, 681–687.
219. Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusion. *Communications in Mathematical Physics* **104**, 1–19.
220. Klein, A. (1994). Multiscale analysis in disordered systems: Percolation and contact process in a random environment. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 139–152.
221. Klein, A. (1994). Extinction of contact and percolation processes in a random environment. *Annals of Probability* **22**, 1227 1251.
222. Kotecký, R. (1994). Geometric representation of lattice models and large volume asymptotics. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 153–176.

223. Kotecký, R. and Shlosman, S. (1982). First order phase transitions in large entropy lattice systems. *Communications in Mathematical Physics* **83**, 493–515.
224. Kuulasmaa, K. (1982). The spatial general epidemic and locally dependent random graphs. *Journal of Applied Probability* **19**, 745–758.
225. Laanait, L., Messager, A., Miracle-Sole, S., Ruiz, J., and Shlosman, S. (1991). Interfaces in the Potts model I: Pirogov–Sinai theory of the Fortuin–Kasteleyn representation. *Communications in Mathematical Physics* **140**, 81–91.
226. Laanait, L., Messager, A., and Ruiz, J. (1986). Phase coexistence and surface tensions for the Potts model. *Communications in Mathematical Physics* **105**, 527–545.
227. Langlands, R. P. (1993). Dualität bei endlichen Modellen der Perkolation. *Mathematische Nachrichten* **160**, 7–58.
228. Langlands, R. P. (1996). An essay on the dynamics and statistics of critical field theories. *50th anniversary volume of the Canadian Mathematical Society* (to appear).
229. Langlands, R. P. and Lafortune, M.-A. (1994). Finite models for percolation. *Contemporary Mathematics* **177**, 227–246.
230. Langlands, R., Pichet, C., Pouliot, P., and Saint-Aubin, Y. (1992). On the universality of crossing probabilities in two-dimensional percolation. *Journal of Statistical Physics* **67**, 553–574.
231. Langlands, R., Pouliot, P., and Saint-Aubin, Y. (1994). Conformal invariance in two-dimensional percolation. *Bulletin of the American Mathematical Society* **30**, 1–61.
232. Lebowitz, J. and Martin-Löf, A. (1972). On the uniqueness of the equilibrium state for Ising spin systems. *Communications in Mathematical Physics* **25**, 276–282.
233. Licea, C. and Newman, C. M. (1996). Geodesics in two-dimensional first-passage percolation. *Annals of Probability* **24**, 399–410.
234. Licea, C., Newman, C. M., and Piza, M. S. T. (1995). Superdiffusivity in first-passage percolation (to appear).
235. Lieb, E. H. (1980). A refinement of Simon's correlation inequality. *Communications in Mathematical Physics* **77**, 127–135.
236. Liggett, T. M. (1985). *Interacting Particle Systems.* Springer-Verlag, Berlin.
237. Liggett, T. M. (1991). Spatially inhomogeneous contact processes. *Spatial Stochastic Processes* (K. S. Alexander and J. C. Watkins, eds.), Birkhäuser, Boston, pp. 105–140.
238. Liggett, T. M. (1991). The periodic threshold contact process. *Random Walks, Brownian Motion and Interacting Particle Systems* (R. T. Durrett and H. Kesten, eds.), Birkhäuser, Boston, pp. 339–358.
239. Liggett, T. M. (1992). The survival of one-dimensional contact processes in random environments. *Annals of Probability* **20**, 696–723.
240. Liggett, T. M. (1994). Coexistence in threshold voter models. *Annals of Probability* **22**, 764–802.
241. Liggett, T. M. (1995). Improved upper bounds for the contact process critical value. *Annals of Probability* **23**, 697–723.

242. Liggett, T. M. (1995). Survival of discrete time growth models, with applications to oriented percolation. *Annals of Applied Probability* **5**, 613–636.
243. Liggett, T. M. (1996). Multiple transition points for the contact process on the binary tree. *Annals of Probability* **24**, 1675–1710.
244. Liggett, T. M., Schonmann, R. H., and Stacey, A. (1997). Domination by product measures. *Annals of Probability* **25** (to appear).
245. Lorentz, H. A. (1905). The motion of electrons in metallic bodies, I, II, and III. *Koninklijke Akademie van Wetenschappen te Amsterdam, Section of Sciences* **7**, 438–453, 585–593, 684–691.
246. Łuczak, T. and Wierman, J. C. (1988). Critical probability bounds for two-dimensional site percolation models. *Journal of Physics A: Mathematical and General* **21**, 3131–3138.
247. Łuczak, T. and Wierman, J. C. (1989). Counterexamples in AB percolation. *Journal of Physics A: Mathematical and General* **22**, 185–191.
248. Lyons, R. (1990). Random walks and percolation on trees. *Annals of Probability* **18**, 931–958.
249. Lyons, R. (1992). Random walks, capacity and percolation on trees. *Annals of Probability* **20**, 2043–2088.
250. Lyons, R. and Pemantle, R. (1992). Random walk in a random environment and first-passage percolation on trees. *Annals of Probability* **20**, 125–137.
251. Lyons, T. (1983). A simple criterion for transience of a reversible Markov chain. *Annals of Probability* **11**, 393–402.
252. Madras, N., Schinazi, R. B., and Schonmann, R. H. (1994). On the critical behavior of the contact process in deterministic inhomogeneous environment. *Annals of Probability* **22**, 1140–1159.
253. Madras, N. and Slade, G. (1993). *The Self-Avoiding Walk.* Birkhäuser, Boston.
254. Mandelbrot, B. (1983). *The Fractal Geometry of Nature.* W. H. Freeman, San Francisco.
255. Martirosian, D. H. (1986). Translation invariant Gibbs states in the q-state Potts model. *Communications in Mathematical Physics* **105**, 281–290.
256. Mauldin, R. D. and Williams, S. C. (1986). Random recursive constructions: asymptotic geometry and topological properties. *Transactions of the American Mathematical Society* **295**, 325–346.
257. Meester, R. (1989). An algorithm for calculating critical probabilities and percolation functions in percolation models defined by rotations. *Ergodic Theory and Dynamical Systems* **8**, 495–509.
258. Meester, R. (1992). Connectivity in fractal percolation. *Journal of Theoretical Probability* **5**, 775–789.
259. Meester, R. (1994). Uniqueness in percolation theory; a review. *Statistica Neerlandica* **48**, 237–252.
260. Meester, R. (1995). Equality of critical densities in continuum percolation. *Journal of Applied Probability* **32**, 90–104.
261. Meester, R. and Häggström, O. (1995). Asymptotic shapes in stationary first passage percolation. *Annals of Probability* **23**, 1511–1522.

262. Meester, R. and Häggström, O. (1995). Nearest neighbour and hard sphere models in continuum percolation (to appear).
263. Meester, R. and Nowicki, T. (1989). Infinite clusters and critical values in two-dimensional circle percolation. *Israel Journal of Mathematics* **68**, 63–81.
264. Meester, R., Penrose, M., and Sarkar, A. (1995). The random connection model in high dimensions (to appear).
265. Meester, R., Roy, R., and Sarkar, A. (1994). Non-universality and continuity of the critical covered volume fraction in continuum percolation. *Journal of Statistical Physics* **75**, 123–134.
266. Meester, R. and Steif, J. (1996). On the continuity of the critical value for long range percolation in the exponential case. *Communications in Mathematical Physics* **180**, 483–504.
267. Meester, R. and Steif, J. (1995). Consistent estimation of percolation quantities (to appear).
268. Menshikov, M. V. (1986). Coincidence of critical points in percolation problems. *Soviet Mathematics Doklady* **33**, 856–859.
269. Menshikov, M. V. (1987). Quantitative estimates and rigorous inequalities for critical points of a graph and its subgraphs. *Theory of Probability and its Applications* **32**, 544–547.
270. Menshikov, M. V. (1987). Strict inequalities for critical probabilities of a regular graph and its subgraphs. *Uspekhy Math. Nauk* **42**, 239–240.
271. Menshikov, M. V., Molchanov, S. A., and Sidorenko, A. F. (1986). Percolation theory and some applications. *Itogi Nauki i Techniki* (Series of Probability Theory, Mathematical Statistics, Theoretical Cybernetics) **24**, 53–110.
272. Menshikov, M. V. and Pelikh, K. D. (1989). Percolation with several types of defects. Estimates of the critical probability on the square lattice, 778–785 (in translation). *Mathematical Notes* **46**, 38–47 (in Russian).
273. Menshikov, M. V. and Pelikh, K. D. (1993). Percolation method in a system of numerous particles. *Proceeding of Third Conference on Probability Methods in Discrete Mathematics, Petrozavodsk*, TVP/VSP, pp. 329–336.
274. Menshikov, M. V. and Zuev, S. A. (1993). Percolation method in a system of numerous particles. *Probabilistic Methods in Discrete Mathematics* (V. F. Kolchin, ed.), TVP/VSP, pp. 329–336.
275. Menshikov, M. V. and Zuev, S. A. (1993). Models of ρ-percolation. *Probabilistic Methods in Discrete Mathematics* (V. F. Kolchin, ed.), TVP Science Publishers, pp. 337–347.
276. Morrow, G., Schinazi, R., and Zhang, Y. (1994). The critical contact process on a homogeneous tree. *Journal of Applied Probability* **31**, 250–255.
277. Nash-Williams, C. St. J. A. (1959). Random walks and electric currents in networks. *Proceedings of the Cambridge Philosophical Society* **55**, 181–194.
278. Newman, C. M. (1986). Some critical exponent inequalities for percolation. *Journal of Statistical Physics* **45**, 359–368.
279. Newman, C. M. (1986). Percolation theory: a selective survey of rigorous results. *Advances in Multiphase Flow and Related Problems* (G. Papanicolaou, ed.), SIAM, Philadelphia, pp. 147–167.

280. Newman, C. M. (1987). Inequalities for γ and related critical exponents in short and long range percolation. *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.), Springer, Berlin, pp. 229–244.
281. Newman, C. M. (1987). Another critical exponent inequality for percolation: $\beta \geq 2/\delta$. *Journal of Statistical Physics* **47**, 695–699.
282. Newman, C. M. (1989). The $1/r^2$ phase transition: a review. *Proceedings of the IXth International Congress on Mathematical Physics* (B. Simon, A. Truman, and I. M. Davies, eds.), Adam Hilger, Bristol, pp. 377–379.
283. Newman, C. M. (1991). Ising models and dependent percolation. *Topics in Statistical Dependence* (H. W. Block, A. R. Sampson, and T. H. Savits, eds.), IMS Lecture Notes, Monograph Series, vol. 16, pp. 395–401.
284. Newman, C. M. (1991). Topics in percolation. *Mathematics of Random Media* (W. Kohler and B. White, eds.), vol. 27, American Mathematical Society, Providence, pp. 41–51.
285. Newman, C. M. (1994). Disordered Ising systems and random cluster representations. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 247–260.
286. Newman, C. M. (1995). A surface view of first-passage percolation. *Proceedings of The International Congress of Mathematicians* (S. D. Chatterji, ed.), Birkhäuser, Boston, pp. 1017–1023.
287. Newman, C. M. and Piza, M. S. T. (1995). Divergence of shape fluctuations in two dimensions. *Annals of Probability* **23**, 977–1005.
288. Newman, C. M. and Schulman, L. S. (1981). Infinite clusters in percolation models. *Journal of Statistical Physics* **26**, 613–628.
289. Newman, C. M. and Stein, D. L. (1990). Broken symmetry and domain structure in Ising-like systems. *Physical Review Letters* **65**, 460–463.
290. Newman, C. M. and Stein, D. L. (1992). Multiple states and thermodynamic limits in short-ranged Ising spin glass models. *Physical Review B* **46**, 973–982.
291. Newman, C. M. and Stein, D. L. (1993). Chaotic size dependence in spin glasses. *Cellular Automata and Cooperative Systems* (N. Boccara, E. Goles, S. Martinez, and P. Picco, eds.), Kluwer, Dordrecht, pp. 525–529.
292. Newman, C. M. and Stein, D. L. (1994). Spin glass model with dimension-dependent ground state multiplicity. *Physical Review Letters* **72**, 2286–2289.
293. Newman, C. M. and Stein, D. L. (1995). Random walk in a strongly inhomogeneous environment and invasion percolation. *Annales de l'Institut Henri Poincaré* **31**, 249–261.
294. Newman, C. M. and Stein, D. L. (1995). Broken ergodicity and the geometry of rugged landscapes. *Physical Review E* **51**, 5228–5238.
295. Newman, C. M. and Stein, D. L. (1996). Non-mean-field behavior of realistic spin glasses. *Physical Review Letters* **76**, 515–518.
296. Newman, C. M. and Stein, D. L. (1996). Ground state structure in a highly disordered spin glass model. *Journal of Statistical Physics* **82**, 1113–1132.
297. Newman, C. M. and Stein, D. L. (1995). Spatial inhomogeneity and thermodynamic chaos (to appear).

298. Newman, C. M. and Volchan, S. B. (1996). Persistent survival of one-dimensional contact processes in random environments. *Annals of Probability* **24**, 411–421.
299. Newman, C. M. and Wu, C. C. (1990). Markov fields on branching planes. *Probability Theory and Related Fields* **85**, 539–552.
300. Onsager, L. (1944). Crystal statistics, I. A two-dimensional model with an order-disorder transition. *The Physical Review* **65**, 117–149.
301. Ornstein, L. S. and Zernike, F. (1915). Accidental deviations of density and opalescence at the critical point of a single substance. *Koninklijke Akademie van Wetenschappen te Amsterdam, Section of Sciences* **17**, 793–806.
302. Orzechowski, M. E. (1996). On the phase transition to sheet percolation in random Cantor sets. *Journal of Statistical Physics* **82**, 1081–1098.
303. Pemantle, R. (1991). Choosing a spanning tree for the integer lattice uniformly. *Annals of Probability* **19**, 1559–1574.
304. Pemantle, R. (1992). The contact process on trees. *Annals of Probability* **20**, 2089–2116.
305. Pemantle, R. (1995). Uniform random spanning trees. *Topics in Contemporary Probability and its Applications* (J. L. Snell, ed.), CRC Press, Boca Raton.
306. Pemantle, R. and Peres, Y. (1994). Planar first-passage percolation times are not tight. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 261–264.
307. Pemantle, R. and Peres, Y. (1994). Domination between trees and application to an explosion problem. *Annals of Probability* **22**, 180–194.
308. Pemantle, R. and Peres, Y. (1996). On which graphs are all random walks in random environments transient?. *Random Discrete Structures* (D. Aldous and R. Pemantle, eds.), IMA Vol. Math. Appl. 76, Springer, New York, pp. 207–211.
309. Pemantle, R. and Peres, Y. (1996). Galton–Watson trees with the same mean have the same polar sets. *Annals of Probability* **23**, 1102–1124.
310. Penrose, M. (1996). Continuum percolation and Euclidean minimal spanning trees in high dimensions. *Annals of Applied Probability* **6**, 528–544.
311. Penrose, M. and Pisztora, A. (1996). Large deviations for discrete and continuous percolation. *Advances in Applied Probability* **28**, 29–52.
312. Peres, Y. (1996). Intersection-equivalence of Brownian paths and certain branching processes. *Communications in Mathematical Physics* **177**, 417–434.
313. Peyrière, J. (1978). Mandelbrot random beadsets and birthprocesses with interaction, I.B.M. Research Report RC-7417.
314. Pirogov, S. A. and Sinai, Ya. G. (1975). Phase diagrams of classical lattice systems. *Theoretical and Mathematical Physics* **25**, 1185–1192.
315. Pirogov, S. A. and Sinai, Ya. G. (1976). Phase diagrams of classical lattice systems, continuation. *Theoretical and Mathematical Physics* **26**, 39–49.
316. Pisztora, A. (1996). Surface order large deviations for Ising, Potts and percolation models. *Probability Theory and Related Fields* **104**, 427–466.
317. Pokorny, M., Newman, C. M., and Meiron, D. (1990). The trapping transition in dynamic (invasion) and static percolation. *Journal of Physics A: Mathematical and General* **23**, 1431–1438.

318. Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society* **48**, 106–109.
319. Preston, C. J. (1974). *Gibbs States on Countable Sets.* Cambridge University Press, Cambridge.
320. Propp J. G. and Wilson, D. B. (1995). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* (to appear).
321. Quas, A. (1996). Some properties of Lorentz lattice gas models (to appear).
322. Reimer, D. (1995). Butterflies (to appear).
323. Roy, R. and Meester, R. (1994). Uniqueness of unbounded occupied and vacant components in Boolean models. *Advances in Applied Probability* **4**, 933–951.
324. Roy, R. and Meester, R. (1996). *Continuum Percolation.* Cambridge University Press, Cambridge.
325. Ruijgrok, T. W. and Cohen, E. G. D. (1988). Deterministic lattice gas models. *Physics Letters A* **133**, 415–418.
326. Russo, L. (1978). A note on percolation. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **43**, 39–48.
327. Russo, L. (1981). On the critical percolation probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **56**, 229–238.
328. Russo, L. (1982). An approximate zero-one law. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61**, 129–139.
329. Scheinerman, E. R. and Wierman, J. C. (1987). Infinite AB percolation clusters exist. *Journal of Physics A: Mathematical and General* **20**, 1305–1307.
330. Schonmann, R. H. (1994). Theorems and conjectures on the droplet-driven relaxation of stochastic Ising models. *Probability and Phase Transition* (G. R. Grimmett, ed.), Kluwer, Dordrecht, pp. 265–310.
331. Seymour, P. D. and Welsh, D. J. A. (1978). Percolation probabilities on the square lattice. *Advances in Graph Theory* (B. Bollobás, ed.), Annals of Discrete Mathematics 3, North-Holland, Amsterdam, pp. 227–245.
332. Simon, B. (1980). Correlation inequalities and the decay of correlations in ferromagnets. *Communications in Mathematical Physics* **77**, 111–126.
333. Sinai, Ya. G. (1982). *Theory of Phase Transitions: Rigorous Results.* Pergamon Press, Oxford.
334. Slade, G. (1991). The lace expansion and the upper critical dimension for percolation. *Mathematics of Random Media* (W. E. Kohler and B. S. White, eds.), Lectures in Applied Mathematics, vol. 27, American Mathematical Society, Providence, pp. 53–63.
335. Slade, G. (1995). Bounds on the self-avoiding walk connective constant, Special Issue: Proceedings of the Conference in Honor of Jean-Pierre Kahane, 1993. *Journal of Fourier Analysis and its Applications*, 525–533.
336. Stacey, A. M. (1996). The existence of an intermediate phase for the contact process on trees. *Annals of Probability* **24**, 1711–1726.
337. Stauffer, D. (1981). Scaling properties of percolation clusters. *Disordered Systems and Localization* (C. Castellani, C. DiCastro, and L. Peliti, eds.), Springer, Berlin, pp. 9–25.

338. Swendsen, R. H. and Wang, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters* **58**, 86–88.
339. Sykes, M. F. and Essam, J. W. (1964). Exact critical percolation probabilities for site and bond problems in two dimensions. *Journal of Mathematical Physics* **5**, 1117–1127.
340. Tanemura, H. (1995). Critical behavior for a continuum percolation model (to appear).
341. Vahidi-Asl, M. Q. and Wierman, J. C. (1990). First-passage percolation on the Voronoi tesselation and Delaunay triangulation. *Random Graphs '87*, Wiley, London, pp. 341–359.
342. Vahidi-Asl, M. Q. and Wierman, J. C. (1990). A shape result for first-passage percolation on the Voronoi tesselation and Delaunay triangulation. *Random Graphs '89*, Wiley, London, pp. 247–262.
343. Vahidi-Asl, M. Q. and Wierman, J. C. (1993). Upper and lower bounds for the route length of first-passage percolation in Voronoi tesselations. *Bulletin of the Iranian Mathematical Society* **19**, 15-28.
344. Wang, F. and Cohen, E. G. D. (1995). Diffusion in Lorentz lattice gas cellular automata: the honeycomb and quasi-lattices compared with the square and triangular lattices. *Journal of Statistical Physics* **81**, 467–495.
345. Watts, G. M. T. (1996). A crossing probability for critical percolation in two dimensions. *Journal of Physics A: Mathematical and General* **29**, L363–L368.
346. Welsh, D. J. A. (1993). Percolation in the random-cluster process. *Journal of Physics A: Mathematical and General* **26**, 2471–2483.
347. Whittle, P. (1986). *Systems in Stochastic Equilibrium*. John Wiley, Chichester.
348. Whittle, P. (1994). Polymer models and generalised Potts–Kasteleyn models. *Journal of Statistical Physics* **75**, 1063–1092.
349. Wierman, J. C. (1987). Directed site percolation and dual filling models. *Annals of Discrete Mathematics* **33**, 339–352.
350. Wierman, J. C. (1987). Duality for k-degree percolation on the square lattice. *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.), Springer, Berlin, pp. 311–323.
351. Wierman, J. C. (1988). Bond percolation critical probability bounds derived by edge contraction. *Journal of Physics A: Mathematical and General* **21**, 1487–1492.
352. Wierman, J. C. (1988). AB percolation on close-packed graphs. *Journal of Physics A: Mathematical and General* **21**, 1939–1944.
353. Wierman, J. C. (1988). On AB percolation in bipartite graphs. *Journal of Physics A: Mathematical and General* **21**, 1945–1949.
354. Wierman, J. C. (1989). AB percolation: a brief survey. *Combinatorics and Graph Theory*, vol. 25, Banach Centre Publications, Warsaw, pp. 241–251.
355. Wierman, J. C. (1990). Bond percolation critical probability bounds for the Kagomé lattice by a substitution method. *Disorder in Physical Systems* (G. R. Grimmett and D. J. A. Welsh, eds.), Clarendon Press, Oxford, pp. 349–360.

356. Wierman, J. C. (1992). Equality of the bond percolation critical exponents for two pairs of dual lattices. *Combinatorics, Probability, Computing* **1**, 95–105.
357. Wierman, J. C. (1994). Equality of directional critical exponents in multiparameter percolation models. *Journal of Physics A: Mathematical and General* **27**, 1851–1858.
358. Wierman, J. C. (1995). Substitution method critical probability bounds for the square lattice site percolation model. *Combinatorics, Probability, Computing* **4**, 181–188.
359. Wilson, R. J. (1979). *Introduction to Graph Theory.* Longman, London.
360. Wu, F. Y. (1982). The Potts model. *Reviews in Modern Physics* **54**, 235–268.
361. Yang, W. and Zhang, Y. (1992). A note on differentiability of the cluster density for independent percolation in high dimensions. *Journal of Statistical Physics* **66**, 1123–1138.
362. Zhang, Y. (1991). A power law for connectedness of some random graphs at the critical point. *Random Structures and Algorithms* **2**, 101–119.
363. Zhang, Y. (1992). Failure of the power laws on some subgraphs of the Z^2 lattice. *Journal of Physics A: Mathematical and General* **25**, 6617–6622.
364. Zhang, Y. (1993). A shape theorem for epidemics and forest fires with finite range interactions. *Annals of Probability* **21**, 1755–1781.
365. Zhang, Y. (1994). A note on inhomogeneous percolation. *Annals of Probability* **22**, 803–820.
366. Zhang, Y. (1994). Analyticity properties at the supercritical state, preprint.
367. Zhang, Y. (1995). The fractal volume of the two-dimensional invasion percolation cluster. *Communications in Mathematical Physics* **167**, 237–254.
368. Zhang, Y. (1995). Supercritical behaviors in first-passage percolation. *Stochastic Processes and their Applications* **59**, 251–266.
369. Zhang, Y. (1995). A limit theorem for matching random sequences allowing deletions. *Annals of Applied Probability* **5**, 1236–1240.
370. Zhang, Y. (1996). The complete convergence theorem on trees. *Annals of Probability* **24**, 1408–1443.
371. Zhang, Y. (1996). Continuity of percolation probability in $\infty + 1$ dimensions. *Journal of Applied Probability* **33**, 427–433.
372. Zhang, Y. (1996). Two critical behaviors of first passage time, preprint.
373. Zhang, Y. (1996). Some power laws on two dimensional critical bond percolation, preprint.
374. Zhang, Y. (1996). Divergence of the bulk resistance at criticality in disordered media. *Journal of Statistical Physics* **84**, 263–267.
375. Zhang, Y. and Zhang, Y. C. (1984). A limit theorem for N_{0n}/n in first passage percolation. *Annals of Probability* **12**, 1068–1076.
376. Ziff, R. M., Kong, X. P., and Cohen, E. G. D. (1991). Lorentz lattice-gas and kinetic-walk model. *Physical Review A* **44**, 2410–2428.

LECTURES ON FINITE MARKOV CHAINS

Laurent SALOFF-COSTE

Contents

1 Introduction and background material **304**
1.1 Introduction . . . 304
1.1.1 My own introduction to finite Markov chains . . . 305
1.1.2 Who cares? . . . 307
1.1.3 A simple open problem . . . 309
1.2 The Perron-Frobenius Theorem . . . 309
1.2.1 Two proofs of the Perron-Frobenius theorem . . . 310
1.2.2 Comments on the Perron-Frobenius theorem . . . 313
1.2.3 Further remarks on strong irreducibility . . . 315
1.3 Elementary functional analysis . . . 316
1.3.1 Operator norms . . . 316
1.3.2 Hilbert space techniques . . . 318
1.4 Notation for finite Markov chains . . . 319
1.4.1 Discrete time versus continuous time . . . 322

2 Analytic tools **326**
2.1 Nothing but the spectral gap . . . 326
2.1.1 The Dirichlet form . . . 326
2.1.2 The spectral gap . . . 327
2.1.3 Chernoff bounds and central limit theorems . . . 331
2.2 Hypercontractivity . . . 332
2.2.1 The log-Sobolev constant . . . 332
2.2.2 Hypercontractivity, α, and ergodicity . . . 333
2.2.3 Some tools for bounding α from below . . . 338
2.3 Nash inequalities . . . 344
2.3.1 Nash's argument for finite Markov chains I . . . 345
2.3.2 Nash's argument for finite Markov chains II . . . 347
2.3.3 Nash inequalities and the log-Sobolev constant . . . 350
2.3.4 A converse to Nash's argument . . . 351
2.3.5 Nash inequalities and higher eigenvalues . . . 352
2.3.6 Nash and Sobolev inequalities . . . 355
2.4 Distances . . . 358
2.4.1 Notation and inequalities . . . 358
2.4.2 The cutoff phenomenon and related questions . . . 361

3 Geometric tools **367**

3.1 Adapted edge sets 368

3.2 Poincaré inequality 368

3.3 Isoperimetry 379

3.3.1 Isoperimetry and spectral gap 379

3.3.2 Isoperimetry and Nash inequalities 386

3.3.3 Isoperimetry and the log-Sobolev constant 393

3.4 Moderate growth 394

4 Comparison techniques **396**

4.1 Using comparison inequalities 396

4.2 Comparison of Dirichlet forms using paths 400

Chapter 1

Introduction and background material

1.1 Introduction

I would probably never have worked on finite Markov chains if I had not met Persi Diaconis. These notes are based on our joint work and owe a lot to his broad knowledge of the subject although the presentation of the material would have been quite different if he had given these lectures.

The aim of these notes is to show how functional analysis techniques and geometric ideas can be helpful in studying finite Markov chains from a quantitative point of view.

A Markov chain will be viewed as a Markov operator K acting on functions defined on the state space. The action of K on the spaces $\ell^p(\pi)$ where π is the stationary measure of K will be used as an important tool. In particular, the Hilbert space $\ell^2(\pi)$ and the Dirichlet form

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x,y)\pi(x)$$

associated to K will play crucial roles. Functional inequalities such as Poincaré inequalities, Sobolev and Nash inequalities, or Logarithmic Sobolev inequalities will be used to study the behavior of the chain.

There is a natural graph structure associated to any finite Markov chain K. The geometry of this graph and the combinatorics of paths enter the game as tools to prove functional inequalities such as Poincaré or Nash inequalities and also to study the behavior of different chains through comparison of their Dirichlet forms.

The potential reader should be aware that these notes contain no probabilistic argument. Coupling and strong stationary times are two powerful techniques that have also been used to study Markov chains. They form a set of techniques that are very different in spirit from the one presented here. See, e.g., [1, 19]. Diaconis' book [17] contains a chapter on these techniques. David Aldous and Jim Fill are writing a book on finite Markov chains [3] that contains many wonderful things.

The tools and ideas presented in these notes have emerged recently as useful techniques to obtain quantitative convergence results for complex finite Markov chains. I have tried to illustrate these techniques by natural, simple but non trivial examples. More complex (and more interesting) examples require too much additional specific material to be treated in these notes. Here are a few references containing compelling examples:

- For eigenvalue estimates using path techniques, see [35, 41, 53, 72].
- For comparison techniques, see [23, 24, 30]
- For other geometric techniques, see [21, 38, 39, 43, 60].

Acknowledgements: Many thanks to Michel Benaim, Sergei Bobkov, Persi Diaconis, Susan Holmes, Michel Ledoux, Pascal Lezaud and Laurent Miclo for their help. Thanks also to David Aldous, Jim Fill, Mark Jerrum, Alistair Sinclair for useful discussions and comments over the years.

1.1.1 My own introduction to finite Markov chains

Finite Markov chains provide nice exercises in linear algebra and elementary probability theory. For instance, they can serve to illustrate diagonalization or triangularization in linear algebra and the notion of conditional probability or stopping times in probability. That is often how the subject is known to professional mathematicians.

The ultimate results then appear to be the classification of the states and, in the ergodic case, the existence of an invariant measure and the convergence of the chain towards its invariant measure at an exponentiel rate (the Perron-Frobenius theorem). Indeed, this set of results describes well the asymptotic behavior of the chain.

I used to think that way, until I heard Persi Diaconis give a couple of talks on card shuffling and other examples.

How many times do you have to shuffle a deck of cards so that the deck is well mixed?

The fact that shuffling many, many times does mix (the Perron-Frobenius Theorem) is reassuring but does not at all answer the question above.

Around the same time I started to read a paper by David Aldous [1] on the subject because a friend of mine, a student at MIT, was asking me questions about it. I was working on analysis on Lie groups and random walk on finitely generated, infinite group under the guidance of Nicolas Varopoulos. I had the vague feeling that the techniques that Varopolous had taught me could also be applied to random walks on finite groups. Of course, I had trouble deciding whether this feeling was correct or not because, on a finite set, everything is always true, any functional inequality is satisfied with appropriate constants.

Consider an infinite group G, generated by a finite symmetric set S. The associated random walk proceeds by picking an element s in S at random and move from the current state x to xs. An important nontrivial result in random walk theory is that the transient/recurrent behavior of these walks depends only on G and not on the choosen generating set S. The proof proceeds by comparison of Dirichlet forms. The Dirichlet form associated to S is

$$\mathcal{E}_S(f,f) = \frac{1}{2|S|} \sum_{g \in G, h \in S} |f(g) - f(gh)|^2.$$

If S and T are two generating sets, one easily shows that there are constants $a, A > 0$ such that

$$a\mathcal{E}_S \leq \mathcal{E}_T \leq A\mathcal{E}_S.$$

To prove these inequalities one writes the elements of S as finite products of elements of T and vice versa. They can be used to show that the behavior of finitely generated symmetric random walks on G, in many respects, depends only on G, not on the generating set.

I felt that this should have a meaning on finite groups too although clearly, on a finite group, different generating finite sets may produce different behaviors.

I went to see Persi Diaconis and we had the following conversation:

L: Do you have an example of finite group on which there are many different walks of interest?
P: Yes, the symmetric group S_n!
L: Is there a walk that you really know well?
P: Yes there is. I know a lot about random transpositions.
L: Now, we need another walk that you do not know as well as you wish.
P: Take the generators $\tau = (1,2)$ and $c^{\pm 1} = (1, \ldots, n)^{\pm 1}$.
L& P: Lets try it. Any transposition can be written as a product of τ and $c^{\pm 1}$ of length at most $10n$. Each of τ, c, c^{-1} is used at most $10n$ times to write a given transposition. Hence, (after some computations) we get

$$\mathcal{E}_T \leq 100\, n^2\, \mathcal{E}_S$$

where $\mathcal{E}_T$ is the Dirichlet form for random transpositions and $S = \{\tau, c, c^{-1}\}$. What can we do with this? Well, the first nontrivial eigenvalue of random transpositions is $1 - 2/n$ by Fourier analysis. This yields a bound of order $1 - 50/n^3$ for the walk based on the generating set S.
L: I have no idea whether this is good or not.
P: Well, I do not know how to get this result any other way (as we later realized $1 - c/n^3$ is the right order of magnitude for the first nontrivial eigenvalue of the walk based on S).
L: Do you have any other example? ....

This took place during the spring of 1991. The conversation is still going on and these notes are based on it.

1.1.2 Who cares?

There are many ways in which finite Markov chains appear as interesting or useful objects. This section presents briefly some of the aspects that I find most compelling.

Random walks on finite groups. I started working on finite Markov chains by looking at random walks on finite groups. This is still one of my favorite aspects of the subject. Given a finite group G and a generating set $S \subset G$, define a Markov chain as follows. If the current state is g, pick s in S uniformly at random and move to gs. For instance, take $G = S_n$ and $S = \{\mathrm{id}\} \cup \{(i,j) : 1 \le i < j \le n\}$. This yields the "random transpositions" walk. Which generating sets of S_n are most efficient? Which sets yield random walks that are slow to converge? How slow can it be? More generally, which groups carry fast generating sets of small cardinality? How does the behavior of random walks relate to the algebraic structure of the group? These are some of the questions that one can ask in this context. These notes do not study finite random walks on groups in detail except for a few examples. The book [17] gives an introduction and develops tools from Fourier analysis and probability theory. See also [42]. The survey paper [27] is devoted to random walks on finite groups. It contains pointers to the literature and some open questions. Many examples of walks on the symmetric group are treated by comparison with random transpositions in [24]. M. Hildebrand [49] studies random transvections in finite linear groups by Fourier analysis. The recent paper of D. Gluck [45] contains results for some classical finite groups that are based on the classification of simple finite groups. Walks on finite nilpotent groups are studied in [25, 26] and in [74, 75, 76].

Markov Chain Monte Carlo. Markov chain Monte Carlo algorithms use a Markov chain to draw from a given distribution π on a state space $\mathcal{X}$ or to approximate π and compute quantities such as $\pi(f)$ for certain functions f. The **Metropolis** algorithm and its variants provide ways of constructing Markov chains which have the desired distribution π as stationary measure. For instance let Λ be a 100 by 100 square grid, $\mathcal{X} = \{x : \Lambda \to \{\pm 1\}\}$ and

$$\pi(x) = z(c)^{-1} \exp\left\{ c \left(\sum_{i,j:i\sim j} x_i x_j + h \sum_i x_i \right) \right\}$$

where $z(c)$ is the unknown normalizing constant. This is the **Gibbs** measure of a finite two-dimentional Ising model with inverse temperature $c > 0$ and external field strength h. In this case the Metropolis chain proceed as follows. Pick a site $i \in \Lambda$ at random and propose the move $x \to x^i$ where x^i is obtained from x by changing $x(i)$ to $-x(i)$. If $\pi(x^i)/\pi(x) \ge 1$ accept this move. If not, flip a coin with probability of heads $\pi(x^i)/\pi(x)$. If the coin comes up heads, move to x^i. If the coins comes up tails, stay at x. It is not difficult to show that this chain has stationary measure π as desired. It can then be used (in principle) to draw from π (i.e., to produce typical configurations), or to estimate the normalizing constant $z(c)$. Observe that running this chain implies computing $\pi(x^i)/\pi(x)$. This is reasonable because the unknown normalizing constant disappears in this ratio and the computation only involves looking at neighbors of the site i.

Application of the Metropolis algorithm are widespread. Diaconis recommends looking at papers in the Journal of the Royal Statistical Society, Series B, 55(3), (1993) for examples and pointers to the literature. Clearly, to validate (from a theoretical point of view) the use of this type of algorithm one needs to be able to answer the question: how many steps are sufficient (necessary) for the chain to yield a good approximation of π? These chains and algorithms are often used without any theoretical knowledge of how long they should be run. Instead, the user most often relies on experimental knowledge, hoping for the best.

Let us emphasize here the difficulties that one encounters in trying to produce theoretical results that bear on applications. In order to be directly relevant to applied work, theoretical results concerning finite Markov chains must not only be quantitative but they must yield bounds that are close to be sharp. If the bounds are not sharp enough, the potential user is likely to disregard them as unreasonably conservative (and too expensive in running time). It turns out that many finite Markov chains are very effective (i.e., are fast to reach stationarity) for reasons that seem to defy naive analysis. A good example is given by the Swendsen-Wang algorithm which is a popular sampling procedure for Ising configuration according to the Gibbs distribution [77]. This algorithm appears to work extremely well but there are no quantitative theoretical results to support this experimental finding. A better understood example of this phenomenon is given by random transpositions (and other walks) on the symmetric group. In this case, a precise analysis can be obtained through the well developed representation theory of the symmetric group. See [17].

Theoretical Computer Science. Much recent progress in quantitative finite Markov chain theory is due to the Computer Science community. I refer the reader to [54, 56, 71, 72] and also [31] for pointers to this literature. Computer scientists are interested in classifying various combinatorial tasks according to their complexity. For instance, given a bipartite connected graph on $2n$ vertices with vertex set $O \cup I$, $\#O = \#I = n$, and edges going from I to O, they ask whether or not there exists a deterministic algorithm in polynomial time in n for the following tasks:

(1) decide whether there exists a perfect matching in this graph

(2) count how many perfect matchings there are.

A perfect matching is a set of n edges such that each vertex appears once. It turn out that the answer is yes for (1) and most probably no for (2) in a precise sense, that is, (2) is an example of a # P-complete problem. See e.g., [72].

Using previous work of Broder, Mark Jerrum and Alistair Sinclair were able to produce a stochastic algorithm which approximate the number of matchings in polynomial time (for a large class of graphs). The main step of their proof consists in studying a finite Markov chain on perfect and near perfect matchings. They need to show that this chain converges to stationarity in polynomial time. They introduce paths and their combinatorics as a tool to solve this problem. See [54, 72]. This technique will be discussed in detail in these notes.

Computer scientists have a host of problems of this type, including the celebrated problem of approximating the volume of a convex set in high dimension. See [38, 39, 56, 60].

To conclude this section I would like to emphasize that although the present notes only contain theoretical results these results are motivated by the question obviously relevant to applied works:

How many steps are needed for a given finite Markov chain to be close to equilibrium?

1.1.3 A simple open problem

I would like to finish this introduction with a simple example of a family of Markov chains for which the asymptotic theory is trivial but satisfactory quantitative results are still lacking. This example was pointed out to me by M. Jerrum.

Start with the hypercube $\mathcal{X} = \{0,1\}^n$ endowed with its natural graph structure where x and y are neighbors if and only if they differ at exactly one coordinate, that is, $|x-y| = \sum |x_i - y_i| = 1$. The simple random walk on this graph can be analysed by commutative Fourier analysis on the group $\{0,1\}^n$ (or otherwise). The corresponding Markov operator has eigenvalues $1 - 2j/n$, $j = 0, 1, \dots, n$, each with multiplicity $\binom{n}{j}$. It can be shown that this walk reaches approximate equilibrium after $\frac{1}{4} n \log n$ many steps in a precise sense.

Now, fix a sequence $\mathbf{a} = (a_i)_1^n$ of non-negative numbers and $b > 0$. Consider

$$\mathcal{X}(\mathbf{a}, b) = \left\{ x \in \{0,1\}^n : \sum a_i x_i \le b \right\}.$$

This is the hypercube chopped by a hyperplane. Consider the chain $K = K_{\mathbf{a},b}$ on this set defined by $K(x,y) = 1/n$ if $|x-y| = 1$, $K(x,y) = 0$ if $|x-y| > 1$ and $K(x,x) = 1 - n(x)/n$ where $n(x) = n_{\mathbf{a},b}(x)$ is the number of y in $\mathcal{X}(\mathbf{a}, b)$ such that $|x-y| = 1$. This chain has the uniform distribution on $\mathcal{X}(\mathbf{a}, b)$ as stationary measure.

At this writing it is an open problem to prove that this chain is close to stationarity after $n^{O(1)}$ many steps, uniformly over all choices of $\mathbf{a}, b$. A partial result when the set $\mathcal{X}(\mathbf{a}, b)$ is large enough will be described in these notes. See also [38].

1.2 The Perron-Frobenius Theorem

One possible approach for studying finite Markov chains is to reduce everything to manipulations of finite-dimensional matrices. Kemeny and Snell [57] is a useful reference written in this spirit. From this point of view, the most basic result concerning the asymptotic behavior of finite Markov chains is a theorem in linear algebra, namely the celebrated Perron-Frobenius theorem.

1.2.1 Two proofs of the Perron-Frobenius theorem

A *stochastic matrix* is a square matrix with nonnegative entries whose rows all sum to 1.

Theorem 1.2.1 *Let M be an n-dimensional stochastic matrix. Assume that there exists k such that M^k has all its entries positive. Then there exists a row vector $m = (m_j)_1^n$ with positive entries summing to 1 such that for each $1 \le i \le n$,*

$$\lim_{\ell \to \infty} M_{i,j}^{\ell} = m_j. \tag{1.2.1}$$

Furthermore, $m = (m_i)_1^n$ is the unique row vector such that $\sum_1^n m_i = 1$ and $mM = m$.

We start with the following Lemma.

Lemma 1.2.2 *Let M be an n-dimensional stochastic matrix. Assume that for each pair $(i,j), 1 \le i,j \le n$ there exists $k = k(i,j)$ such that $M_{i,j}^k > 0$. Then there exists a unique row vector $m = (m_j)_1^n$ with positive entries summing to 1 such that $mM = m$. Furthermore, 1 is a simple root of the characteristic polynomial of M.*

PROOF: By hypothesis, the column vector $\mathbf{1}$ with all entries equal to 1 satisfies $M\mathbf{1} = \mathbf{1}$. By linear algebra, the transpose M^t of M also has 1 as an eigenvalue, i.e., there exists a row vector v such that $vM = v$. We claim that $|v|$ also satisfies $|v|M = |v|$. Indeed, we have $\sum_i |v_i| M_{i,j} \ge |v_j|$. If $|v|M \ne |v|$, there exists j_0 such that $\sum_i |v_i| M_{i,j_0} > |v_{j_0}|$. Hence, $\sum_i |v_i| = \sum_j \sum_i |v_i| M_{i,j} > \sum_j |v_j|$, a contradiction. Set $m_j = v_j / (\sum_i |v_i|)$. The weak irreducibility hypothesis in the lemma suffices to insure that there exists ℓ such that $A = (I+M)^\ell$ has all its entries positive. Now, $mA = 2^\ell m$ implies that m has positive entries.

Let u be such that $uM = u$. Since $|u|$ is also an eigenvector its follows that the vector u^+ with entries $u_i^+ = \max\{u_i, 0\}$ is either trivial or an eigenvector. Hence, u^+ is either trivial or equal to u (because it must have positive entries). We thus obtain that each vector $u \ne 0$ satisfying $uM = u$ has entries that are either all positive or all negative. Now, if m, m' are two normalized eigenvectors with positive entries then $m - m'$ is either trivial or an eigenvector. If $m - m'$ is not trivial its entries must change sign, a contradiction. So, in fact, $m = m'$.

To see that 1 has geometric multiplicity one, let V be the space of column vectors. The subspace $V_0 = \{v : \sum_i v_i = 0\}$ is stable under M: $MV_0 \subset V_0$ and $V = \mathbb{R}\mathbf{1} \oplus V_0$. So either $M - I$ is invertible on V_0 or there is a $0 \ne v \in V_0$ such that $Mv = v$. The second possibility must be ruled out because we have shown that the entries of such a v would have constant sign. This ends the proof of Lemma 1.2.2. We now complete the proof of Theorem 1.2.1 in two different ways.

PROOF (1) OF THEOREM 1.2.1: Using the strong irreducibility hypothesis of the theorem, let k be such that $\forall\, i,j \;\; M^k_{i,j} > 0$. Let $m = (m_i)^n_1$ be the row vector constructed above and set $M^\infty_{i,j} = m_j$ so that M^∞ is the matrix with all rows equal to m. Observe that

$$MM^\infty = M^\infty M = M^\infty \tag{1.2.2}$$

and that $M^k_{i,j} \geq cM^\infty_{i,j}$ with $c = \min_{i,j}\{M^\infty_{i,j}/M^k_{i,j}\} > 0$. Consider the matrix

$$N = \frac{1}{1-c}\left(M^k - cM^\infty\right)$$

with the convention that $N = 0$ if $c = 1$ (in which case we must indeed have $M^k = M^\infty$). If $0 < c < 1$, N is a stochastic matrix and $NM^\infty = M^\infty N = M^\infty$. In all cases, the entries of $(N - M^\infty)^\ell = N^\ell - M^\infty$ are bounded by 1, in absolute value, for all $\ell = 1, 2, \ldots$. Furthermore

$$\begin{aligned} M^k - M^\infty &= (1-c)(N - M^\infty) \\ M^{k\ell} - M^\infty &= (M^k - M^\infty)^\ell = (1-c)^\ell (N - M^\infty)^\ell. \end{aligned}$$

Thus

$$|M^{k\ell}_{i,j} - M^\infty_{i,j}| \leq (1-c)^\ell.$$

Consider the norm $\|A\|_\infty = \max_{i,j} |A_{i,j}|$ on matrices. The function

$$\ell \to \|M^\ell - M^\infty\|_\infty$$

is nonincreasing because $M^{\ell+1} - M^\infty = M(M^\ell - M^\infty)$ implies

$$\begin{aligned} (M^{\ell+1} - M^\infty)_{i,j} &= \sum_s M_{i,s}(M^\ell - M^\infty)_{s,j} \\ &\leq \left(\sum_s M_{i,s}\right) \|M^\ell - M^\infty\|_\infty = \|M^\ell - M^\infty\|_\infty. \end{aligned}$$

Hence,

$$\max_{i,j}\left\{|M^\ell_{i,j} - m_j|\right\} \leq (1-c)^{\lfloor \ell/k \rfloor}.$$

In particular $\lim_{\ell\to\infty} M^\ell_{i,j} = m_j$. This argument is pushed further in Section 1.2.3 below.

PROOF (2) OF THEOREM 1.2.1: For any square matrix let

$$\rho(A) = \max\{|\lambda| : \lambda \text{ an eigenvalue of } A\}.$$

Observe that any norm $\|\cdot\|$ on matrices that is submultiplicative (i.e., $\|AB\| \leq \|A\|\|B\|$) must satisfy $\rho(A) \leq \|A\|$.

Lemma 1.2.3 *For any square matrix A and any $\epsilon > 0$ there exists a submultiplicative matrix norm $\|\cdot\|$ such that $\|A\| \leq \rho(A) + \epsilon$.*

PROOF: Let U be a unitary matrix such that $A' = UAU^*$ with A' upper-triangular. Let $D = D(t)$, $t > 0$, be the diagonal matrix with $D_{i,i} = t^i$. Then $A'' = DA'D^{-1}$ is upper-triangular with $A''_{i,j} = t^{-(j-i)}A'_{i,j}$, $j \geq i$. Note that, by construction, the diagonal entries are the eigenvalues of A. Consider the matrix norm (induced by the vector norm $\|v\|_1 = \sum |v_i|$)

$$\|B\|_1 = \max_j \sum_i |B_{i,j}|.$$

Then $\|A''\|_1 = \rho(A) + O(t^{-1})$. Pick $t > 0$ large enough so that $\|A''\|_1 \leq \rho + \epsilon$. For U, D fixed as above, define a matrix norm by setting, for any matrix B,

$$\|B\| = \|DUBU^*D^{-1}\|_1 = \|(UD)B(DU)^{-1})\|_1.$$

This norm satisfies the conclusion of the lemma (observe that it depends very much on A and ϵ).

Lemma 1.2.4 *We have* $\lim_{\ell\to\infty} \max_{i,j} A^\ell_{i,j} = 0$ *if and only if* $\rho(A) < 1$.

For each $\epsilon > 0$, the submultiplicative norm of Lemma 1.2.3 satisfies

$$\|A\| \leq \rho(A) + \epsilon.$$

If $\rho(A) < 1$, then we can pick $\epsilon > 0$ so that $\|A\| < 1$. Then $\lim_{\ell\to\infty} \|A^\ell\| \leq \lim_{\ell\to\infty} \|A\|^k = 0$. The desired conclusion follows from the fact that all norms on a finite dimensional vector space are equivalent. Conversely, if

$$\lim_{\ell\to\infty} \left(\max_{i,j} A^\ell_{i,j}\right) = 0$$

then $\lim_{\ell\to\infty} \|A^\ell\|_1 = 0$. Since $\|\cdot\|_1$ is multiplicative, $\rho(A) \leq \|A^\ell\|_1^{1/\ell} < 1$ for ℓ large enough.

Let us pause here to see how the above argument translates in quantitative terms. Let $\|A\|_\infty = \max_{i,j} |A_{i,j}|$ and $|||A|||^2 = \sum_{i,j} |A_{i,j}|^2$. We want to bound $\|A^\ell\|_\infty$ in terms of the norm $\|A^\ell\|$ of Lemma 1.2.3.

Lemma 1.2.5 *For any* $n \times n$ *matrix* A *and any* $\epsilon > 0$, *we can choose the norm* $\|\cdot\|$ *of Lemma* 1.2.3 *so that*

$$\|A^\ell\|_\infty \leq n^{1/2}(1 + |||A|||/\epsilon)^n \|A^\ell\|.$$

PROOF: With the notation of the proof of Lemma 1.2.3, we have

$$\begin{aligned}
|A'_{i,j}| &= \sum_{s,t} U_{i,s}A_{s,t}\overline{U}_{j,t} \\
&\leq \left(\sum_{s,t} |A_{s,t}|^2\right)^{1/2} \left(\sum_{s,t} |U_{i,s}|^2|U_{j,t}|^2\right)^{1/2} \\
&\leq \left(\sum_{s,t} |A_{s,t}|^2\right)^{1/2} \leq |||A|||
\end{aligned}$$

because U is unitary. It follows that

$$\sum_i |A''_{i,j}| \leq \rho(A) + |\!|\!|A|\!|\!|(t-1)^{-1}.$$

Hence, for $t = 1 + |\!|\!|A|\!|\!|/\epsilon$, we get

$$\|A\| = \|A''\|_1 \leq \rho(A) + \epsilon$$

as desired. Now, for any ℓ, set $B = A^\ell, B' = (A')^\ell, B'' = (A'')^\ell$. Then $\|A^\ell\| = \|B''\|_1$ and $A^\ell = U^*B'U = U^*D^{-1}B''DU$. The matrix $B' = D^{-1}B''D$ is upper-triangular with coefficients $B'_{i,j} = t^{j-i}B''_{i,j}$ for $j \geq i$. This yields

$$\begin{aligned} \|A^\ell\|_\infty &\leq \left(\sum_{\substack{i,j:\\ i\leq j}} t^{2(j-i)}|B''_{i,j}|^2\right)^{1/2} \\ &\leq n^{1/2}(1+|\!|\!|A|\!|\!|/\epsilon)^n\|B''\|_1 \\ &= n^{1/2}(1+|\!|\!|A|\!|\!|/\epsilon)^n\|A^\ell\|. \end{aligned}$$

With this material at hand the following lemma suffices to finish the second proof or the Perron-Frobenius theorem.

Lemma 1.2.6 *Let M be a stochastic matrix satisfying the strong irreducibility condition of Theorem* 1.2.1. *Let $M^\infty_{i,j} = m_j$ where $m = (m_j)$ is the unique normalized row vector with positive entries such that $mM = m$. Then $\rho(M - M^\infty) < 1$.*

PROOF: Let λ be an eigenvalue of M with left eigenvector v. Assume that $|\lambda| = 1$. Then, again, $|v|$ is a left eigenvector with eigenvalue 1. Let k be such that $M^k > 0$. It follows that

$$|\sum_j M^k_{i,j}v_j| = \sum_j M^k_{i,j}|v_j|.$$

Since $M^k_{i,j} > 0$ for all j, this implies that $v_j = e^{i\theta}|v_j|$ for some fixed θ. Hence $\lambda = 1$. Let $\lambda_1 = 1$ and λ_i, $i = 2, \ldots, n$ be the eigenvalues of M repeated according to there geometric multiplicities. By Lemma 1.2.2, $|\lambda_i| < 1$ for $i = 2, \ldots, n$. The eigenvalues of M^∞ are 1 with eigenspace $\mathbb{R}1$ and 0 with eigenspace $V_0 = \{v : \sum_i v_i = 0\}$. By (1.2.2) it follows that the eigenvalues of $M - M^\infty$ are $0 = \lambda_1 - 1$ and $\lambda_i = \lambda_i - 0$, $i = 2, \ldots, n$. Hence $\rho(M - M^\infty) < 1$.

1.2.2 Comments on the Perron-Frobenius theorem

Each of the two proofs of Theorem 1.2.1 outlined above provides existence of $A > 0$ and $0 < \epsilon < 1$ such that

$$|M^\ell_{i,j} - m_j| \leq A(1-\epsilon)^\ell. \tag{1.2.3}$$

However, it is rather dishonest to state the conclusion (1.2.1) in this form without a clear WARNING:

$$\boxed{\text{the proof does not give a clue on how large } A \text{ and how small } \epsilon \text{ can be.}}$$

Indeed, "Proof (1)" looks like a quantitative proof since it shows that

$$|M^{\ell}_{i,j} - m_j| \leq (1-c)^{\lfloor \ell/k \rfloor} \tag{1.2.4}$$

whenever $M^k \geq cM^{\infty}$. But, in general, it is hard to find explicit reasonable k and c such that the condition $M^k \geq cM^{\infty}$ is satisfied.

EXAMPLE 1.2.1: Consider the random walk on $\mathbb{Z}/n\mathbb{Z}, n = 2p+1$, where, at each step, we add 1 or substract 1 or do nothing each with probability 1/3. Then M is an $n \times n$ matrix with $M_{i,j} = 1/3$ if $|i-j| = 0, 1$, $M_{1,n} = M_{n,1} = 1/3$, and all the orther entries equal to zero. The matrix M^{∞} has all its entries equal to $1/n$. Obviously, $M^p \geq n\,3^{-p} M^{\infty}$, hence $|M^{\ell}_{i,j} - (1/n)| \leq 2(1-n3^{-p})^{\lfloor \ell/p \rfloor}$. This is a very poor estimate. It is quite typical of what can be obtained by using (1.2.4).

Still, there is an interesting conclusion to be drawn from (1.2.4). Let

$$k_0 = \inf\{\ell : M^{\ell} \geq (1-1/e)M^{\infty}\}$$

where the constant $c = 1 - 1/e$ as been chosen for convenience. This k_0 can be interpreted as a measure of how long is takes for the chain to be close to equilibrium in a crude sense. Then (1.2.4) says that this crude estimate suffices to obtain the exponential decay with rate $1/k_0$

$$|M^{\ell}_{i,j} - m_j| \leq 3e^{-\ell/k_0}$$

"Proof (2)" has the important theoretical advantage of indicating what is the best exponential rate in (1.2.3). Namely, for any norm $\|\cdot\|$ on matrices, we have

$$\lim \|M^{\ell} - M^{\infty}\|^{1/\ell} = \rho \tag{1.2.5}$$

where

$$\rho = \rho(M - M^{\infty}) = \max\{|\lambda| : \lambda \neq 1, \lambda \text{ an eigenvalue of } M\}.$$

Comparing with (1.2.4) we discover that $M^k \geq cM^{\infty}$ implies

$$\rho \leq \frac{1}{k}\log(1-c).$$

Of course (1.2.5) shows that, for all $\epsilon > 0$, there exists $C(\epsilon)$ such that

$$|M^{\ell}_{i,j} - m_j| \leq C(\epsilon)\,(\rho + \epsilon)^{\ell}.$$

The constant $C(\epsilon)$ can be large and is dificult to bound. Since $\|M^\ell - M^\infty\| \leq 2n^{1/2}$ (in the notation of the proof of Lemma 1.2.5), Lemma 1.2.5 yields

$$|M^\ell_{i,j} - m_j| \leq n^{1/2}\left(1 + \frac{2n^{1/2}}{\epsilon}\right)^n (\rho+\epsilon)^\ell. \tag{1.2.6}$$

This is quantitative, but essentially useless. I am not sure what is the best possible universal estimate of this sort but I find the next example quite convincing in showing that "Proof (2)" is not satisfactory from a quantitative point of view.

EXAMPLE 1.2.2: Let $\mathcal{X} = \{0,1\}^n$. Define a Markov chain with state space $\mathcal{X}$ as follows. If the current state is $x = (x_1, \ldots, x_n)$ then move to $y = (y_1, \ldots, y_n)$ where $y_i = x_{i+1}$ for $i = 1, \ldots, n-1$ and $y_n = x_1$ or $y_n = x_1 + 1 \pmod 2$, each with equal probability $1/2$. It is not hard to verify that this chain is irreducible. Let M denote the matrix of this chain for some ordering of the state space. Then the left normalized eigenvector m with eigenvalue 1 is the constant vector with $m_i = 2^{-n}$. Furthermore, a moment of thought shows that $M^n = M^\infty$. Hence $\rho = \rho(M - M^\infty) = 0$. Now, $\max_{i,j} |M^{n-1}_{i,j} - m_j|$ is of order 2^{-n}. So, in this case, $C(\varepsilon)$ of order $(2\epsilon)^{-n}$ is certainly needed for the inequality $|M^\ell_{i,j} - m_j| \leq C(\epsilon)\,(\rho+\epsilon)^\ell$ to be satisfied for all ℓ.

1.2.3 Further remarks on strong irreducibility

A n-dimensional stochastic matrix M is strongly irreducible if there exists an integer k such that, for all i, j, $M^k_{i,j} > 0$. This is related to what is known as the **Doeblin** condition. Say that M satisfies the Doeblin condition if there exist an integer k, a positive c, and a probability measure q on $\{1, \ldots, n\}$ such that

(D) $\qquad$ for all $i \in \{1, \ldots, n\}$, $\quad M^k_{i,j} \geq c q_j$.

Proof (1) of Theorem 1.2.1 is based on the fact that strong irreducibility implies the Doeblin condition (**D**) with $q = m$ (the stationary measure) and some $k, c > 0$. The argument developed in this case yields the following well known result.

Theorem 1.2.7 *If M satisfies* (**D**) *for some $k, c > 0$ and a some probability q then*

$$\sum_j |M^\ell_{i,j} - m_j| \leq 2(1-c)^{\lfloor \ell/k \rfloor}$$

for all integer ℓ. Here $m = (m_j)_1^n$ is the vector appearing in Lemma 1.2.2, i.e., the stationary measure of M.

PROOF: Using (1.2.1), observe that (**D**) implies $m_j \geq c q_j$. Let M^∞ be the matrix with all rows equal to m, let Q be the matrix with all rows equal to q and set

$$N = \frac{1}{1-c}\left(M^k - cQ\right), \quad N^\infty = \frac{1}{1-c}\left(M^\infty - cQ\right).$$

These two matrices are stochatic. Furthermore

$$M^k - M^\infty = (1-c)(N - N^\infty)$$

and

$$\begin{aligned} M^{k\ell} - M^\infty &= \left(M^k - M^\infty\right)^\ell \\ &= (1-c)^\ell (N - N^\infty)^\ell. \end{aligned}$$

Observe that $(N - N^\infty)^2 = (N - N^\infty)N$ because N^∞ has constant columns so that $PN^\infty = N^\infty$ for any stochastic matrix P. It follows that $(N - N^\infty)^\ell = (N - N^\infty)N^{\ell-1}$. If we set $\|A\|_1 = \max_i \sum_j |A_{i,j}|$ for any matrix A and recall that $\|AB\|_1 \leq \|A\|_1 \|B\|_1$ we get

$$\|M^{k\ell} - M^\infty\|_1 \leq (1-c)^\ell \|N - N^\infty\|_1 \|N^{\ell-1}\|_1.$$

Since N is stochastic, we have $\|N\|_1 = 1$. Also $\|N - N^\infty\|_1 \leq 2$. Hence

$$\max_i \sum_j |M^{k\ell} - M^\infty| \leq 2(1-c)^\ell.$$

This implies the stated result because $\ell \to \|M^\ell - M^\infty\|_1$ is nonincreasing.

This section introduces notation and concepts from elementary functional analysis such as operator norms, interpolation, and duality. This tools turn out to be extremely useful in manipulating finite Markov chains.

1.2.4 Operator norms

Let A, B be two Banach spaces with norms $\|\cdot\|_A$, $\|\cdot\|_B$. Let $K : A \to B$ be a linear operator. We set

$$\|K\|_{A\to B} = \sup_{\substack{f\in A:\\ \|f\|_A \leq 1}} \{\|Kf\|_B\} = \sup_{f\in A: f\neq 0} \left\{\frac{\|Kf\|_B}{\|f\|_A}\right\}.$$

If A^*, B^* are the (topological) duals of A, B, the dual operator $K^* : B^* \to A^*$ defined by $K^*b^*(a) = b^*(Ka), a \in A$, satisfies

$$\|K^*\|_{B^*\to A^*} \leq \|K\|_{A\to B}.$$

In particular, if $\mathcal{X}$ is a countable set equipped with a positive measure π and if $A = \ell^p(\pi)$ and $B = \ell^q(\pi)$ with

$$\|f\|_p = \|f\|_{\ell^p(\pi)} = \left(\sum_{x\in\mathcal{X}} |f(x)|^p \pi(x)\right)^{1/p} \quad \text{and} \quad \|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)|,$$

we write

$$\|K\|_{p\to q} = \|K\|_{\ell^p(\pi)\to\ell^q(\pi)}.$$

Let

$$\langle f, g\rangle = \langle f, g\rangle_\pi = \sum_{\mathcal{X}} f(x)\overline{g(x)}\pi(x)$$

be the scalar product on $\ell^2(\pi)$. For $1 \le p < \infty$, this scalar product can be used to identify $\ell^p(\pi)^*$ with $\ell^q(\pi)$ where p, q are Hölder conjugate exponents, that is $1/p + 1/q = 1$. Furthermore, for all $1 \le p \le \infty$, $\ell^q(\pi)$ norms $\ell^p(\pi)$. Namely,

$$\|f\|_p = \sup_{\substack{g \in \ell^q(\pi) \\ \|g\|_q \le 1}} \langle f, g\rangle_\pi.$$

It follows that for any linear operator $K : \ell^p(\pi) \to \ell^r(\pi)$ with $1 \le p, r \le +\infty$,

$$\|K\|_{p\to r} = \|K^*\|_{s\to q}$$

where $1/p + 1/q = 1$, $1/r + 1/s = 1$. Assume now that the operator K is defined by

$$Kf(x) = \sum_{y \in \mathcal{X}} K(x, y) f(y)$$

for any finitely supported function f. Then the norm $\|K\|_{p\to\infty}$ is given by

$$\|K\|_{p\to\infty} = \max_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{X}} |K(x, y)/\pi(y)|^q \pi(y) \right)^{1/q} \tag{1.2.7}$$

where $1/p + 1/q = 1$. In particular,

$$\|K\|_{2\to\infty} = \|K^*\|_{1\to 2} = \max_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{X}} |K(x, y)/\pi(y)|^2 \pi(y) \right)^{1/2} \tag{1.2.8}$$

and

$$\|K\|_{1\to\infty} = \|K^*\|_{1\to\infty} = \max_{x, y \in \mathcal{X}} \{|K(x, y)/\pi(y)|\}. \tag{1.2.9}$$

For future reference we now recall the Riesz-Thorin interpolation theorem (complex method). It is a basic tools in modern analysis. See, e.g., Theorem 1.3, page 179 in [73].

Theorem 1.2.8 *Fix $1 \le p_i, q_i \le \infty$, $i = 1, 2$, with $p_1 \le p_2$, $q_1 \le q_2$. Let K be a linear operator acting on functions by $Kf(x) = \sum_y K(x, y) f(y)$. For any p such that $p_1 \le p \le p_2$ let θ be such that $1/p = \theta/p_1 + (1-\theta)/p_2$ and define $q \in [q_1, q_2]$ by $1/q = \theta/q_1 + (1-\theta)/q_2$. Then*

$$\|K\|_{p\to q} \le \|K\|_{p_1\to q_1}^{\theta} \|K\|_{p_2\to q_2}^{1-\theta}.$$

1.2.5 Hilbert space techniques

For simplicity we assume now that $\mathcal{X}$ is finite of cardinality $n = |\mathcal{X}|$ and work on the (n-dimensional) Hilbert space $\ell^2(\pi)$. An operator $K : \ell^2(\pi) \to \ell^2(\pi)$ is self-adjoint if it satisfies

$$\langle Kf, g\rangle_\pi = \langle f, Kg\rangle_\pi, \quad \text{i.e.,} \quad K^* = K.$$

Let $K(x,y)$ be the kernel of the operator K. Then K^* has kernel

$$K^*(x,y) = \pi(y)\overline{K(y,x)}/\pi(x)$$

and it follows that K is selfadjoint if and only if

$$K(x,y) = \pi(y)\overline{K(y,x)}/\pi(x).$$

Lemma 1.2.9 *Assume that K is self-adjoint on $\ell^2(\pi)$. Then K is diagonalizable in an orthonormal basis of $\ell^2(\pi)$ and has real eigenvalues $\beta_0 \geq \beta_1 \ldots \geq \beta_{n-1}$. For any associated orthonormal basis $(\psi_i)_0^{n-1}$ of eigenfunctions, we have*

$$K(x,y)/\pi(y) = \sum_i \beta_i \psi_i(x)\overline{\psi_i(y)}. \tag{1.2.10}$$

$$\|K(x,\cdot)/\pi(\cdot)\|_2^2 = \sum_i \beta_i^2 |\psi_i(x)|^2. \tag{1.2.11}$$

$$\sum_{x\in\mathcal{X}} \|K(x,\cdot)/\pi(\cdot)\|_2^2 \pi(x) = \sum_i \beta_i^2. \tag{1.2.12}$$

PROOF: We only prove the set of equalities. Let $z \to \mathbf{1}_x(z)$ be the function which is equal to 1 at x and zero everywhere else. Then $K(x,y) = K\mathbf{1}_y(x)$. The function $\mathbf{1}_y$ has coordinates $\langle \mathbf{1}_y, \psi_i\rangle_\pi = \overline{\psi_i(y)}\pi(y)$ in the orthonormal basis $(\psi_i)_0^{n-1}$. Hence $K\mathbf{1}_y(x) = \pi(y)\sum_i \beta_i\psi_i(x)\overline{\psi_i(y)}$. The second and third results follow by using the fact that $(\psi_i)_0^{n-1}$ is orthonormal.

We now turn to an important tool known as the Courant-Fischer min-max theorem. Let $\mathcal{E}$ be a (positive) Hermitian form on $\ell^2(\pi)$. For any vector space $W \subset \ell^2(\pi)$, set

$$M(W) = \max_{\substack{f\in W\\ f\neq 0}} \left\{\frac{\mathcal{E}(f,f)}{\|f\|_2^2}\right\}, \quad m(W) = \min_{f\in W} \left\{\frac{\mathcal{E}(f,f)}{\|f\|_2^2}\right\}.$$

Recall from linear algebra that there exists a unique Hermitian matrice A such that $\mathcal{E}(f,f) = \langle Af, f\rangle_\pi$ and that, by definition, the eigenvalues of $\mathcal{E}$ are the eigenvalues of A. Furthermore, these are real.

Theorem 1.2.10 *Let $\mathcal{E}$ be a quadratic form on $\ell^2(\pi)$, with eigenvalues*

$$\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}.$$

Then

$$\lambda_k = \min_{\substack{W\subset\ell^2(\pi):\\ \dim(W)\geq k+1}} M(W) = \max_{\substack{W\subset\ell^2(\pi):\\ \dim(W^\perp)\leq k}} m(W). \tag{1.2.13}$$

For a proof, see [51], page 179-180. Clearly, the minimum of $M(W)$ with $\dim(W) \geq k+1$ is obtained when W is the linear space spanned by the $k+1$ first eigenvectors ψ_i associated with λ_i, $i = 0, \ldots, k$. Similarly, the maximum of $m(W)$ with $\dim(W^{\perp}) \leq k$ is attained when W is spanned by the ψ_i's, $i = k, \ldots, n$. This result also holds in infinite dimension. It has the following corollary.

Theorem 1.2.11 *Let $\mathcal{E}, \mathcal{E}'$ be two quadratic forms on different Hilbert spaces $\mathcal{H}$, $\mathcal{H}'$ of dimension $n \leq n'$. Assume that there exists a linear map $f \to \tilde{f}$ from $\mathcal{H}$ into $\mathcal{H}'$ such that, for all $f \in \mathcal{H}$,*

$$\mathcal{E}'(\tilde{f}, \tilde{f}) \leq A\mathcal{E}(f, f) \quad \text{and} \quad a\|f\|_{\mathcal{H}}^2 \leq \|\tilde{f}\|_{\mathcal{H}'}^2 \tag{1.2.14}$$

for some constants $0 < a, A < \infty$. Then

$$\frac{a}{A}\,\lambda'_\ell \leq \lambda_\ell \quad \text{for} \quad \ell = 1, \ldots, n-1. \tag{1.2.15}$$

PROOF: Fix $\ell = 0, 1, \ldots, n-1$ and let ψ_i be orthonormal eigenvectors associated to λ_i, $i = 0, \ldots, n-1$. Observe that the second condition in (1.2.14) implies that $f \to \tilde{f}$ is one to one. Let $W \subset \mathcal{H}$ be the vector space spanned by $(\psi_i)_0^{\ell-1}$, and let $\widetilde{W} \subset \mathcal{H}'$ be its image under the one to one map $f \to \tilde{f}$. Then $\widetilde{W}$ has dimension ℓ and by (3.7)

$$\begin{aligned}
\lambda'_\ell \;&\leq\; M(\widetilde{W}) = \max_{f \in W} \left\{ \frac{\mathcal{E}'(\tilde{f}, \tilde{f})}{\|\tilde{f}\|_{\mathcal{H}'}^2} \right\} \\
&\leq\; \max_{f \in W} \left\{ \frac{A\mathcal{E}(f, f)}{a\|f\|_{\mathcal{H}}^2} \right\} \geq \frac{A\lambda_\ell}{a}.
\end{aligned}$$

1.3 Notation for finite Markov chains

Let $\mathcal{X}$ be a finite space of cardinality $|\mathcal{X}| = n$. Let $K(x, y)$ be a Markov kernel on $\mathcal{X}$ with associated Markov operator defined by

$$Kf(x) = \sum_{y \in \mathcal{X}} K(x, y) f(y).$$

That is, we assume that

$$K(x, y) \geq 0 \quad \text{and} \quad \sum_y K(x, y) = 1.$$

The operator K^ℓ has a kernel $K^\ell(x, y)$ which satisfies

$$K^\ell(x, y) = \sum_{z \in \mathcal{X}} K^{\ell-1}(x, z) K(z, y).$$

Properly speaking, the Markov chain with initial distribution q associated with K is the sequence of $\mathcal{X}$-valued random variables $(X_n)_0^\infty$ whose law $\mathbf{P}_q$ is determined by

$$\forall \ell = 1, 2, \dots, \quad \mathbf{P}_q(X_i = x_i, 1 \le i \le \ell) = q(x_0)K(x_0, x_1)\cdots K(x_{\ell-1}, x_\ell).$$

With this notation the probability measure $K^\ell(x, \cdot)$ is the law of X_ℓ for the Markov chain started at x:

$$\mathbf{P}_x(X_\ell = y) = K^\ell(x, y).$$

However, this language will almost never be used in these notes.

The continuous time semigroup associated with K is defined by

$$H_t f(x) = e^{-t(I-K)} = e^{-t} \sum_0^\infty \frac{t^i K^i f}{i!}. \tag{1.3.1}$$

Obviously, it has kernel

$$H_t(x, y) = e^{-t} \sum_0^\infty \frac{t^i K^i(x, y)}{i!}.$$

Observe that this is indeed a semigroup of operators, that is,

$$\begin{aligned} H_{t+s} &= H_t H_s \\ \lim_{t \to 0} H_t &= I. \end{aligned}$$

Furthermore, for any f, the function $u(t, x) = H_t f(x)$ solves

$$\begin{cases} (\partial_t + (I - K))\, u(t, x) = 0 \text{ on } (0, \infty) \times \mathcal{X} \\ u(0, x) = f(x). \end{cases}$$

Set $H_t^x(y) = H_t(x, y)$. Then $H_t^x(\cdot)$ is a probability measure on $\mathcal{X}$ which represents the distribution a time t of the continuous Markov chain $(X_t)_{t>0}$ associated with K and started at x. This process can be described as follows. The moves are those of the discrete time Markov chain with transition kernel K started at x, but the jumps occur after independent Poison(1) waiting times. Thus, the probability that there have been exactly i jumps at time t is $e^{-t} t^i/i!$ and the probability to be at y after exactly i jumps at time t is $e^{-t} t^i K^i(x, y)/i!$.

The operators K, H_t also acts on measures. If μ is a measure then μK (resp. μH_t) is defined by setting

$$\mu K(f) = \mu(Kf) \quad (\text{resp. } \mu H_t(f) = \mu(H_t f))$$

for all functions f. Thus

$$\mu K(x) = \sum_y \mu(y) K(y, x).$$

Definition 1.3.1 *A Markov kernel K on a finite set $\mathcal{X}$ is said to be irreducible if for any x, y there exists $j = j(x, y)$ such that $K^j(x, y) > 0$.*

Assume that K is irreducible and let π be the unique stationary measure for K, that is, the unique probability measure satisfying $\pi K = \pi$ (see Lemma 1.2.2). We will use the notation

$$\pi(f) = \sum_x f(x)\pi(x) \quad \text{and} \quad \mathrm{Var}_\pi(f) = \sum_x |f(x) - \pi(f)|^2 \pi(x).$$

We also set

$$\pi_* = \min_{x \in \mathcal{X}} \{\pi(x)\}. \tag{1.3.2}$$

Throughout these notes we will work with the Hilbert space $\ell^2(\pi)$ with scalar product

$$\langle f, g \rangle = \sum_{x \in \mathcal{X}} f(x)\overline{g(x)}\pi(x),$$

and with the space $\ell^p(\pi)$, $1 \le p \le \infty$, with norm

$$\|f\|_p = \left(\sum_{x \in \mathcal{X}} |f(x)|^p \pi(x) \right)^{1/p}, \quad \|f\|_\infty = \max_{x \in \mathcal{X}} \{|f(x)|\}.$$

In this context, it is natural and useful to consider the densities of the probability measures K_x^ℓ, H_t^x with respect to π which will be denoted by

$$k_x^\ell(y) = k^\ell(x, y) = \frac{K^\ell(x, y)}{\pi(y)}$$

and

$$h_t^x(y) = h_t(x, y) = \frac{H_t^x(y)}{\pi(y)}.$$

Observe that the semigroup property implies that, for all $t, s > 0$,

$$h_{t+s}(x, y) = \sum_z h_t(x, z) h_s(z, y) \pi(z).$$

The operator K (hence also H_t) is a contraction on each $\ell^p(\pi)$ (i.e., $\|Kf\|_p \le \|f\|_p$). Indeed, by Jensen's inequality, $|Kf(x)|^p \le K(|f|^p)(x)$ and thus

$$\|Kf\|_p^p \le \sum_{x,y} K(x, y)|f(y)|^p \pi(x) = \sum_y |f(y)|^p \pi(y) = \|f\|_p^p.$$

The adjoint K^* of K on $\ell^2(\pi)$ has kernel

$$K^*(x, y) = \pi(y)K(y, x)/\pi(x).$$

Since π is the stationary measure of K, it follows that K^* is a Markov operator. The associated semigroup is $H_t^* = e^{-t(I - K^*)}$ with kernel

$$H_t^*(x, y) = \pi(y)H_t(y, x)/\pi(x)$$

and density

$$h_t^*(x,y) = h_t(y,x).$$

The Markov process associated with H_t^* is the time reversal of the process associated to H_t.

If a measure μ has density f with respect to π, that is, if $\mu(x) = f(x)\pi(x)$, then μK (resp. μH_t) has density K^*f (resp. $H_t^* f$) with respect to π. Thus acting by K (resp. H_t) on a measure is equivalent to acting by K^* (resp H_t^*) on its density with respect to π. In particular, the density $h_t(x,\cdot)$ of the measure H_t^x with respect to π is $H_t^*\delta_x$ where $\delta_x = \mathbf{1}_x/\pi(x)$. Indeed, the measure $\mathbf{1}_x$ has density $\delta_x = \mathbf{1}_x/\pi(x)$ with respect to π. Hence $H_t^x = \mathbf{1}_x H_t$ has density

$$H_t^*\delta_x(y) = \frac{H_t^*(y,x)}{\pi(x)} = h_t^*(y,x) = h_t(x,y)$$

with respect to π.

Recall the following classic definition.

Definition 1.3.2 *A pair (K,π) where K is Markov kernel and π a positive probability measure on $\mathcal{X}$ is reversible if*

$$\pi(x)K(x,y) = \pi(y)K(y,x).$$

This is sometimes called the detailed balance condition.

If (K,π) is reversible then $\pi K = \pi$. Furthermore, (K,π) is reversible if and only if K is self-adjoint on $\ell^2(\pi)$.

1.3.1 Discrete time versus continuous time

These notes are written for continuous time finite Markov chains. The reason of this choice is that it makes life easier from a technical point of view. This will allow us hopefully to stay more focussed on the main ideas. This choice however is not very satisfactory because in some respects (e.g., implementation of algorithms) discrete time chains are more natural. Furthermore, since the continuous time chain is obtained as a function of the discrete time chain through the formula $H_t = e^{-t(I-K)}$ it is often straightforward to transfer information from discrete time to continuous time whereas the converse can be more difficult. Thus, let us emphasize that the techniques presented in these lectures are not confined to continuous time and work well in discrete time. Treatments of discrete time chains in the spirit of these notes can be found in [23, 24, 25, 26, 27, 28, 29, 35, 41, 63].

For reversible chains, it is possible to relate precisely the behavior of H_t to that of K^ℓ through eigenvalues and eigenvectors as follows. Assuming that (K,π) is reversible and $|\mathcal{X}| = n$, let $(\lambda_i)_0^{n-1}$ be the eigenvalues of $I-K$ in non-decreasing order and let $(\psi_i)_0^{n-1}$ be an orthonormal basis of $\ell^2(\pi)$ made of real eigenfuntions associated to the eigenvalues $(\lambda_i)_0^{n-1}$ with $\psi_0 \equiv 1$.

Lemma 1.3.3 *If (K,π) is reversible, it satisfies*

(1) $k^\ell(x,y) = \sum_0^{n-1}(1-\lambda_i)^\ell \psi_i(x)\psi_i(y), \quad \|k_x^\ell - 1\|_2^2 = \sum_1^{n-1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2.$

(2) $h_t(x,y) = \sum_0^{n-1} e^{-t\lambda_i}\psi_i(x)\psi_i(y), \quad \|h_t^x - 1\|_2^2 = \sum_1^{n-1} e^{-2t\lambda_i}|\psi_i(x)|^2.$

This classic result follows from Lemma 1.2.9. The next corollary gives a useful way of transferring information between discrete and continuous time. It separates the effects of the largest eigenvalue λ_{n-1} from those of the rest of the spectrum.

Corollary 1.3.4 *Assume that (K,π) is reversible and set $\beta_- = \max\{0, -1 + \lambda_{n-1}\}$. Then*

(1) $\|h_t^x - 1\|_2^2 \le \frac{1}{\pi(x)}e^{-t} + \|k_x^{[t/2]} - 1\|_2^2$.

(2) $\|k_x^N - 1\|_2^2 \le \beta_-^{2m}\left(1 + \|h_\ell^x - 1\|_2^2\right) + \|h_N^x - 1\|_2^2$ *for* $N = m + \ell + 1$.

Proof: For (1), use Lemma 1.3.3,

$$(1-\lambda_i)^{2\ell} = e^{2\ell\log(1-\lambda_i)}$$

and the inequality $\log(1-x) \ge -2x$ for $0 \le x \le 1/2$. For (2), observe that

$$k^{2\ell+1}(x,x) = \sum_0^{n-1}(1-\lambda_i)^{2\ell+1}|\psi_i(x)|^2 \ge 0.$$

This shows that

$$-\sum_{i:\lambda_i>1}(1-\lambda_i)^{2\ell+1}|\psi_i(x)|^2 \le \sum_{i:\lambda_i<1}(1-\lambda_i)^{2\ell+1}|\psi_i(x)|^2.$$

Hence

$$\sum_{i:\lambda_i>1}(1-\lambda_i)^{2\ell+2}|\psi_i(x)|^2 \le \sum_{i:\lambda_i<1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2.$$

Now, for those λ_i that are smaller than 1, we have

$$(1-\lambda_i)^{2\ell} = e^{2\ell\log(1-\lambda_i)} \le e^{-2\ell\lambda_i}$$

so that

$$\sum_{i:\lambda_i<1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2 \le \|h_\ell^x\|_2^2$$

and

$$\sum_{i\ne 0,\lambda_i<1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2 \le \|h_\ell^x - 1\|_2^2.$$

Putting these pieces together, we get for $N = m + \ell + 1$,

$$\begin{aligned}
\|k_x^N - 1\|_2^2 &= \sum_1^{n-1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 \\
&= \sum_{i:\lambda_i > 1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 + \sum_{i \neq 0: \lambda_i < 1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 \\
&\leq \beta_-^{2m} \left(\sum_{i:\lambda_i > 1} (1-\lambda_i)^{2\ell+2} |\psi_i(x)|^2 \right) + \sum_{i \neq 0: \lambda_i < 1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 \\
&\leq \beta_-^{2m} \|h_\ell^x\|_2^2 + \|h_N^x - 1\|_2^2 \\
&= \beta_-^{2m} \left(1 + \|h_\ell^x - 1\|_2^2\right) + \|h_N^x - 1\|_2^2.
\end{aligned}$$

Observe that, according to Corrolary 1.3.4, it is useful to have tools to bound $1 - \lambda_{n-1}$ away from -1.

Corollary 1.3.4 says that the behavior of a discrete time chain and of its associated continuous time chain can not be too different in the reversible case. It is interesting to see that this fails to be satisfied for nonreversible chains.

EXAMPLE 1.3.1: Consider the chain K on $\mathcal{X} = \mathbb{Z}/m\mathbb{Z}$ with $m = n^2$ an odd integer and

$$K(x,y) = \begin{cases} 1/2 & \text{if } y = x+1 \\ 1/2 & \text{if } y = x+n \end{cases}.$$

On one hand, the discrete time chain takes order $m^2 \approx n^4$ steps to be close to stationarity. Indeed, there exists an affine bijection from $\mathcal{X}$ to $\mathcal{X}$ that send 1 to 1 and n to -1. On the other hand, one can show that the associated continuous time process is close to stationarity after a time of order $m = n^2$. See [25].

Lemma 1.3.3 is often hard to use directly because it involves both eigenvalues and eigenvectors. To have a similar statement involving only eigenvalues one has to work with the distance

$$|\!|\!| f - g |\!|\!| = \left(\sum_{x,y} |f(x,y) - g(x,y)|^2 \pi(x)\pi(y) \right)^{1/2}$$

between functions on $\mathcal{X} \times \mathcal{X}$.

Lemma 1.3.5 *If (K,π) is reversible, it satisfies*

$$|\!|\!| k^\ell - 1 |\!|\!|^2 = \sum_1^{n-1} (1-\lambda_i)^{2\ell} \quad \textit{and} \quad |\!|\!| h_t - 1 |\!|\!|^2 = \sum_1^{n-1} e^{-2t\lambda_i}.$$

It is possible to bound $|\!|\!| k^\ell - 1 |\!|\!|$ using only $\beta_* = \max\{1-\lambda_1, -1+\lambda_{n-1}\}$ and the eigenvalues λ_i such that $\lambda_i < 1$. It is natural to state this result in terms of the eigenvalues $\beta_i = 1 - \lambda_i$ of K. Then $\beta_* = \max\{\beta_1, |\beta_{n-1}|\}$ and $\lambda_i < 1$ corresponds to the condition $\beta_i > 0$.

Corollary 1.3.6 *Assume that* (K, π) *is reversible. With the notation introduced above we have, for* $N = m + \ell + 1$,

$$\|k^m - 1\|^2 \leq 2\beta_*^{2\ell} \left(\sum_{i:0<\beta_i\leq 1} \beta_i^{2m} \right).$$

PROOF: We have

$$\sum_{\mathcal{X}} k^{2m+1}(x,x)\pi(x) = \sum_{0}^{n-1} \beta_i^{2m+1} \geq 0.$$

Hence

$$\sum_{\beta_i<0} \beta_i^{2m+2} \leq \sum_{\beta_i>0} \beta_i^{2m}.$$

It follows that

$$\begin{aligned} \|k^N - 1\| &= \sum_{1}^{n-1} \beta_i^{2m+2\ell+2} \\ &\leq \beta_*^{2\ell} \left(\sum_{0}^{n-1} \beta_i^{2m+2} \right) \leq 2\beta_*^{2\ell} \left(\sum_{i:\beta_i>0} \beta_i^{2m} \right). \end{aligned}$$

Chapter 2

Analytic tools

This chapter uses semigroup techniques to obtain quantitative estimates on the convergence of continuous time finite Markov chain in terms of various functional inequalities. The same ideas and techniques apply to discrete time but the details are somewhat more tedious. See [28, 29, 35, 41, 63, 72].

2.1 Nothing but the spectral gap

2.1.1 The Dirichlet form

Classicaly, the notion of *Dirichlet form* is introduced in relation with reversible Markov semigroups. The next definition coincides with the classical notion when (K, π) is reversible.

Definition 2.1.1 *The form*

$$\mathcal{E}(f,g) = \Re(\langle (I-K)f, g\rangle)$$

is called the Dirichlet form associated with $H_t = e^{-t(I-K)}$

The notion of Dirichlet form will be one of our main technical tools.

Lemma 2.1.2 *The Dirichlet form* $\mathcal{E}$ *satisfies* $\mathcal{E}(f,f) = \langle (I - \frac{1}{2}(K+K^*))f, f\rangle$,

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x,y)\pi(x) \tag{2.1.1}$$

and

$$\frac{\partial}{\partial t}\|H_t f\|_2^2 = -2\,\mathcal{E}(H_t f, H_t f). \tag{2.1.2}$$

PROOF: The first equality follows from $\langle Kf, f\rangle = \langle f, K^* f\rangle = \overline{\langle K^* f, f\rangle}$. For the second, observe that $\mathcal{E}(f,f) = \|f\|_2^2 - \Re(\langle Kf, f\rangle)$ and

$$\frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x,y)\pi(x)$$

$$= \frac{1}{2}\sum_{x,y}\left(|f(x)|^2 + |f(y)|^2 - 2\Re(\overline{f(x)}f(y))\right)K(x,y)\pi(x)$$
$$= \|f\|_2^2 - \Re(\langle Kf, f\rangle).$$

The third is calculus. In a sense, (2.1.2) is the definition of $\mathcal{E}$ as the Dirichlet form of the semigroup H_t since

$$\mathcal{E}(f,f) = -\left.\partial_t\|H_t f\|_2^2\right|_{t=0} = -\lim_{t\to 0}\frac{1}{t}\langle (I-H_t)f, f\rangle.$$

Lemma 2.1.2 shows that the Dirichlet forms of H_t, H_t^* and $S_t = e^{-t(I-R)}$ whith $R = \frac{1}{2}(K+K^*)$ are equal. Let us emphasize that equalities (2.1.1) and (2.1.2) are crucial in most developments involving Dirichlet forms. Equality (2.1.1) expresses the Dirichlet form as a sum of positive terms. It will allow us to estimate $\mathcal{E}$ in geometric terms and to compare different Dirichlet forms. Equality (2.1.2) is the key to translating functional inequalities such as Poincaré or logarithmic Sobolev inequalities into statements about the behavior of the semigroup H_t.

2.1.2 The spectral gap

This section introduces the notion of spectral gap and gives bounds on convergence that depend only on the spectral gap and the stationary measure.

Definition 2.1.3 *Let K be a Markov kernel with Dirichlet form $\mathcal{E}$. The spectral gap $\lambda = \lambda(K)$ is defined by*

$$\lambda = \min\left\{\frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)}; \mathrm{Var}_\pi(f) \neq 0\right\}$$

Observe that λ is not, in general, an eigenvalue of $(I-K)$. If K is self-adjoint on $\ell^2(\pi)$ (that is, if (K,π) is reversible) then λ is the smallest non zero eigenvalue of $I-K$. In general λ is the smallest non zero eigenvalue of $I - \frac{1}{2}(K+K^*)$. Note also that the Dirichlet forms of K^* and K satisfy

$$\mathcal{E}_K(f,f) = \mathcal{E}_{K^*}(f,f).$$

It follows that $\lambda(K) = \lambda(K^*)$. Clearly, we also have

$$\lambda = \min\left\{\mathcal{E}(f,f); \|f\|_2 = 1,\ \pi(f) = 0\right\}.$$

Furthermore, if one wishes, one can impose that f be real in the definition of λ. Indeed, let λ_r be the quantity obtained for real f. Then $\lambda_r \geq \lambda$ and, if $f = u + iv$ with u, v real functions, then $\lambda_r \mathrm{Var}_\pi(f) = \lambda_r(\mathrm{Var}_\pi(u) + \mathrm{Var}_\pi(v)) \leq \mathcal{E}(v,v) + \mathcal{E}(u,u) = \mathcal{E}(f,f)$. Hence $\lambda_r \leq \lambda$ and finally $\lambda_r = \lambda$.

Lemma 2.1.4 *Let K be a Markov kernel with spectral gap $\lambda = \lambda(K)$. Then the semigroup $H_t = e^{-t(I-K)}$ satisfies*

$$\forall\, f \in \ell^2(\pi), \quad \|H_t f - \pi(f)\|_2^2 \leq e^{-2\lambda t}\,\mathrm{Var}_\pi(f).$$

PROOF: Set $u(t) = \mathrm{Var}_\pi(H_t f) = \|H_t(f - \pi(f))\|_2^2 = \|H_t f - \pi(f)\|_2^2$. Then

$$u'(t) = -2\,\mathcal{E}\,(H_t(f - \pi(f)), H_t(f - \pi(f))) \leq -2\lambda u(t).$$

It follows that

$$u(t) \leq e^{-2\lambda\, t} u(0)$$

which is the desired inequality because $u(0) = \mathrm{Var}_\pi(f)$.

As a corollary we obtain one of the simplest and most useful quantitative results in finite Markov chain theory.

Corollary 2.1.5 *Let K be a Markov kernel with spectral gap $\lambda = \lambda(K)$. Then the density $h_t^x(\cdot) = H_t^x(\cdot)/\pi(\cdot)$ satisfies*

$$\|h_t^x - 1\|_2 \leq \sqrt{1/\pi(x)}\, e^{-\lambda\, t}.$$

It follows that

$$|H_t(x,y) - \pi(y)| \leq \sqrt{\pi(y)/\pi(x)}\, e^{-\lambda\, t}.$$

PROOF: Let H_t^* be the adjoint of H_t on $\ell^2(\pi)$ (see Section 2.1.1). This is a Markov semigroup with spectral gap $\lambda(K^*) = \lambda(K)$. Set $\delta_x(y) = 1/\pi(x)$ if $y = x$ and $\delta_x(y) = 0$ otherwise. Then

$$h_t^x(y) = \frac{H_t^x(y)}{\pi(y)} = H_t^* \delta_x(y)$$

and, by Lemma 2.1.4 applied to K^*,

$$\|H_t^* \delta_x - 1\|_2^2 \leq e^{-2\lambda\, t} \mathrm{Var}_\pi(\delta_x).$$

Hence

$$\|h_t^x - 1\|_2 \leq \sqrt{\frac{1 - \pi(x)}{\pi(x)}}\, e^{-\lambda\, t} \leq \frac{1}{\sqrt{\pi(x)}}\, e^{-\lambda\, t}.$$

Of course, the same result holds for H_t^*. Hence

$$\begin{aligned} |h_t(x,y) - 1| &= \left| \sum_z (h_{t/2}(x,z) - 1)(h_{t/2}(z,y) - 1)\pi(z) \right| \\ &\leq \|h_{t/2}^x - 1\|_2 \|h_{t/2}^{*y} - 1\|_2 \\ &\leq \frac{1}{\sqrt{\pi(x)\pi(y)}}\, e^{-\lambda\, t}. \end{aligned}$$

Multiplying by $\pi(y)$ yields the desired inequality. This ends the proof of Corollary 2.1.5.

Definition 2.1.6 *Let $\omega = \omega(K) = \min\{\Re(\zeta) : \zeta \neq 0$ an eigenvalue of $I - K\}$.*

Let $\mathcal{S}$ denote the spectrum of $I - K$. Since $H_t = e^{-t(I-K)}$, the spectrum of H_t is $\{e^{-t\xi} : \xi \in \mathcal{S}\}$. It follows that the spectral radius of $H_t - E_\pi$ in $\ell^2(\pi)$ is $e^{-t\omega}$. Using (1.2.5) we obtain the following result.

Theorem 2.1.7 *Let K be an irreducible Markov kernel. Then*

$$\forall\, 1 \le p \le \infty, \quad \lim_{t\to\infty} \frac{-1}{t} \log\left(\max_x \|h_t^x - 1\|_p\right) = \omega.$$

In particular, $\lambda \le \omega$ with equality if (K,π) is reversible. Furthermore, if we set

$$T_p = T_p(K, 1/e) = \min\left\{t > 0 : \max_x \|h_t^x - 1\|_p \le 1/e\right\}, \tag{2.1.3}$$

and define π_* as in (1.3.2) *then, for $1 \le p \le 2$,*

$$\frac{1}{\omega} \le T_p \le \frac{1}{2\lambda}\left(2 + \log\frac{1}{\pi_*}\right),$$

whereas, for for $2 < p \le \infty$,

$$\frac{1}{\omega} \le T_p \le \frac{1}{\lambda}\left(1 + \log\frac{1}{\pi_*}\right).$$

EXAMPLE 2.1.1: Let $\mathcal{X} = \{0,\dots,n\}$. Consider the Kernel $K(x,y)) = 1/2$ if $y = x \pm 1$, $(x,y) = (0,0)$ or (n,n), and $K(x,y) = 0$ otherwise. This is a symmetric kernel with uniform stationary distribution $\pi \equiv 1/(n+1)$. Feller [40], page 436, gives the eigenvalues and eigenfunctions of K. For $I - K$, we get the following:

$$\lambda_0 = 0, \quad \psi_0(x) \equiv 1$$

$$\lambda_j = 1 - \cos\frac{\pi j}{n+1}, \quad \psi_j(x) = \sqrt{2}\cos(\pi j(x+1/2)/(n+1)) \quad \text{for} \quad j = 1,\dots,n.$$

Let $H_t = e^{-t(I-K)}$ and write (using $\cos(\pi x) \le 1 - 2x^2$ for $0 \le x \le 1$)

$$\begin{aligned} |h_t(x,y) - 1| &= \left|\sum_{j=1}^{n} \psi_j(x)\psi_j(y) e^{-t(1-\cos(\pi j/(n+1)))}\right| \\ &\le 2\sum_{j=1}^{n} e^{-2tj^2/(n+1)^2} \\ &\le 2e^{-2t/(n+1)^2}\left(1 + \sqrt{(n+1)^2/2t}\right). \end{aligned}$$

To obtain the last inequality, use

$$\sum_2^n e^{-2tj^2/(n+1)^2} \le \int_1^\infty e^{-2ts^2/(n+1)^2} ds = \frac{n+1}{\sqrt{2t}} \int_{\frac{\sqrt{2t}}{n+1}}^\infty e^{-u^2} du$$

and

$$\frac{2}{\sqrt{\pi}} \int_z^\infty e^{-u^2} du = \frac{2e^{-z^2}}{\sqrt{\pi}} \int_z^\infty e^{-(u-z)^2 - 2(u-z)z} du \le e^{-z^2}.$$

In particular,

$$\max_{x,y}|h_{2t}(x,y)-1| = \max_x \|h_t^x - 1\|_2^2 \le 2e^{-c} \quad \text{for} \quad t = \frac{1}{4}(n+1)^2(1+c)$$

and $T_2(K, 1/e) \le 3(n+1)^2/4$. Also, $\omega = \lambda = 1 - \cos\frac{\pi}{n+1} \le \pi^2/(n+1)^2$. Hence in this case, the lower bound for $T_2(K, 1/e)$ given by Theorem 2.1.6 is of the right order of magnitude whereas the upper bound

$$T_2 \le \frac{1}{2\lambda}\left(2 + \log\frac{1}{\pi_*}\right) \le \frac{1}{4}(n+1)^2(2+\log(n+1))$$

is off by a factor of $\log(n+1)$.

EXAMPLE 2.1.2: Let $\mathcal{X} = \{0,1\}^n$ and $K(x,y) = 0$ unless $|x-y| = \sum_i |x_i - y_i| = 1$ in which case $K(x,y) = 1/n$. Viewing $\mathcal{X}$ as an Abelian group it is not hard to see that the characters

$$\chi_y : x \to (-1)^{y.x}, \quad y \in \{0,1\}^n$$

where $x.y = \sum_i x_i y_i$, form an orthonormal basis of $\ell^2(\pi)$, $\pi \equiv 2^{-n}$. Also

$$\begin{aligned} K\chi_y(x) &= \sum_z K(x,z)\chi_y(z) \\ &= \left(\frac{1}{n}\sum_i (-1)^{e_i.y}\right)\chi_y(x) = \frac{n-2|y|}{n}\chi_y(x). \end{aligned}$$

This shows that χ_y is an eigenfunction of $I-K$ with eigenvalue $2|y|/n$ where $|y|$ is the number of 1's in y. Thus the eigenvalue $2j/n$ has multiplicity $\binom{n}{j}$ $0 \le j \le n$. This information leads to the bound

$$\begin{aligned} \|h_t^x - 1\|_2^2 &= \sum_1^n \binom{n}{j} e^{-4tj/n} \\ &\le \sum_1^n \frac{n^j}{j!} e^{-4tj/n} \\ &\le e^{ne^{-4t/n}} - 1. \end{aligned}$$

Hence

$$\|h_t^x - 1\|_2^2 \le e^{1-c} \quad \text{for} \quad t = \frac{1}{4}n(\log n + c), \quad c > 0.$$

It follows that $T_2(K, 1/e) \le \frac{1}{4}n(2+\log n)$. Also, $\|h_t^x - 1\|_2^2 \ge ne^{-4t/n}$ hence $T_2 = T_2(K, 1/e) \ge \frac{1}{4}n(1+\log n)$. In this case, the lower bound

$$T_2 \ge \frac{1}{\lambda} = \frac{1}{\omega} = \frac{4}{n}$$

is off by a factor of $\log n$ whereas the upper bound

$$T_2 \le \frac{2}{\lambda}\left(2 + \log\frac{1}{\pi_*}\right) = \frac{n}{2}(2+n).$$

is off by a factor of $n/\log n$.

2.1.3 Chernoff bounds and central limit theorems

It is well established that ergodic Markov chains satisfy large deviation bounds of Chernoff's type for

$$\mathbf{P}_q\left(\frac{1}{t}\int_0^t f(X_s)ds - \pi(f) > \gamma\right)$$

as well as central limit theorems to the effect that

$$\left|\mathbf{P}_q\left(\int_0^t f(X_s)ds - t\pi(f) \le \sigma t^{1/2}\gamma\right) - \Phi(\gamma)\right| \to 0$$

where $\Phi(\gamma)$ is the cumulative Gaussian distribution and σ is an appropriate number depending on f and K (the asymptotic variance).

The classical treatment of these problems leads to results having a strong asymptotic flavor. Turning these results into quantitative bounds is rather frustrating even in the context of finite Markov chains.

Some progress has been made recently in this direction. This short section presents without any detail two of the main results obtained by Pascal Lezaud [59] and Brad Mann [61] in their Ph.D. theses respectively at Toulouse and Harvard universities.

The work of Lezaud clarifies previous results of Gillman [44] and Dinwoodie [36, 37] on quantitative Chernoff bounds for finite Markov chains. A typical result is as follows (there are also discrete time versions).

Theorem 2.1.8 *Let (K,π) be a finite irreducible Markov chain. Let q denote the initial distribution and $\mathbf{P}_q$ be the law of the associated continuous time process $(X_t)_{t>0}$. Then, for all functions f such that $\pi(f) = 0$ and $\|f\|_\infty \le 1$,*

$$\mathbf{P}_q\left(\frac{1}{t}\int_0^t f(X_s)ds > \gamma\right) \le \|q/\pi\|_2 \exp\left(-\frac{\gamma^2\lambda t}{10}\right).$$

Concerning the Berrry-Essen central limit theorem, we quote a continuous time version of one of Brad Mann's result which has been obtained by Pascal Lezeaud.

Theorem 2.1.9 *Let (K,π) be a finite irreducible reversible Markov chain. Let q denote the initial distribution and $\mathbf{P}_q$ be the law of the associated continuous time process $(X_t)_{t>0}$. Then, for $t > 0$, $-\infty < \gamma < \infty$ and for all functions f such that $\pi(f) = 0$ and $\|f\|_\infty \le 1$,*

$$\left|\mathbf{P}_q\left(\frac{1}{\sigma\sqrt{t}}\int_0^t f(X_s)ds \le \gamma\right) - \Phi(\gamma)\right| \le \frac{100\|q/\pi\|_2\|f\|_2^2}{\lambda^2\sigma^3 t^{1/2}}$$

where

$$\sigma^2 = \lim_{t\to\infty}\frac{1}{t}\mathrm{Var}_\pi\left(\int_0^t f(X_s)ds\right).$$

See [41, 61, 59, 28] for details and examples. There are non-reversible and/or discrete time versions of the last theorem. Mann's Thesis contains a nice discussion of the history of the subject and many references.

2.2 Hypercontractivity

This section introduces the notions of logarithmic Sobolev constant and of hypercontractivity and shows how they enter convergence bounds. A very informative account of the development of hypercontractivity and logarithmic Sobolev inequalities can be found in L. Gross survey paper [47]. See also [7, 8, 15, 16, 46]. The paper [29] develops applications of these notions to finite Markov chains.

2.2.1 The log-Sobolev constant

The definition of the logarithmic Sobolev constant α is similar to that of the spectral gap λ where the variance has been replaced by

$$\mathcal{L}(f) = \sum_{x \in X} |f(x)|^2 \log \left(\frac{|f(x)|^2}{\|f\|_2^2} \right) \pi(x).$$

Observe that $\mathcal{L}(f)$ is nonnegative. This follows from Jensen's inequality applied to the convex function $\phi(t) = t^2 \log t^2$. Furthermore $\mathcal{L}(f) = 0$ if and only if f is constant.

Definition 2.2.1 *Let K be an irreducible Markov chain with stationary measure π. The logarithmic constant $\alpha = \alpha(K)$ is defined by*

$$\alpha = \min \left\{ \frac{\mathcal{E}(f,f)}{\mathcal{L}(f)}; \mathcal{L}(f) \neq 0 \right\}.$$

It follows from the definition that α is the largest constant c such that the logarithmic Sobolev inequality

$$c\mathcal{L}(f) \leq \mathcal{E}(f,f)$$

holds for all functions f. Observe that one can restrict f to be real nonnegative in the definition of α since $\mathcal{L}(f) = \mathcal{L}(|f|)$ and $\mathcal{E}(|f|,|f|) \leq \mathcal{E}(f,f)$.

To get a feel for this notion we prove the following result.

Lemma 2.2.2 *For any chain K the log-Sobolev constant α and the spectral gap λ satisfy $2\alpha \leq \lambda$.*

PROOF: We follow [67]. Let g be real and set $f = 1 + \varepsilon g$ and write, for ε small enough

$$\begin{aligned} |f|^2 \log |f|^2 &= 2\left(1 + 2\varepsilon g + \varepsilon^2 |g|^2\right)\left(\varepsilon g - \frac{\varepsilon^2 |g|^2}{2} + O(\varepsilon^3)\right) \\ &= 2\varepsilon g + 3\varepsilon^2 |g|^2 + O(\varepsilon^3) \end{aligned}$$

and

$$\begin{aligned} |f|^2 \log \|f\|_2^2 &= \left(1 + 2\varepsilon g + \varepsilon^2 |g|^2\right)\left(2\varepsilon \pi(g) + \varepsilon^2 \|g\|_2^2 - 2\varepsilon^2 (\pi(g))^2 + O(\varepsilon^3)\right) \\ &= 2\varepsilon \pi(g) + 4\varepsilon^2 g \pi(g) + \varepsilon^2 \|g\|_2^2 - 2\varepsilon^2 (\pi(g))^2 + O(\varepsilon^3). \end{aligned}$$

Thus,

$$|f|^2 \log \frac{|f|^2}{\|f\|_2^2} = 2\varepsilon(g - \pi(g)) + \varepsilon^2 \left(3|g|^2 - \|g\|_2^2 - 4g\pi(g) + 2(\pi(g))^2\right) + O(\epsilon^3)$$

and

$$\begin{aligned} \mathcal{L}(f) &= 2\varepsilon^2 \left(\|g\|^2 - (\pi(g))^2\right) + O(\varepsilon^3) \\ &= 2\varepsilon^2 \mathrm{Var}(g) + O(\varepsilon^3). \end{aligned}$$

To finish the proof, observe that $\mathcal{E}(f,f) = \varepsilon^2 \mathcal{E}(g,g)$, multiply by ε^{-2}, use the variational characterizations of α and λ, and let ε tend to zero.

It is not completely obvious from the definition that $\alpha(K) > 0$ for any finite irreducible Markov chain. The next result, adapted from [65, 66, 67], yields a proof of this fact.

Theorem 2.2.3 *Let K be an irreducible Markov chain with stationary measure π. Let α be its logarithmic Sobolev constant and λ its spectral gap. Then either $\alpha = \lambda/2$ or there exists a positive non-constant function u which is solution of*

$$2u \log u - 2u \log \|u\|_2 - \frac{1}{\alpha}(I - K)u = 0, \tag{2.2.1}$$

and such that $\alpha = \mathcal{E}(u,u)/\mathcal{L}(u)$. In particular $\alpha > 0$.

PROOF: Looking for a minimizer of $\mathcal{E}(f,f)/\mathcal{L}(f)$, we can restrict ourselves to non-negative functions satisfying $\pi(f) = 1$. Now, either there exists a non-constant non-negative minimizer (call it u), or the minimum is attained at the constant function 1 where $\mathcal{E}(1,1) = \mathcal{L}(1) = 0$. In this second case, the proof of Lemma 2.2.2 shows that we must have $\alpha = \lambda/2$ since, for any function $g \not\equiv 0$ satisfying $\pi(g) = 0$,

$$\lim_{\varepsilon \to 0} \frac{\mathcal{E}(1+\varepsilon g, 1+\varepsilon g)}{\mathcal{L}(1+\varepsilon g)} = \lim_{\varepsilon \to 0} \frac{\varepsilon^2 \mathcal{E}(g,g)}{2\varepsilon^2 \mathrm{Var}_\pi(g)} \geq \frac{\lambda}{2}.$$

Hence, either $\alpha = \lambda/2$ or there must exist a non-constant non-negative function u which minimizes $\mathcal{E}(f,f)/\mathcal{L}(f)$. It is not hard to show that any minimizer of $\mathcal{E}(f,f)/\mathcal{L}(f)$ must satisfy (2.2.1). Finally, if $u \geq 0$ is not constant and satisfies (2.2.1) then u must be positive. Indeed, if it vanishes at $x \in \mathcal{X}$ then $Ku(x) = 0$ and u must vanishe at all points y such that $K(x,y) > 0$. By irreducibility, this would imply $u \equiv 0$, a contradiction.

2.2.2 Hypercontractivity, α, and ergodicity

We now recall the main result relating log-Sobolev inequalities to the so-called hypercontractivity of the semigroup H_t. For a history of this result see Gross' survey [47]. See also [7, 8, 16, 46]. A proof can also be found in [29].

Theorem 2.2.4 *Let (K,π) be a finite Markov chain with log-Sobolev constant α.*

1. *Assume that there exists $\beta > 0$ such that $\|H_t\|_{2\to q} \le 1$ for all $t > 0$ and $2 \le q < +\infty$ satisfying $e^{4\beta t} \ge q-1$. Then $\beta\mathcal{L}(f) \le \mathcal{E}(f,f)$ for all f and thus $\alpha \ge \beta$.*

2. *Assume that (K,π) is reversible. Then $\|H_t\|_{2\to q} \le 1$ for all $t > 0$ and all $2 \le q < +\infty$ satisfying $e^{4\alpha t} \ge q-1$.*

3. *For non-reversible chains, we still have $\|H_t\|_{2\to q} \le 1$ for all $t > 0$ and all $2 \le q < +\infty$ satisfying $e^{2\alpha t} \ge q-1$.*

We will not prove this result but only comment on the different statements. First let us assume that (K,π) is reversible. The first two statements show that α can also be characterized as the largest β such that

$$\|H_t\|_{2\to q} \le 1 \text{ for all } t > 0 \text{ and all } 2 \le q < +\infty \text{ satisfying } e^{4\beta t} \ge q-1. \quad (2.2.2)$$

Recall that H_t is always a contraction on $\ell^2(\pi)$ and that, in fact, $\|H_t\|_{2\to 2} = 1$ for all $t > 0$. Also, (1.2.8) and (1.2.11) easily show that $\|H_t\|_{2\to\infty} > 1$ for all $t > 0$ and tends to 1 as t tends to infinity. Thus, even in the finite setting, it is rather surprising that for each $2 < q < \infty$ there exists a finite $t_q > 0$ such that $\|H_t\|_{2\to q} \le 1$ for $t \ge t_q$. The fact that such a t_q exists follows from Theorem 2.2.3 and Theorem 2.2.4(2).

Statements 2 and 3 in Theorem 2.2.4 are the keys of the following theorem which describes how α enters quantitative bounds on convergence to stationarity.

Theorem 2.2.5 *Let (K,π) be a finite Markov chain. Then, for $\varepsilon, \theta, \sigma \ge 0$ and $t = \varepsilon + \theta + \sigma$,*

$$\|h_t^x - 1\|_2 \le \begin{cases} \|h_\varepsilon^x\|_2^{2/(1+e^{4\alpha\theta})}\, e^{-\lambda\sigma} & \text{if } (K,\pi) \text{ is revesible} \\ \|h_\varepsilon^x\|_2^{2/(1+e^{2\alpha\theta})}\, e^{-\lambda\sigma} & \text{in general.} \end{cases} \quad (2.2.3)$$

In particular,

$$\|h_t^x - 1\|_2 \le e^{1-c} \quad (2.2.4)$$

for all $c \ge 0$ and

$$t = \begin{cases} (4\alpha)^{-1} \log_+ \log(1/\pi(x)) + \lambda^{-1} c & \text{for reversible chains} \\ (2\alpha)^{-1} \log_+ \log(1/\pi(x)) + \lambda^{-1} c & \text{in general} \end{cases}$$

where $\log_+ t = \max\{0, \log t\}$.

PROOF: We treat the general case. The improvement for reversible chains follows from Theorem 2.2.4(2). For $\theta > 0$, set $q(\theta) = 1 + e^{2\alpha\theta}$. The third statement of Theorem 2.2.4(3) gives $\|H_\theta\|_{2\to q(\theta)} \le 1$. By duality, it follows that $\|H_\theta^*\|_{q'(\theta)\to 2} \le 1$ where $q'(\theta)$ is the Hölder conjugate of $q(\theta)$ defined by $1/q'(\theta) + 1/q(\theta) = 1$. Write

$$\begin{aligned} \|h^x_{\varepsilon+\theta+\sigma} - 1\|_2 &= \|(H^*_{\theta+\sigma} - \pi)h^x_\varepsilon\|_2 \le \|H^*_\theta h^x_\varepsilon\|_2 \|H^*_\sigma - \pi\|_{2\to 2} \\ &\le \|h^x_\varepsilon\|_{q'(\theta)} \|H^*_\theta\|_{q'(\theta)\to 2} \|H^*_\sigma - \pi\|_{2\to 2} \le \|h^x_\varepsilon\|_2^{2/q(\theta)} e^{-\lambda\sigma}. \end{aligned}$$

Here we have used $1 \le q' \le 2$ and the Hölder inequality

$$\|f\|_{q'} \le \|f\|_1^{1-2/q} \|f\|_2^{2/q}$$

with $f = h^x_\varepsilon$, $\|h^x_\varepsilon\|_1 = 1$ to obtain the last inequality.

Consider the function δ_x defined by $\delta_x(x) = 1/\pi(x)$ and $\delta_x(y) = 0$ for $x \ne y$ and observe that $h^x_0 = \delta_x$, $\|h^x_0\|_2 = \|\delta_x\|_2 \le 1/\pi(x)^{1/2}$. Hence, for $t = \theta + \sigma$,

$$\|h^x_t - 1\|_2 \le \left(\frac{1}{\pi(x)}\right)^{1/(1+e^{2\alpha\theta})} e^{-\lambda\sigma}.$$

Assuming $\pi(x) < 1/e$ and choosing

$$\theta = \frac{1}{2\alpha} \log\log\frac{1}{\pi(x)}, \quad \sigma = \frac{c}{\lambda}$$

we obtain $\|h_t - 1\|_2 \le e^{1-c}$ which is the desired inequality. When $\pi(x) \ge 1/e$, simply use $\theta = 0$.

Corollary 2.2.6 *Let (K, π) be a finite Markov chain. Then*

$$\left|\frac{H_t(x,y)}{\pi(y)} - 1\right| = |h_t(x,y) - 1| \le e^{2-c} \tag{2.2.5}$$

for all $c > 0$ and

$$t = \begin{cases} (4\alpha)^{-1}\left(\log_+ \log(1/\pi(x)) + \log_+\log(1/\pi(y))\right) + \lambda^{-1} c & \textit{(reversible)} \\ (2\alpha)^{-1}\left(\log_+ \log(1/\pi(x)) + \log_+\log(1/\pi(y))\right) + \lambda^{-1} c & \textit{(general)}. \end{cases}$$

PROOF: Use Theorem 2.2.5 for both H_t and H_t^* together with

$$|h_{t+s}(x,y) - 1| \le \|h^x_t - 1\|_2 \|h^{*y}_s - 1\|_2.$$

The next result must be compared with Theorem 2.1.7.

Corollary 2.2.7 *Let (K, π) be a finite reversible Markov chain. For $1 \le p \le \infty$, let T_p be defined by* (2.1.3). *Then, for $1 \le p \le 2$,*

$$\frac{1}{2\alpha} \le T_p \le \frac{1}{4\alpha}\left(4 + \log_+\log\frac{1}{\pi_*}\right)$$

and for $2 < p \le \infty$,

$$\frac{1}{2\alpha} \le T_p \le \frac{1}{2\alpha}\left(3 + \log_+\log\frac{1}{\pi_*}\right)$$

where $\pi_* = \min_x \pi(x)$ as in (1.3.2). *Similar upper bounds holds in the non-reversible case (simply multiply the right-hand side by 2).*

This result shows that α is closely related to the quantity we want to bound, namely the "time to equilbrium" T_2 (more generally T_p) of the chain (K, π). The natural question now is:

$$\boxed{\text{can one compute or estimate the constant } \alpha?}$$

Unfortunately, the present answer is that it seems to be a very difficult problem to estimate α. To illustrate this point we now present what, in some sense, is the only example of finite Markov chain for which α is known explicitely.

EXAMPLE 2.2.1: Let $\mathcal{X} = \{0,1\}$ be the two point space. Fix $0 < \theta \leq 1/2$. Consider the Markov kernel $K = K_\theta$ given by $K(0,0) = K(1,0) = \theta$, $K(0,1) = K(1,1) = 1-\theta$. The chain K_θ is reversible with respect to π_θ where $\pi_\theta(0) = (1-\theta)$, $\pi_\theta(1) = \theta$.

Theorem 2.2.8 *The log-Sobolev constant of the chain (K_θ, π_θ) on $\mathcal{X} = \{0,1\}$ is given by*

$$\alpha_\theta = \frac{1-2\theta}{\log[(1-\theta)/\theta]}$$

with $\alpha_{1/2} = 1/2$.

PROOF: The case $\theta = 1/2$ is due to Aline Bonami [10] and is well known since the work of L. Gross [46]. The case $\theta < 1/2$ has only been worked out recently in [29] and independently in [48]. The present elegant proof is due to Sergei Bobkov. He kindly authorized me to include his argument in these notes.

First, linearize the problem by observing that

$$\mathcal{L}(f) = \sup\left\{\langle f^2, g\rangle : g \neq 0,\ \|e^g\|_1 = 1\right\}.$$

Hence

$$\alpha = \inf\{\alpha(g) : g \neq 0,\ \|e^g\|_1 = 1\}$$

with

$$\alpha(g) = \inf\left\{\frac{\mathcal{E}_\theta(f,f)}{\langle f^2, g\rangle} : f \neq 0\right\}$$

where $\mathcal{E}_\theta$ is the Dirichlet form $\mathcal{E}_\theta(f,f) = \theta(1-\theta)|f(0) - f(1)|^2$. This is valid for any Markov chain.

We now return to the two point space. Fix $g \neq 0$ and set $g(0) = b$, $g(1) = a$ with $\theta e^a + (1-\theta)e^b = 1$. Observe that this implies $ab < 0$. To find $\alpha_\theta(g)$ we can assume $f > 0$, $f(0) = \sqrt{x}$, $f(1) = \sqrt{y} = 1$ with $x > 0$. Then

$$\alpha_\theta(g) = \inf_{x>0}\left\{\frac{\theta(1-\theta)(\sqrt{x}-1)^2}{\theta x a + (1-\theta)b}\right\}.$$

One easily checks that the infimum is attained for $x = [(1-\theta)b/\theta a]^2$. Therefore

$$\alpha_\theta(g) = \frac{\theta}{b} + \frac{1-\theta}{a}.$$

It follows that

$$\alpha_\theta = \inf\left\{\frac{\theta}{b} + \frac{1-\theta}{a} : \theta e^a + (1-\theta)e^b = 1\right\}.$$

We set

$$t = e^a, \quad s = e^b$$

and

$$h(t) = \frac{\theta}{\log s} + \frac{1-\theta}{\log t} \text{ with } \theta t + (1-\theta)s = 1,$$

so that

$$\alpha_\theta = \inf\{h(t) : t \in (0,1) \cup (1, 1/\theta)\}.$$

By Taylor expansion at $t = 1$,

$$h(t) = \frac{1}{2} + \frac{2\theta - 1}{12(1-\theta)}(t-1) + \frac{\theta^3 + (1-\theta)^3}{24(1-\theta)^2}(t-1)^2 + O((t-1)^3).$$

So, we extend h as a continuous function on $[0, 1/\theta]$ by setting

$$h(0) = -\theta/\log(1-\theta), \quad h(1) = 1/2, \quad h(1/\theta) = -(1-\theta)/\log\theta.$$

Observe that $h(1)$ is not a local minimum if $\theta \neq 1/2$. We have

$$h'(t) = \frac{\theta^2}{(1-\theta)s[\log s]^2} - \frac{(1-\theta)}{t[\log t]^2}.$$

This shows that neither $h(0)$ nor $h(1/\theta)$ are minima of h since $h'(0) = -\infty$, $h'(1/\theta) = +\infty$.

Let us solve $h'(t) = 0$ and show that this equation has a unique solution in $(0, 1/\theta)$. The condition $h'(t) = 0$ is equivalent to (recall that $(\log s)(\log t) < 0$)

$$\begin{cases} \theta\sqrt{t}\log t = -(1-\theta)\sqrt{s}\log s \\ \theta t + (1-\theta)s = 1 \end{cases}.$$

Since $\theta t + (1-\theta)s = 1$, we have $\theta = (1-s)/(t-s)$, $1 - \theta = (1-t)/(s-t)$. Hence $h'(t) = 0$ implies $s = t = 1$ or

$$\frac{\sqrt{t}\log t}{1-t} = \frac{\sqrt{s}\log s}{1-s}.$$

The function $t \to v(t) = \frac{\sqrt{t}\log t}{1-t}$ satisfies $v(0) = v(+\infty) = 0$, $v(1) = -1$ and $v(1/t) = v(t)$. It is decreasing on $(0,1)$ and increasing on $(1, +\infty)$. It follows that $h'(t) = 0$ implies that either $s = t = 1$ or $t = 1/s = (1-\theta)/\theta$ (because $\theta t + (1-\theta)s = 1$). If $\theta \neq 1/2$ then $h'(1) \neq 0$, the equation $h'(t) = 0$ has a unique solution $t = (1-\theta)/\theta$ and

$$\min_{t\in(0,1/\theta)} h(t) = h((1-\theta)/\theta) = \frac{1-2\theta}{\log[(1-\theta)/\theta]}.$$

If $\theta = 1/2$, then $h'(1) = 0$ and 1 is the only solution of $h'(t) = 0$ so that $\min_{t\in(0,2)} h(t) = h(1) = 1/2$ in this case. This proves Theorem 2.2.8.

EXAMPLE 2.2.2: Using Theorems 2.2.3 and 2.2.8, one obtains the following result.

Theorem 2.2.9 *Let π be a positive probability measure on $\mathcal{X}$. Let $K(x,y) = \pi(y)$. Then the log-Sobolev constant of (K,π) is given by*

$$\alpha = \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]}$$

where $\pi_* = \min_{\mathcal{X}} \pi$.

PROOF: Theorem 2.2.3 shows that any non trivial minimizer must take only two values. The desired result then follows from Theorem 2.2.8. See [29] for details. THeorem 2.2.9 yields a sharp universal lower bound on α in terms of λ.

Corollary 2.2.10 *The log-Sobolev constant α and the spectral gap λ of any finite Markov chain K with stationary measure π satisfy*

$$\alpha \geq \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]}\,\lambda.$$

PROOF: The variance $\mathrm{Var}_\pi(f)$ is nothing else than the Dirichlet form of the chain considered in Theorem 2.2.9. Hence

$$\frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]}\,\mathcal{L}_\pi(f) \leq \mathrm{Var}_\pi(f) \leq \frac{1}{\lambda}\mathcal{E}_{K,\pi}(f,f).$$

The desired result follows.

2.2.3 Some tools for bounding α from below

The following two results are extremely useful in providing examples of chains where α can be either computed or bounded from below. Lemma 2.2.11 computes the log-Sobolev constant of products chains. This important result is due (in greater generality) to I Segal and to W. Faris, see [47]. Lemma 2.2.12 is a comparison result.

Lemma 2.2.11 *Let (K_i,π_i), $i = 1,\ldots,d$, be Markov chains on finite sets $\mathcal{X}_i$ with spectral gaps λ_i and log-Sobolev constants α_i. Fix $\mu = (\mu_i)_1^d$ such that $\mu_i > 0$ and $\sum \mu_i = 1$. Then the product chain (K,π) on $\mathcal{X} = \prod_1^d \mathcal{X}_i$ with Kernel*

$$\begin{aligned} K_\mu(x,y) &= K(x,y) \\ &= \sum_1^d \mu_i \delta(x_1,y_1)\ldots\delta(x_{i-1},y_{y_i-1})K_i(x_i,y_i)\delta(x_{i+1},y_{i+1})\ldots\delta(x_d,y_d) \end{aligned}$$

(where $\delta(x,y)$ vanishes for $x \neq y$ and $\delta(x,x) = 1$) and stationary measure $\pi = \bigotimes_1^d \pi_i$ satisfies

$$\lambda = \min_i\{\mu_i\lambda_i\}\ , \quad \alpha = \min_i\{\mu_i\alpha_i\}.$$

PROOF: Let $\mathcal{E}_i$ denote the Dirichlet form associated to K_i, then the product chain K has Dirichlet form

$$\mathcal{E}(f,f) = \sum_1^d \mu_i \left(\sum_{x_j : j\neq i} \mathcal{E}_i(f,f)(x^i)\pi^i(x^i) \right)$$

where x^i is the sequence $(x_1,\ldots,x_d)$ with x_i omitted, $\pi^i = \bigotimes_{\ell:\ell\neq i}\pi_\ell$ and $\mathcal{E}_i(f,f)(x^i) = \mathcal{E}_i(f(x_1,\ldots,x_d), f(x_1,\ldots,x_d))$ has the obvious meaning: $\mathcal{E}_i$ acts on the i^{th} coordinate whereas the other coordinates are fixed. It is enough to prove the Theorem when $d = 2$. We only prove the statement for α. The proof for λ is similar. Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ be a nonnegative function and set $F(x_2) = \left(\sum_{x_1} f(x_1,x_2)^2\pi_1(x_1)\right)^{1/2}$. Write

$$\begin{aligned}
\mathcal{L}(f) &= \sum_{x_1,x_2} |f(x_1,x_2)|^2 \log \frac{f(x_1,x_2)^2}{\|f\|_{2,\pi}^2}\, \pi(x_1,x_2) \\
&= \sum_{x_2} |F(x_2)|^2 \log \frac{F(x_2)^2}{\|F\|_{2,\pi_2}^2}\, \pi_2(x_2) \\
&\quad + \sum_{x_1,x_2} |f(x_1,x_2)|^2 \log \frac{f(x_1,x_2)^2}{F(x_2)^2}\, \pi(x_1,x_2) \\
&\leq [\mu_2\alpha_2]^{-1}\mu_2\mathcal{E}_2(F,F) + [\mu_1\alpha_1]^{-1} \sum_{x_2} \mu_1\mathcal{E}_1(f(\cdot,x_2), f(\cdot,x_2))\, \pi_2(x_2).
\end{aligned}$$

Now, the triangle inequality

$$\begin{aligned}
|F(x_2) - F(y_2)| &= \left|\, \|f(\cdot,x_2)\|_{2,\pi_1} - \|f(\cdot,y_2)\|_{2,\pi_1} \right| \\
&\leq \|f(\cdot,x_2) - f(\cdot,y_2)\|_{2,\pi_1}
\end{aligned}$$

implies that

$$\mathcal{E}_2(F,F) \leq \sum_{x_1} \mathcal{E}_2(f(x_1,\cdot), f(x_1,\cdot))\, \pi_1(x_1).$$

Hence

$$\begin{aligned}
\mathcal{L}(f) \leq\ & [\mu_2\alpha_2]^{-1} \sum_{x_1} \mu_2\mathcal{E}_2(f(x_1,\cdot), f(x_1,\cdot))\, \pi_1(x_1) \\
& + [\mu_1\alpha_1]^{-1} \sum_{x_2} \mu_1\mathcal{E}_1(f(\cdot,x_2), f(\cdot,x_2))\pi_2(x_2)
\end{aligned}$$

which yields

$$\mathcal{L}(f) \leq \max_i \{1/[\mu_i \alpha_i]\} \mathcal{E}(f, f).$$

This shows that $\alpha \geq \min_i [\mu_i \alpha_i]$. Testing on functions that depend only on one of the two variables shows that $\alpha = \min_i [\mu_i \alpha_i]$.

EXAMPLE 2.2.3: Fix $0 < \theta < 1$. Take each $\mathcal{X}_i = \{0, 1\}$, $\mu_i = 1/d$, $K_i = K_\theta$ as in Theorem 2.2.8. We obtain a chain on $\mathcal{X} = \{0, 1\}^d$ which proceeds as follows. If the current state is x, we pick a coordinate, say i, uniformly at random. If $x_i = 0$ we change it to 1 with probability $1 - \theta$ and do nothing with probability θ. If $x_i = 1$ we change it to 0 with probability θ and do nothing with pobability $1 - \theta$. According to Lemma 2.2.11, this chain has spectral gap $\lambda = 1/d$ and log-Sobolev constant

$$\alpha = \frac{1 - 2\theta}{d \log[(1 - \theta)/\theta]}.$$

Observe that the function $F(t) : t \to c(1 - \theta - t)$ with $c = (\theta(1 - \theta))^{-1/2}$ is an eigenfunction of K_i (for each i) with eigenvalue $0 = 1 - \lambda$ satisfying $\|F_i\|_2 = 1$. It follows that the eigenvalues of $I - K$ are the numbers j/d each with multiplicity $\binom{d}{j}$. The corresponding orthonormal eigenfunctions are

$$F_I : (x)_1^d \to \prod_{i \in I} F_i(x)$$

where $I \subset \{1, \dots, d\}$, $F_i(x) = F(x_i)$ and $\#I = j$. The product structure of the chain K yields

$$\|h_t^x - 1\|_2^2 = h_{2t}(x, x) - 1 = \prod_1^d (1 + |F_i(x)|^2 e^{-2t/d})^d - 1.$$

For instance,

$$\begin{aligned} \|h_t^0 - 1\|_2^2 &= \left(1 + \frac{1 - \theta}{\theta} e^{-2t/d}\right)^d - 1 \\ &\leq \frac{(1 - \theta)d}{\theta} e^{-2t/d} e^{\frac{(1-\theta)d}{\theta} e^{-2t/d}}. \end{aligned}$$

In particular

$$\|h_t^0 - 1\|_2 \leq e^{\frac{1}{2} - c} \quad \text{for} \quad t = \frac{d}{2} (\log[(1 - \theta)d/\theta] + 2c), \quad c > 0.$$

Hence

$$T_2(K_\theta, 1/e) \leq \frac{d}{2} (3 + \log[(1 - \theta)d/\theta]), \quad c > 0.$$

Also, we have

$$\|h_t^0 - 1\|_2^2 \geq \frac{(1 - \theta)d}{\theta} e^{-2t/d}$$

which shows that the upper bound obtained above is sharp and that

$$T_2(K,1/e) \geq \frac{d}{2}\left(2+\log[(1-\theta)d/\theta]\right).$$

It is instructive to compare these precise results with the upper bound which follows from Theorem 2.2.5. In the present case this theorem yields

$$\|h_t^0 - 1\|_2 \leq e^{1-c} \quad \text{for } t = \frac{d}{2}\left(\frac{1}{2(1-2\theta)}\left(\log\frac{1-\theta}{\theta}\right)\log d \ + 2c\right).$$

For any fixed $\theta < 1/2$, this is slightly off, but of the right order of magnitude. For $\theta = 1/2$ this simplifies to

$$\|h_t^0 - 1\|_2 \leq e^{1-c} \quad \text{for } t = \frac{d}{2}(\log d \ + 2c)$$

which is very close to the sharp result described above. In this case, the upper bound

$$T_2 = T_2(K_{1/2},1/e) \leq \frac{1}{4\alpha}\left(4+\log_+\log\frac{1}{\pi_*}\right) \leq \frac{d}{2}\left(4+\log d\right)$$

of Corollary 2.2.7 compares well with the lower bound

$$T_2 \geq \frac{d}{2}(2+\log d).$$

EXAMPLE 2.2.4: Consider now $|x| = \sum_1^d x_i$, that is, the number of 1's in the chain in the preceding example, as random variable taking values in $\mathcal{X}_0 = \{0,\dots,d\}$. Clearly, this defines a Markov chain on $\mathcal{X}_0$ with stationary measure

$$\pi_0(j) = \theta^j(1-\theta)^{d-j}\binom{d}{j}$$

and kernel

$$K_0(i,j) = \begin{cases} 0 & \text{if } |i-j|>1 \\ (1-\theta)(1-i/d) & \text{if } j=i+1 \\ \theta i/d & \text{if } j=i-1 \\ (1-\theta)i/d+\theta(1-i/d) & \text{if } i=j. \end{cases}$$

All the eigenvalues of $I-K_0$ are also eigenvalues of $I-K$. It follows that $\lambda_0 \geq 1/d$. Furthermore, the function $F: i \to c_0[d(1-\theta)-i]$ with $c_0 = (d\theta(1-\theta))^{-1/2}$ is an eigenfunction with eigenvalue $1/d$ and $\|F\|_2 = 1$. Hence, $\lambda_0 = 1/d$. Concerning α_0, all we can say is that

$$\alpha_0 \geq \frac{1-2\theta}{d\log[(1-\theta)/\theta]}.$$

When $\theta = 1/2$ this inequality and Lemma 2.2.2 show that $\alpha_0 = 1/(2d) = \lambda/2$.

The next result allows comparison of the spectral gaps and log-Sobolev constants of two chains defined on different state spaces.

Lemma 2.2.12 *Let (K,π), (K',π') be two Markov chains defined respectively on the finite sets $\mathcal{X}$ and $\mathcal{X}'$. Assume that there exists a linear map*

$$\ell^2(\mathcal{X},\pi) \to \ell^2(\mathcal{X}',\pi') : f \to \tilde{f}$$

and constants $A, B, a > 0$ such that, for all $f \in \ell^2(\mathcal{X},\pi)$

$$\mathcal{E}'(\tilde{f},\tilde{f}) \le A\mathcal{E}(f,f) \quad \text{and} \quad a\mathrm{Var}_\pi(f) \le \mathrm{Var}_{\pi'}(\tilde{f}) + B\mathcal{E}(f,f)$$

then

$$\frac{a\lambda'}{A+B\lambda'} \le \lambda\,.$$

Similarly, if

$$\mathcal{E}'(\tilde{f},\tilde{f}) \le A\mathcal{E}(f,f) \quad \text{and} \quad a\mathcal{L}_\pi(f) \le \mathcal{L}_{\pi'}(\tilde{f}) + B\mathcal{E}(f,f),$$

then

$$\frac{a\alpha'}{A+B\alpha'} \le \alpha.$$

In particular, if $\mathcal{X} = \mathcal{X}'$, $\mathcal{E}' \le A\mathcal{E}$ and $a\pi \le \pi'$, then

$$\frac{a\lambda'}{A} \le \lambda, \quad \frac{a\alpha'}{A} \le \alpha.$$

PROOF: The two first assertions follow from the variational definitions of λ and α. For instance, for λ we have

$$\begin{aligned} a\mathrm{Var}_\pi(f) &\le \mathrm{Var}_{\pi'}(\tilde{f}) + B\mathcal{E}(f,f) \\ &\le \frac{1}{\lambda'}\mathcal{E}'(\tilde{f},\tilde{f}) + B\mathcal{E}(f,f) \\ &\le \left(\frac{A}{\lambda'} + B\right)\mathcal{E}(f,f). \end{aligned}$$

The desired inequality follows.

To prove the last assertion, use $a\pi \le \pi'$ and the formula

$$\mathrm{Var}_\pi(f) = \min_{c\in\mathbb{R}} \sum_x |f(x) - c|^2\pi(x)$$

to see that $a\mathrm{Var}_\pi(f) \le \mathrm{Var}_{\pi'}(f)$. The inequality between log-Sobolev constants follows from $\xi\log\xi - \xi\log\zeta - \xi + \zeta \ge 0$ for all $\xi,\zeta > 0$ and

$$\begin{aligned} \mathcal{L}_\pi(f) &= \sum_x \left(|f(x)|^2\log|f(x)|^2 - |f(x)|^2\log\|f\|_2^2 - |f(x)|^2 + \|f\|_2^2\right)\pi(x) \\ &= \min_{c>0}\sum_x \left(|f(x)|^2\log|f(x)|^2 - |f(x)|^2\log c - |f(x)|^2 + c\right)\pi(x). \end{aligned}$$

This useful observation is due to Holley and Stroock [50].

EXAMPLE 2.2.5: Let $\mathcal{X} = \{0,1\}^n$ and set $|x-y| = \sum_i |x_i - y_i|$. Let $\tau : \mathcal{X} \to \mathcal{X}$ be the map defined by $\tau(x) = y$ where $y_i = x_{i-1}$, $1 < i \leq n$, $y_1 = x_n$. Consider the chain

$$K(x,y) = \begin{cases} 1/(n+1) & \text{if } |x-y| = 1 \\ 1/(n+1) & \text{if } y = \tau(x) \\ 0 & \text{oherwise.} \end{cases}$$

It is not hard to check that the uniform distribution $\pi \equiv 2^{-n}$ is the stationary measure of K. Observe that K is neither reversible nor an invariant chain on the group $\{0,1\}^n$. We will study this chain by comparison with the classic chain K' whose kernel vanishes if $|x-y| \neq 1$ and is equal to $1/n$ if $|x-y| = 1$. These two chains have the same stationary measure $\pi \equiv 2^{-n}$. Obviously the Dirichlet forms $\mathcal{E}'$ and $\mathcal{E}$ satisfy

$$\mathcal{E}' \leq \frac{n+1}{n}\mathcal{E}(f,f).$$

Applying Lemma 2.2.12, and using the known values $\lambda' = 2/n$, $\alpha' = 1/n$ of the spectral gap and log Sobolev constant of the chain K', we get

$$\lambda \geq \frac{2}{n+1}, \quad \alpha \geq \frac{1}{n+1}.$$

To obtain upper bounds, we use the test function $f = \sum_i (x_i - 1/2)$. This has $\pi(f) = 0$. Also

$$\mathcal{E}(f,f) = \frac{n}{n+1}\,\mathcal{E}'(f,f) = \frac{n}{n+1}\,\frac{2}{n}\,\mathrm{Var}_\pi(f).$$

The first equality follows from the fact that $f(\tau(x)) = f(x)$. The second follows from the fact that f is an eigenvalue of $I - K'$ associated with the eigenvalue $2/n$ (in fact, one can check that f is an eigenfunction of K itself). Hence $\lambda \leq 2/(n+1)$. This implies

$$\lambda = \frac{2}{n+1}, \quad \alpha = \frac{1}{n+1}.$$

Applying Theorem 2.2.5 we get

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for} \quad t = \frac{n+1}{4}\,(2c + \log n)\,, \quad c > 0.$$

The test function f used above has $\|f\|_\infty = n/2$ and $\|f\|_2^2 = n/4$ and is an eigenfunction associated with λ. Hence

$$\max_x \|h_t^x - 1\|_2 = \|H_t - \pi\|_{2\to\infty} \geq \frac{\|H_t f\|_\infty}{\|f\|_2} = n^{1/2} e^{-2t/(n+1)}.$$

This proves the sharpness of our upper bound. A lower bound in ℓ^1 can be obtained by observing that the number of 1's in x, that is $|x|$, evolves has a Markov chain on $\{0,\dots,n\}$ which is essentially the classic Ehrenfest's urn Markov chain.

This example generalizes easily as follows. The permutaion τ can be replaced by any other permutation without affecting the analysis presented above. We can also pick at random among several permutations of the coordinates. This will simply change the factor of comparison between $\mathcal{E}$ and $\mathcal{E}'$.

We end this section with a result that bounds α in terms of $\max_x \|h_t^x - 1\|_2 = \|H_t - \pi\|_{2\to\infty}$. See [29] for a proof. Similar results can be found in [8, 16]

Theorem 2.2.13 *Assume that (K, π) is reversible. Fix $2 < q \le +\infty$ and assume that t_q, M_q satisfy $\|H_{t_q} - \pi\|_{2\to q} \le M_q$. Then*

$$\alpha \ge \frac{(1-\frac{2}{q})\lambda}{2(\lambda t_q + \log M_q + \frac{q-2}{q})} .$$

In particular, if $q = \infty$ and t is such that $\max_x \|h_t^x - 1\|_2 \le M$, we have

$$\alpha \ge \frac{\lambda}{2(\lambda t + \log M)} .$$

EXAMPLE 2.2.6: Consider the nearest neighbor chain K on $\{0, \dots, n\}$ with loops at the ends. Then $\lambda = 1 - \cos\frac{\pi}{n+1}$. At the end of Section 2.1 it is proved that

$$\|H_t - \pi\|_{2\to\infty}^2 = \max_x \|h_t^x - 1\|_2^2 \le 2e^{-4t/(n+1)^2}\left(1 + \sqrt{(n+1)^2/4t}\right).$$

Thus, for $t = \frac{1}{2}(n+1)^2$, $\|H_t - \pi\|_{2\to\infty} \le 1$. Using this and $\lambda \ge 2/(n+1)^2$ in Theorem 2.2.13 give

$$\frac{1}{2(n+1)^2} \le \alpha \le \frac{1}{2}\left(1 - \cos\frac{\pi}{n+1}\right) = \frac{\pi^2}{4(n+1)^2} + O(1/n^4).$$

The exact value of α is not known.

2.3 Nash inequalities

A Nash inequality for the finite Markov chain (K, π) is an inequality of the type

$$\forall\, f \in \ell^2(\mathcal{X}, \pi), \quad \|f\|_2^{2(1+2/d)} \le C\left(\mathcal{E}(f,f) + \frac{1}{T}\|f\|_2^2\right)\|f\|_1^{4/d}$$

where d, C, T are constants depending on K. The size of these constants is of course crucial in our applications. This inequality implies (in fact, is equivalent to)

$$H_t(x,y) \le B(d)\pi(y)\,(C/t)^{d/2} \ \text{ for } \ 0 < t \le T$$

where $B(d)$ depends only on d and d, C, T are as above. This is discussed in detail in this section. Nash inequalities have received considerable attention in recent years. I personally learned about them from Varopoulos [78]. Their use is emphasized in [11]. Applications to finite Markov chains are presented in [28], with many examples. See also [69]

2.3.1 Nash's argument for finite Markov chains I

Nash introduced his inequality in [64] to study the decay of the heat kernel of certain parabolic equations in Euclidean space. His argument only uses the formula 2.1.2 for the time derivative of $u(t) = \|H_t f\|_2^2$ which reads $u'(t) = -2\mathcal{E}(H_t f, H_t)$. This formula shows that any functional inequality between the ℓ^2 norm of g and the Dirichlet form $\mathcal{E}(g,g)$ (for all g, thus $g = H_t f$) can be translated into a differential inequation involving u. Namely, assume that the Dirichlet form $\mathcal{E}$ satisfies the inequality

$$\forall\, g,\ \ \mathrm{Var}_\pi(g)^{1+2/d} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

Then fix f satisfying $\|f\|_1 = 1$ and set $u(t) = \|H_t(f-\pi(f))\|_2^2 = \mathrm{Var}_\pi(H_t f)$. In terms of u, the Nash's inequality above gives

$$\forall\, t,\ \ u(t)^{1+2/d} \le -\frac{C}{2}u'(t),$$

since $\|f\|_1 = 1$ implies $\|H_t f\|_1 \le 1$ for all $t > 0$. Setting $v(t) = \frac{dC}{4}u(t)^{-2/d}$ this differential inequality implies $v'(t) \ge 1$. Thus $v(t) \ge t$ (because $v(0) \ge 0$). Finally,

$$\forall\, t > 0,\ \ u(t) \le \left(\frac{dC}{4t}\right)^{d/2}.$$

Taking the supremum over all functions f with $\|f\|_1 = 1$ yields

$$\forall\, t,\ \ \|H_t - \pi\|_{1\to 2} \le \left(\frac{dC}{4t}\right)^{d/4}.$$

The same applies to adjoint H_t^* and thus

$$\forall\, t > 0,\ \ \|H_t - \pi\|_{2\to\infty} \le \left(\frac{dC}{4t}\right)^{d/4}.$$

Finally, using $H_t - \pi = (H_{t/2} - \pi)(H_{t/2} - \pi)$, we get

$$\forall\, t > 0,\ \ \|H_t - \pi\|_{1\to\infty} \le \left(\frac{dC}{2t}\right)^{d/2}$$

which is the same as

$$|h_t(x,y) - 1| \le (dC/2t)^{d/2}.$$

Theorem 2.3.1 *Assume that the finite Markov chain (K,π) satisfies*

$$\forall\, g \in \ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}. \tag{2.3.1}$$

Then

$$\forall\, t > 0,\ \ \|h_t^x - 1\|_2 \le \left(\frac{dC}{4t}\right)^{d/4}$$

and

$$\forall\, t > 0,\ \ |h_t(x,y) - 1| \le \left(\frac{dC}{2t}\right)^{d/2}.$$

Let us discuss what this says. First, the hypothesis 2.3.1 and Jensen's inequality imply $\forall\ g \in \ell^2(\pi)$, $\mathrm{Var}_\pi(g) \leq C\mathcal{E}(g,g)$. This is a Poincaré inequality and it shows that $\lambda \geq 1/C$. Thus, the conclusion of Theorem 2.3.1 must be compared with

$$\forall\, t > 0, \;\; \|h_t^x - 1\|_2 \leq \pi(x)^{-1/2} e^{-t/C} \tag{2.3.2}$$

which follows from Corollary 2.1.5 when $\lambda \geq 1/C$. This last inequality looks better than the conclusion of Theorem 2.3.1 as it gives an exponential rate. However, Theorem 2.3.1 gives $\|h_t^x - 1\|_2 \leq 1$ for $t = dC/4$ whereas, for the same t, the right hand side of (2.3.2) is equal to $\pi(x)^{-1/2}e^{-d/4}$. Thus, if d is small and $1/\pi(x)$ large, the conclusion of Theorem 2.3.1 improves up on (2.3.2) at least for relatively small value of t. Assume for instance that (2.3.1) holds with $C = A/\lambda$ where we think of A as a numerical constant. Then, for $\theta = dA/(4\lambda)$, $\|H_\theta - \pi\|_{2\to\infty} = \max_x \|h_\theta^x - 1\|_2 \leq 1$. Hence, for $t = s + \theta = s + dA/(4\lambda)$

$$\begin{aligned}
\|h_t^x - 1\|_2 &\leq \|(H_s - \pi)(H_\theta - \pi)\|_{2\to\infty} \\
&\leq \|H_s - \pi\|_{2\to 2}\|H_\theta - \pi\|_{2\to\infty} \\
&\leq e^{-\lambda s}.
\end{aligned}$$

This yields

Corollary 2.3.2 *If (K,π) satisfies (2.3.1) with some constants $C, d > 0$. Then $\lambda \geq 1/C$ and*

$$\forall\, t > 0, \;\; \|h_t^x - 1\|_2 \leq \min\left\{(dC/4t)^{d/4}, e^{-(t-\frac{dC}{4})\lambda}\right\}.$$

If (K,π) is reversible, then K is self-adjoint on $\ell^2(\pi)$ and $1-\lambda$ is the second largest eigenvalue of K. Consider an eigenfunction ψ for the eigenvalue $1-\lambda$, normalized so that $\max|\psi| = 1$. Then,

$$\begin{aligned}
\max_x \|H_t^x - \pi\|_1 &= \max_{\|f\|_\infty \leq 1} \|(H_t - \pi)f\|_\infty \\
&\geq \|(H_t - \pi)\psi\|_\infty \\
&= e^{-t\lambda}.
\end{aligned}$$

Hence

Corollary 2.3.3 *Assume that (K,π) is a reversible Markov chain. Then*

$$e^{-\lambda t} \leq \max_x \|H_t^x - \pi\|_1.$$

Furthermore, if (K,π) satisfies (2.3.1) with $C = A/\lambda$ then

$$e^{-\lambda t} \leq \max_x \|H_t^x - \pi\|_1 \leq 2\, e^{-\lambda t + \frac{dA}{4}}$$

for all $t > 0$.

This illustrates well the strength of Nash inequalities. They produce sharp results in certain circumstances where the time needed to reach stationarity is approximatively $1/\lambda$.

2.3.2 Nash's argument for finite Markov chains II

We now presents a second version of Nash's argument for finite Markov chains which turns out to be often easier to use than Theorem 2.3.1 and Corollary 2.3.2.

Theorem 2.3.4 *Assume that the finite Markov chain (K,π) satisfies*

$$\forall\, g \in \ell^2(\pi),\ \ \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d}. \tag{2.3.3}$$

Then

$$\forall\, t \le T,\ \ \|h_t^x\|_2 \le e\left(\frac{dC}{4t}\right)^{d/4}$$

and

$$\forall\, t \le T,\ \ h_t(x,y) \le e\left(\frac{dC}{2t}\right)^{d/2}.$$

The idea behind Theorem 2.3.4 is that Nash inequalities are most useful to capture the behavior of the chain for relatively small time, i.e., time smaller than T. In contrast with (2.3.1) the Nash inequality (2.3.3) implies no lower bound on the spectral gap. This is an advantage as it allows (2.3.3) to reflect the early behavior of the chain without taking into account the asymptotic behavior. This is well illustrated by two examples that will be treated later in these notes. Consider the natural chain on a square grid $\mathcal{G}_n$ of side length n and the natural chain on the n-dog $\mathcal{D}_n$ obtained by gluing together two copies of $\mathcal{G}_n$ at one of their corners. On one hand the spectral gap of $\mathcal{G}_n$ is of order $1/n^2$ whereas the spectral gap of $\mathcal{D}_n$ is of order $1/[n^2\log n]$ (these facts will be proved later on). On the other hand, $\mathcal{G}_n$ and $\mathcal{D}_n$ both satisfy a Nash inequality of type (2.3.3) with C and T of order n^2. That is, the chains on $\mathcal{G}_n$ and $\mathcal{D}_n$ have similar behaviors for t less than n^2 whereas their asymptotic behavior as t goes to infinity are different. This is not surprising since the local structure of these two graphs are the same. For $\mathcal{D}_n$ a constant C of order $n^2\log n$ is necessary for an inequality of type (2.3.1) to hold true.

PROOF OF THEOREM 2.3.4: Fix f satisfying $\|f\|_1 = 1$ and set

$$u(t) = e^{-2t/T}\|H_t f\|_2^2.$$

Then

$$u'(t) = -2e^{-2t/T}\left(\mathcal{E}(H_t f, H_t f) + \frac{1}{T}\|H_t f\|_2^2\right).$$

Thus, Nash's argument yields

$$u(t) \le \left(\frac{dC}{4t}\right)^{d/2}$$

which implies

$$\|H_t\|_{1\to 2} \le e^{t/T}\left(\frac{dC}{4t}\right)^{d/4}.$$

The announced results follow since

$$\max_x \|h_t^x\|_2 = \|H_t^*\|_{1\to 2} \le e^{t/T}\left(\frac{dC}{4t}\right)^{d/4}$$

by the same argument applied to H_t^*.

Corollary 2.3.5 *Assume that* (K,π) *satisfies* (2.3.3) *and has spectral gap* λ. *Then for all* $c \ge 0$ *and all* $0 < t_0 \le T$,

$$\|h_t^x - 1\|_2 \le e^{1-c}$$

and

$$|h_{2t}(x,y) - 1| \le e^{2-2c}$$

for

$$t = t_0 + \frac{1}{\lambda}\left(\frac{d}{4}\log\left(\frac{dC}{4t_0}\right) + c\right).$$

PROOF: Write $t = s + t_0$ with $t_0 \le T$ and

$$\begin{aligned} \|h_t^x - 1\|_2 &\le \|(H_s - \pi)H_{t_0}\|_{2\to\infty} \\ &\le \|H_s - \pi\|_{2\to 2}\|H_{t_0}\|_{2\to\infty} \\ &\le e(dC/4t_0)^{d/4}\, e^{-\lambda s}. \end{aligned}$$

The result easily follows.

In practice, a "good" Nash inequality is (2.3.3) with a small value of d and $C \approx T$. Indeed, if (2.3.3) holds with, say $d = 4$ and $C = T$, then taking $t_0 = T$ in Corollary 2.3.5 yields

$$\|h_t^x - 1\|_2 \le e^{1-c} \text{ for } t = T + c/\lambda.$$

We now give a simple example that illustrates the strength of a good Nash inequality.

EXAMPLE 2.3.1: Consider the Markov chain on $\mathcal{X} = \{-n,\dots,n\}$ with Kernel $K(x,y) = 0$ unless $|x-y| = 1$ or $x = y = \pm n$ in which cases $K(x,y) = 1/2$. This is an irreducible chain which is reversible with respect to $\pi \equiv (2n+1)^{-1}$. The Dirichlet form of this chain is given by

$$\mathcal{E}(f,f) = \frac{1}{2n+1}\sum_{-n}^{n-1} |f(i+1) - f(i)|^2.$$

For any $u, v \in \mathcal{X}$, and any function f, we have

$$|f(v) - f(u)| \le \sum_{i,i+1 \text{ between } u,v} |f(i+1) - f(i)|.$$

Hence, if f is not of constant sign,

$$\|f\|_\infty \le \sum_{-n}^{n-1} |f(i+1) - f(i)|.$$

To see this take u to be such that $\|f\|_\infty = f(u)$ and v such that $f(v)f(u) \le 0$ so that $|f(u) - f(v)| \ge |f(u)|$. Fix a function g such that $\pi(g > 0) \le 1/2$ and $\pi(g < 0) \le 1/2$ (i.e., 0 is a median of g). Set $f = \operatorname{sgn}(g)|g|^2$. Then f changes sign. Observe also that

$$\begin{aligned} |f(i+1) - f(i)| &= |\operatorname{sgn}(g(i+1))g(i+1)^2 - \operatorname{sgn}(g(i))g(i)^2| \\ &\le |g(i+1) - g(i)|(|g(i+1)| + |g(i)|). \end{aligned}$$

Hence

$$\begin{aligned} \|f\|_\infty &\le \sum_{-n}^{n-1} |f(i+1) - f(i)| \\ &\le \sum_{-n}^{n-1} |g(i+1) - g(i)|(|g(i+1)| + |g(i)|) \\ &\le \left(\sum_{-n}^{n-1} |g(i+1) - g(i)|^2\right)^{1/2} \left(\sum_{-n}^{n-1} (|g(i+1)| + |g(i)|)^2\right)^{1/2} \\ &\le 2^{1/2}(2n+1)\mathcal{E}(g,g)^{1/2}\|g\|_2. \end{aligned}$$

That is

$$\|g\|_\infty^2 \le 2^{1/2}(2n+1)\mathcal{E}(g,g)^{1/2}\|g\|_2.$$

It follows that

$$\begin{aligned} \|g\|_2^4 &\le \|g\|_\infty^2 \|g\|_1^2 \\ &\le 2^{1/2}(2n+1)\mathcal{E}(g,g)^{1/2}\|g\|_2\|g\|_1. \end{aligned}$$

Hence for any g with median 0,

$$\|g\|_2^6 \le 2(2n+1)^2\mathcal{E}(g,g)\|g\|_1^4.$$

For any f with median c, we can apply the above to $g = f - c$ to get

$$\|f - c\|_2^6 \le 2(2n+1)^2\mathcal{E}(f,f)\|f - c\|_1^4 \le 2(2n+1)^2\mathcal{E}(f,f)\|f\|_1^4.$$

Hence

$$\forall f, \quad \mathrm{Var}_\pi(f)^3 \le 2(2n+1)^2\mathcal{E}(f,f)\|f\|_1^4.$$

This is a Nash inequality of type (2.3.1) with $C = 2(2n+1)^2$ and $d = 1$. It implies that

$$\lambda \ge \frac{1}{2(2n+1)^2}$$

and, by Theorem 2.3.1 and Corollary 2.3.2

$$\forall\, t > 0, \qquad \|h_t^x - 1\|_2 \le \left(\frac{(2n+1)^2}{2t}\right)^{1/4}$$

and

$$\forall\, c > 0, \quad \|h_t^x - 1\|_2 \le e^{-c} \quad \text{with } t = \frac{1}{2(2n+1)^2}(4+c).$$

The test function $f(i) = \operatorname{sgn}(i)|i|$ shows that

$$\lambda \le \frac{12}{(2n+1)^2}$$

(in fact $\lambda = 1 - \cos(\pi/(2n+1))$). By Corollary 2.3.3 it follows that

$$e^{-\frac{12t}{(2n+1)^2}} \le \max_{\mathcal{X}} \|h_t^x - 1\|_1 \le 2e^{-\frac{t}{2(2n+1)^2} + \frac{1}{4}}.$$

This shows that a time of order n^2 is necessary and sufficient for approximate equilibrium. This conclusion must be compare with

$$\|h_t^x - 1\|_1 \le \sqrt{2n+1}\, e^{-\frac{t}{2(2n+1)^2}}$$

which follows by using only the spectral gap estimate $\lambda \ge 1/(2(2n+1)^2)$ and Corollary 2.1.5. This last inequality only shows that a time of order $n^2 \log n$ is sufficient for approximate equilibrium.

2.3.3 Nash inequalities and the log-Sobolev constant

Thanks to Theorem 2.2.13 and Nash's argument it is possible to bound the log-Sobolev constant α in terms of a Nash inequality.

Theorem 2.3.6 *Let (K,π) be a finite reversible Markov chain.*

1. *Assume that (K,π) satisfies* (2.3.1), *that is,*

$$\forall\, g \in \ell^2(\pi), \quad \operatorname{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

Then the log-Sobolev constant α of the chain is bounded below by

$$\alpha \ge \frac{2}{dC}.$$

2. *Assume instead that (K,π) satisfies* (2.3.3), *that is,*

$$\forall\, g \in \ell^2(\pi), \quad \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d},$$

and has spectral gap λ. Then the log-Sobolev constant α is bounded below by

$$\alpha \ge \frac{\lambda}{2\left[1 + \lambda t_0 + \frac{d}{4}\log\left(\frac{dC}{4t_0}\right)\right]}$$

for any $0 < t_0 \le T$.

PROOF: For the first statement, observe that Theorem 2.3.1 gives $\|H_t - \pi\|_{2\to\infty} \leq 1$ for $t = dC/4$. Pluging this into Theorem 2.2.13 yields $\alpha \geq 2/(dC)$, as desired.

For the second inequality use Theorem 2.3.3 with $t = t_0 \leq T$ and Theorem 2.2.13.

EXAMPLE 2.3.2: Consider the Markov chain of Example 2.3.1 on $\mathcal{X} = \{-n, \dots, n\}$ with Kernel $K(x,y) = 0$ unless $|x-y| = 1$ or $x = y = \pm n$ in which cases $K(x,y) = 1/2$. We have proved that it satisfies the Nash inequality

$$\forall f, \quad \mathrm{Var}_\pi(f)^3 \leq 2(2n+1)^2 \mathcal{E}(f,f)\|f\|_1^4$$

of type (2.3.1) with $C = 2(2n+1)^2$ and $d = 1$. Hence Theorem 2.3.6 yields

$$\alpha \geq \frac{1}{(2n+1)^2}.$$

2.3.4 A converse to Nash's argument

Carlen et al. [11] found that there is a converse to Nash's argument. We now present a version of their result.

Theorem 2.3.7 *Assume that (K, π) is reversible and satisfies*

$$\forall\, t \leq T, \ \ \|H_t\|_{1\to 2} \leq \left(\frac{C}{t}\right)^{d/4}.$$

Then

$$\forall \overset{f}{g} \in \ell^2(\pi), \quad \|f\|_2^{2(1+2/d)} \leq C' \left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2 \right) \|f\|_1^{4/d}$$

with $C' = 2^{2(1+2/d)}C$.

PROOF: Fix f with $\|f\|_1 = 1$ and write, for $0 < t \leq T$,

$$\begin{aligned} \|f\|_2^2 &= \|H_t f\|_2^2 - \int_0^t \partial_s \|H_s f\|_2^2 ds \\ &= \|H_t f\|_2^2 + 2\int_0^t \mathcal{E}(H_s f, H_s f) ds \\ &\leq (C/t)^{d/2} + 2t\mathcal{E}(f,f). \end{aligned}$$

The inequality uses the hypothesis (which implies $\|H_t f\|_2 \leq (C/t)^{d/4}$ because $\|f\|_1 \leq 1$) and the fact that $t \to \mathcal{E}(H_t f, H_t f)$ is nonincreasing, a fact that uses reversibility. This can be proved by writing

$$\mathcal{E}(H_t f, H_t f) = \|(I-K)^{1/2} H_t f\|_2^2 \leq \|(I-K)^{1/2} f\|_2^2 = \mathcal{E}(f,f).$$

It follows that

$$\|f\|_2^2 \leq (C/t)^{d/2} + 2t\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)$$

for all $t > 0$. The right-hand side is a minimum for

$$\frac{dC^{d/2}}{2} t^{-(1+d/2)} = 2\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)$$

and the minimum is

$$\left[(2/d)^{1/(1+2/d)} + (d/2)^{1/(1+d/2)}\right]\left[2C\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)\right]^{1/(1+2/d)}.$$

This yields

$$\|f\|_2^{2(1+2/d)} \le B\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)$$

with

$$\begin{aligned} B &= 2C\left[(2/d)^{1/(1+2/d)} + (d/2)^{1-1/(1+2/d)}\right]^{1+2/d} \\ &= 2C(1+2/d)(1+d/2)^{2/d} \le 2^{2+2/d}C. \end{aligned}$$

2.3.5 Nash inequalities and higher eigenvalues

We have seen that a Poincaré inequality is equivalent to a lower bound on the spectral gap λ (i.e., the smallest non-zero eigenvalue of $I-K$). It is interesting to note that Nash inequalities imply bounds on higher eigenvalues. Compare with [14].

Let (K,π) be a finite reversible Markov chain. Let $1 = \lambda_0 \le \lambda_1 \le \dots \le \lambda_{n-1}$ be the eigenvalues of $I-K$ and

$$N(s) = N_K(s) = \#\{i \in \{0,\dots,n-1\} : \lambda_i \le s\}, \quad s \ge 0,$$

be the eigenvalue counting function. Thus, N is a step function with $N(s) = 1$ for $0 \le s < \lambda_1$ if (K,π) is irreducible. It is easy to relate the function N to the trace of the semigroup $H_t = e^{-t(I-K)}$. Since (K,π) is reversible, we have

$$\zeta(t) = \sum_x h_t(x,x)\pi(x) = \sum_x \|h_{t/2}^x\|_2^2\pi(x) = \sum_{i=0}^{n-1} e^{-t\lambda_i}.$$

If $\lambda_i \le 1/t$ then $e^{-t\lambda_i} \ge e^{-1}$. Hence

$$N(1/t) \le e\zeta(t).$$

Now, it is clear that Theorems 2.3.1, 2.3.4 give upper bounds on ζ in terms of Nash inequalities.

Theorem 2.3.8 *Let (K,π) be a finite reversible Markov chain.*

1. *Assume that* (K, π) *satisfies* (2.3.1), *that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

Then the counting function N *satisfies*

$$N(s) \le 1 + e(dCs/2)^{d/2}$$

for all $s \ge 0$.

2. *Assume instead that* (K, π) *satisfies* (2.3.3), *that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d}.$$

Then

$$N(s) \le e^3(dCs/2)^{d/2}$$

for all $s \ge 1/T$.

Clearly, if $M(s)$ is a continuous increasing function such that $N(s) \le M(s)$, $s \ge 1/T$, then

$$\lambda_i = \max\{s : N(s) \le i\} \ge M^{-1}(i+1)$$

for all $i > M(1/T) - 1$. Hence, we obtain

Corollary 2.3.9 *Let* (K, π) *be a finite reversible Markov chain. Let* $1 = \lambda_0 \le \lambda_1 \le \ldots \le \lambda_{n-1}$ *be the eigenvalues of* $I - K$.

1. *Assume that* (K, π) *satisfies* (2.3.1), *that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

Then

$$\lambda_i \ge \frac{2i^{2/d}}{e^{2/d}dC}$$

for all $i \in 1, \ldots, n-1$.

2. *Assume instead that* (K, π) *satisfies* (2.3.3), *that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d}.$$

Then

$$\lambda_i \ge \frac{2(i+1)^{2/d}}{e^{6/d}dC}$$

for all $i > e^3(dC/(2T))^{d/2} - 1$.

EXAMPLE 2.3.3: Assume that (K, π) is reversible, has spectral gap λ, and satisfies the Nash inequality (2.3.1) with $C = A/\lambda$ and some d, where we think of A as a numerical constant (e.g., $A = 100$) and d as fixed. Then, the corollary above says that

$$\lambda_i \geq c\lambda i^{2/d}$$

for all $0 \leq i \leq n-1$ with $c^{-1} = e^{2/d} dA$.

EXAMPLE 2.3.4: For the natural graph structure on $\mathcal{X} = \{-n, \ldots, n\}$, we have shown in Example 2.3.1 that the Nash inequality

$$\mathrm{Var}_\pi(f)^3 \leq 2(2n+1)^2 \mathcal{E}(f,f) \|f\|_1^4$$

holds. Corollary 2.3.9 gives

$$\lambda_j \geq \left(\frac{j}{e^2(2n+1)} \right)^2 .$$

In this case, all the eigenvalues are known. They are given by

$$\lambda_j = 1 - \cos \frac{\pi j}{2n+1}, \quad 0 \leq j \leq 2n.$$

This compares well with our lower bound.

EXAMPLE 2.3.5: For a square grid on $\mathcal{X} = \{0, \ldots, n\}^2$, we will show later (Theorem 3.3.14) that

$$\mathrm{Var}_\pi(f)^2 \leq 64(n+1)^2 \mathcal{E}(f,f) \|f\|_1^2.$$

¿From this and corollary 2.3.9 we deduce

$$\lambda_i \geq \frac{i}{e 2^7 (n+1)^2}$$

for all $0 \leq i \leq (n+1)^2 - 1$. One can show that this lower bound is of the right order of magnitude for all i, n. Indeed the eigenvalues of this chain are the numbers

$$1 - \frac{1}{2} \left(\cos \frac{\pi \ell}{n+1} + \cos \frac{\pi k}{n+1} \right), \quad \ell, k \in \{0, \ldots, n\}$$

which are distributed roughly like

$$\frac{\ell^2 + k^2}{(n+1)^2}, \quad \ell, k \in \{0, \ldots, n\}$$

and we have

$$\# \left\{ (\ell, k) \in \{0, \ldots, n\}^2 : \ell^2 + k^2 \leq j \right\} \simeq j.$$

2.3.6 Nash and Sobolev inequalities

Nash inequalities are closely related to the better known Sobolev inequalities (for some fixed $d > 2$)

$$\|f - \pi(f)\|_{2d/(d-2)}^2 \leq C\mathcal{E}(f,f), \tag{2.3.4}$$

$$\|f\|_{2d/(d-2)}^2 \leq C\left\{\mathcal{E}(f,f) + \frac{1}{T}\|f\|_2^2\right\}. \tag{2.3.5}$$

Indeed, the Hölder inequality

$$\|f\|_2^{2(1+2/d)} \leq \|f\|_{2d/(d-2)}^2 \|f\|_1^{4/d}$$

shows that the Sobolev inequality (2.3.4) (resp. (2.3.5)) implies the Nash inequality (2.3.1) (resp. (2.3.3)) with the same constants d, C, T. The converse is also true. (2.3.1) (resp. (2.3.3)) implies (2.3.4) (resp. (2.3.5)) with the same d, T and a C that differ only by a numerical multiplicative factor for large d. See [9].

We now give a complete argument showing that (2.3.1) implies (2.3.4), in the spirit of [9]. The same type of argument works for (2.3.3)) implies (2.3.5).

For any function $f \geq 0$ and any k, we set $f_k = (f - 2^k)_+ \wedge 2^k$ where $(t)_+ = \max\{0, t\}$ and $t \wedge s = \min\{t, s\}$. Thus, f_k has support in $\{x : f(x) \geq 2^k\}$, $f_k(x) = 2^k$ if $x \in \{z : f(z) \geq 2^{k+1}\}$ and $f_k = f - 2^k$ on $\{x : 2^k \leq f \leq 2^{k+1}\}$.

Lemma 2.3.10 *Let K be a finite Markov chain with stationary measure π. With the above notation, for any function f,*

$$\sum_k \mathcal{E}(|f|_k, |f|_k) \leq 2\mathcal{E}(f,f).$$

PROOF: Since $\mathcal{E}(|f|, |f|) \leq \mathcal{E}(f,f)$, we can assume that $f \geq 0$. We can also assume that $K(x,y)\pi(x)$ is symmetric (if not use $\frac{1}{2}(K(x,y)\pi(x) + K(y,x)\pi(y))$). Observe that $|f_k(x) - f_k(y)| \leq |f(x) - f(y)|$ for all x, y. Write

$$\mathcal{E}(f_k, f_k) = \sum_{\substack{x,y \\ f(x) > f(y)}} (f_k(x) - f_k(y))^2 K(x,y)\pi(x).$$

Set

$$\begin{aligned} B_k &= \{x : 2^k < f(x) \leq 2^{k+1}\}, \\ B_k^- &= \{x : f(x) \leq 2^k\}, \\ B_k^+ &= \{x : 2^{k+1} < f(x)\}. \end{aligned}$$

Then

$$\mathcal{E}(f_k, f_k) =$$

$$
\begin{aligned}
& 2^{2k} \sum_{\substack{x \in B_k^+ \\ y \in B_k^-}} K(x,y)\pi(x) + \sum_{\substack{x \in B_k, y \in B_{k+1}^- \\ f(x) > f(y)}} (f_k(x) - f_k(y))^2 K(x,y)\pi(x) \\
\leq \; & 2^{2k} \sum_{\substack{x \in B_k^+ \\ y \in B_k^-}} K(x,y)\pi(x) + \sum_{\substack{x \in B_k, y \in \mathcal{X} \\ f(x) > f(y)}} (f(x) - f(y))^2 K(x,y)\pi(x) \\
= \; & A_1(k) + A_2(k).
\end{aligned}
$$

We now bound $\sum_k A_1(k)$ and $\sum_k A_2(k)$ separately.

$$\sum_k A_1(k) = \sum_{\substack{x,y \\ f(x) > f(y)}} \sum_{k : f(y) \leq 2^k < f(x)/2} 2^{2k}.$$

For x, y fixed, let k_0 be the smallest integer such that $f(y) \leq 2^{k_0}$ and k_1 be the largest integer such that $2^{k_1} < f(x)$. Then

$$\sum_{k : f(y) \leq 2^k < f(x)/2} 2^{2k} = \sum_{k=k_0}^{k_1 - 1} 4^k = \frac{1}{3}(4^{k_1} - 4^{k_0}) \leq (f(x) - f(y))^2.$$

The last inequality follows from the elementary inequality

$$a^2 - b^2 \leq 3(a-b)^2 \text{ if } a \geq 2b \geq 0.$$

This shows that

$$\sum_k A_1(k) \leq \mathcal{E}(f,f).$$

To finish the proof, note that

$$\sum_k A_2(k) = \sum_k \sum_{\substack{x \in B_k, y \in \mathcal{X} \\ f(x) > f(y)}} (f(x) - f(y))^2 K(x,y)\pi(x) = \mathcal{E}(f,f).$$

Lemma 2.3.10 is a crucial tool for the proof of the following theorem.

Theorem 2.3.11 *Assume that (K, π) satisfies the Nash inequality* (2.3.1), *that is,*

$$\mathrm{Var}_\pi(g)^{(1+2/d)} \leq C\mathcal{E}(g,g)\|g\|_1^{4/d}$$

for some $d > 2$ and all functions g. Then

$$\|g - \pi(g)\|_{2d/(d-2)}^2 \leq B(d) C \mathcal{E}(g,g)$$

where $B(d) = 4^{6+2d/(d-2)}$.

PROOF: Fix a function g and let c denote a median of g. Consider the functions $f_\pm = (g - c)_\pm$ where $(t)_\pm = \max\{0, \pm t\}$. By definition of a median, we have

$$\pi\left(\{x : f_\pm(x) = 0\}\right) \geq 1/2.$$

For simplicity of notation, we set $f = f_+$ or f_-. For each k we define $f_k = (f - 2^k)_+ \wedge 2^k$ as in the proof of Lemma 2.3.10. Applying (2.3.1) to each f_k and setting $\pi_k = \pi(f_k)$, we obtain

$$\left[2^{2(k-1)}\pi(|f_k - \pi_k| \geq 2^{k-1})\right]^{1+2/d} \leq C\mathcal{E}(f_k, f_k)\left[2^k\pi(f \geq 2^k)\right]^{4/d}. \tag{2.3.6}$$

Observe that

$$\pi(\{x : f_k(x) = 0\}) \geq 1/2$$

and that, for any function $h \geq 0$ such that $\pi(\{x : h(x) = 0\}) \geq 1/2$ we have

$$\forall\, s \geq 0,\ \forall\, a,\ \ \pi(\{h \geq s\}) \leq 2\pi(\{|h - a| \geq s/2\}). \tag{2.3.7}$$

Indeed, if $a \leq s/2$ then $\pi(\{|h - a| \geq s/2\}) \geq \pi(h \geq s)$ whereas if $a \geq s/2$ then $\pi(\{|h - a| \geq s/2\}) \geq \pi(h = 0) \geq 1/2$. Using (2.3.6) and (2.3.7) with $h = f_k$, $a = \pi_k$ we obtain

$$\left[2^{2(k-1)}\pi(f_k \geq 2^k)\right]^{1+2/d} \leq 2^{1+2/d}C\mathcal{E}(f_k, f_k)\left[2^k\pi(f \geq 2^k)\right]^{4/d}.$$

Now, set $q = 2d/(d-2)$, $b_k = 2^{qk}\pi(\{f \geq 2^k\})$ and $\theta = d/(d+2)$. The last inequality (raised to the power θ) yields, after some algebra,

$$b_{k+1} \leq 2^{3+q}C^\theta\mathcal{E}(f_k, f_k)^\theta\, b_k^{2(1-\theta)}.$$

By Hölder's inequality

$$\begin{aligned}\sum_k b_k = \sum_k b_{k+1} &\leq 2^{3+q}C^\theta\left(\sum_k \mathcal{E}(f_k, f_k)\right)^\theta \left(\sum_k b_k^2\right)^{1-\theta} \\ &\leq 2^{3+q+\theta}C^\theta\mathcal{E}(f,f)^\theta\left(\sum_k b_k\right)^{2(1-\theta)}.\end{aligned}$$

It follows that

$$\left(\sum_k b_k\right)^{2\theta-1} \leq 2^{3+q+\theta}C^\theta\left(\sum_k \mathcal{E}(f_k, f_k)\right)^\theta.$$

Furthermore $2\theta - 1 = 2\theta/q$ and

$$\begin{aligned}(2^q - 1)\sum_k b_k &= \sum_k (2^{q(k+1)} - 2^{qk})\pi(\{f \geq 2^k\}) \\ &= \sum_k (2^{q(k+1)}\pi(\{2^k \leq g < 2^{k+1}\}) \geq \|f\|_q^q.\end{aligned}$$

Hence

$$\|f\|_q^2 \leq 2^{1+(3+q)/\theta}(2^q - 1)^{2/q}C\mathcal{E}(f,f).$$

Recall that $f = f_+$ or f_- with $f_\pm = (g-c)_\pm$, c a median of g. Note also that $\theta > 1/2$ when $d > 2$. Adding the inequalities for f_+ and f_- we obtain

$$\|g-c\|_q^2 \le 2(\|f_+\|_q^2 + \|f_-\|_q^2) \le 4^{5+q} C\mathcal{E}(g,g)$$

because $\mathcal{E}(f_+,f_+) + \mathcal{E}(f_-,f_-) \le \mathcal{E}(g,g)$. This easily implies that

$$\|g-\pi(g)\|_q^2 \le 4^{6+q} C\mathcal{E}(g,g)$$

which is the desired inequality. The constant 4^{6+q} can be improved by using a ρ-cutting, $\rho > 1$, instead of a dyadic cutting in the above argument. See [9].

2.4 Distances

This section discusses the issue of choosing a distance between probability distribution to study the convergence of finite Markov chains to their stationary measure. From the asymptotic point of view, this choice does not matter much. ¿From a more quantitative point of view, it does matter sometimes but it often happen that different choices lead to similar results. This is a phenomenon which is not yet well understood. Many aspects of this question will not be considered here.

2.4.1 Notation and inequalities

Let μ, π be two probability measures on a finite set $\mathcal{X}$ (we work with a finite $\mathcal{X}$ but most of what is going to be said holds without any particular assumption on $\mathcal{X}$). We consider π has the reference measure. Total variation is arguably the most natural distance between probability measures. It is defined by

$$\|\mu-\pi\|_{\mathrm{TV}} = \max_{A\subset\mathcal{X}} |\mu(A)-\pi(A)| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x)-\pi(x)|.$$

To see the second equality, use $\sum_x(\mu(x)-\pi(x)) = 0$. Note also that

$$\|\mu-\pi\|_{\mathrm{TV}} = \max\{|\mu(f)-\pi(f)| : |f| \le 1\}$$

where $\mu(f) = \sum_x f(x)\mu(x)$. A well known result in Markov chain theory relates total variation with the coupling technique. See, e.g., [4, 17] and the references therein.

All the others metrics or metric type quantities that we will consider are defined in terms of the density of μ with respect to π. Hence, set $h = \mu/\pi$. The ℓ^p distances

$$\|h-1\|_p = \left(\sum_{x\in\mathcal{X}} |h(x)-1|^p \pi(x)\right)^{1/p}, \quad \|h-1\|_\infty = \max_{x\in\mathcal{X}} |h(x)-1|$$

are natural choices for the analyst and will be used throughout these notes. The case $p = 2$ is of special interest as it brings in a useful Hilbert space structure.

It is known to statisticians as the chi-square distance. The case $p = 1$ is nothing else that total variation since

$$\|h-1\|_1 = \sum_{x\in\mathcal{X}} |h(x)-1|\pi(x) = \sum_{x\in\mathcal{X}} |\mu(x)-\pi(x)| = 2\|\mu-\pi\|_{\mathrm{TV}}.$$

Jensen's inequality yields a clear ordering between these distances since it implies

$$\|h-1\|_r \le \|h-1\|_s \ \text{ for all } \ 1 \le r \le s \le \infty.$$

If we view (as we may) μ, π as linear functionals $\mu, \pi : \ell^p(\pi) \to \mathbb{R}$, $f \to \mu(f), \pi(f)$, then

$$\|\mu-\pi\|_{\ell^p(\pi)\to\mathbb{R}} = \sup\{|\mu(f)-\pi(f)| : \|f\|_p \le 1\} = \|h-1\|_q$$

where q is given by $1/p+1/q = 1$ (see also Section 1.3.1). Most of the quantitative results described in these notes are stated in terms of the ℓ^2 and ℓ^∞ distances.

There are at least three more quantities that appear in the literature. The Kullback-Leibler separation, or entropy, is defined by

$$\mathrm{Ent}_\pi(h) = \sum_{x\in\mathcal{X}} [h(x)\log h(x)]\pi(x).$$

Observe that $\mathrm{Ent}_\pi(h) \ge 0$ by Jensen inequality. The Hellinger distance is

$$\begin{aligned}\|\mu-\pi\|_H &= \sum_{x\in\mathcal{X}} \left|\sqrt{h(x)}-1\right|^2 \pi(x) = \sum_{x\in\mathcal{X}} \left|\sqrt{\mu(x)}-\sqrt{\pi(x)}\right|^2 \\ &= 2\left(1-\sum_{x\in\mathcal{X}} \sqrt{h(x)}\,\pi(x)\right).\end{aligned}$$

It is not obvious why this distance should be of particular interest. However, Kakutani proved the following. Consider an infinite sequence $(\mathcal{X}_i, \pi_i)$ of probability spaces each of which carries a second probability measure $\mu_i = h_i\pi_i$ which is absolutely continuous with respect to π_i. Let $\mathcal{X} = \prod_i \mathcal{X}_i$, $\mu = \prod_i \mu_i$, $\pi = \prod_i \pi_i$. Kakutani's theorem asserts that μ is absolutely continuous with respect to π if and only if the product $\prod_i \left(\int_{\mathcal{X}_i} \sqrt{h_i}\, d\pi_i\right)$ converges.

Finally Aldous and Diaconis [4] introduces the notion of separation distance

$$d_{\mathrm{sep}}(\mu,\pi) = \max_{x\in\mathcal{X}}\{1-h(x)\}$$

in connection with strong stationary (or uniform) stopping times. See [4, 17, 19]. Observe the absence of absolute value in this definition.

The next lemma collects inequalities between the various distances introduced above. These inequalities are all well known except possibly for the strange looking lower bounds in (2.4.2) and (2.4.4). The only inequality that uses the fact that $\mathcal{X}$ is discrete and finite is the upper bound in (2.4.1).

Lemma 2.4.1 *Let π and $\mu = h\pi$ be two probability measures on a finite set $\mathcal{X}$.*

1. *Set $\pi_* = \min_{\mathcal{X}} \pi$. For $1 \le r \le s \le \infty$,*

$$\|h-1\|_r \le \|h-1\|_s \le \pi_*^{1/s-1/r}\|h-1\|_r. \tag{2.4.1}$$

Also

$$\left(\|h-1\|_2^2 - \|h-1\|_3^3\right) \le \|h-1\|_1 \le \|h-1\|_2. \tag{2.4.2}$$

2. *The Hellinger distance satisfies*

$$\frac{1}{4}\|h-1\|_1^2 \le \|\mu-\pi\|_H \le \frac{1}{4}\|h-1\|_1 \tag{2.4.3}$$

and

$$\frac{1}{8}\left(\|h-1\|_2^2 - \|h-1\|_3^3\right) \le \|\mu-\pi\|_H \le \|h-1\|_2^2 \tag{2.4.4}$$

3. *The entropy satisfies*

$$\frac{1}{2}\|h-1\|_1^2 \le \operatorname{Ent}_\pi(h) \le \frac{1}{2}\left(\|h-1\|_1 + \|h-1\|_2^2\right). \tag{2.4.5}$$

4. *The separation $d_{\mathrm{sep}}(\mu,\pi)$ satisfies*

$$\frac{1}{2}\|h-1\|_1 \le d_{\mathrm{sep}}(\mu,\pi) \le \|h-1\|_\infty. \tag{2.4.6}$$

PROOF: The inequalities in (2.4.1) are well known (the first follows from Jensen's inequality). The inequalities in (2.4.6) are elementary.

The upper bound in (2.4.5) uses

$$\forall\, u > 0, \quad (1+u)\log(1+u) \le u + \frac{1}{2}u^2$$

to bound the positive part of the entropy. The lower bound is more tricky. First, observe that

$$\forall\, u > 0, \quad 3(u-1)^2 \le (4+2u)(u\log(u) - u + 1).$$

Then take square roots and use Cauchy-Schwarz to obtain

$$3\|h-1\|_1^2 \le \|4+2h\|_1\ \|h\log(h) - h + 1\|_1.$$

Finally observe that $u\log(u) - u + 1 \ge 0$ for $u \ge 0$. Hence $\|h\log(h) - h + 1\|_1 = \operatorname{Ent}_\pi(f)$ and

$$3\|h-1\|_1^2 \le 6\operatorname{Ent}_\pi(f)$$

which gives the desired inequality. In his Ph. D. thesis, F. Su noticed the complementary bound

$$\operatorname{Ent}_\pi(h) \le \log\left(1 + \|h-1\|_2^2\right).$$

The upper bound in (2.4.3) follows from $|\sqrt{u}-1|^2 \leq |\sqrt{u}-1|(\sqrt{u}+1) = |u-1|$, $u \geq 0$. The lower bound in (2.4.3) uses $|u-1| = |\sqrt{u}-1|(\sqrt{u}+1)$, $u \geq 0$, Cauchy-Schwarz, and $\|\sqrt{h}+1\|_2^2 \leq 4$.

The upper bound in (2.4.4) follows from $|\sqrt{u}-1| \leq |u-1|$, $u \geq 0$. For the lower bound note that

$$\sqrt{1+u} \leq \begin{cases} 1 + \frac{1}{2}u - \frac{1}{16}u^2 & \text{for } -1 \leq u \leq 1 \\ 1 + \frac{1}{2}u \leq 1 + \frac{1}{2}u - \frac{1}{16}u^2 + \frac{1}{16}u^3 & \text{for } 1 \leq u. \end{cases}$$

It follows that

$$\forall\, , u \geq -1, \quad \sqrt{1+u} \leq 1 + \frac{1}{2}u - \frac{1}{16}u^2 + \frac{1}{16}|u|^3.$$

Now, $\|\mu - \pi\|_H = 2(1 - \|\sqrt{h}\|_1) = 2(1 - \|\sqrt{1+(h-1)}\|_1)$. Hence

$$\|\mu - \pi\|_H \geq \frac{1}{8}(\|h-1\|_2^2 - \|h-1\|_3^3)).$$

Finally, the upper bound in (2.4.2) is a special case of (2.4.1). The lower bound follows from the elementary inequality: $\forall\, u \geq -1, \quad |u| \geq \frac{3}{4}u + u^2 - |u|^3$. This ends the proof of Lemma 2.4.1.

2.4.2 The cutoff phenomenon and related questions

This Section describe briefly a surprising property appearing in number of examples of natural finite Markov chains where a careful study is possible. We refer the reader to [4, 17] and the more recent [18] for further details and references.

Consider the following example of finite Markov chain. The state space $\mathcal{X} = \{0,1\}^n$ is the set of all binary vectors of length n. At each step, we pick a coordinate at random and flip it to its opposite. Hence, the kernel K of the chain is $K(x,y) = 0$ unless $|x-y| = 1$ in which case $K(x,y) = 1/n$. This chain is symmetric, irreducible but periodic. It has the uniform distribution $\pi \equiv 2^{-n}$ as stationary measure. Let $H_t = e^{-t}\sum_0^\infty \frac{t^i}{i!}K^i$ be the associated continuous time chain. Then, by the Perron-Frobenius theorem $H_t(x,y) \to 2^{-n}$ as t tends to infinity. This can be quantified very precisely.

Theorem 2.4.2 *For the continuous time chain on the hypercube $\{0,1\}^n$ described above, let $t_n = \frac{1}{4}n\log n$. Then for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \|H^x_{(1-\varepsilon)t_n} - 2^{-n}\|_{\mathrm{TV}} = 1$$

whereas

$$\lim_{n\to\infty} \|H^x_{(1+\varepsilon)t_n} - 2^{-n}\|_{\mathrm{TV}} = 0$$

In fact, a more precise description is feasible in this case. See [20, 18]. This theorem exhibits a typical case of the so called *cutoff phenomenon.* For n large enough, the graph of $t \to y(t) = \|H^x_t - 2^{-n}\|_{\mathrm{TV}}$ stays very close to the line $y = 1$ for a long time, namely for about $t_n = \frac{1}{4}n\log n$. Then, it falls off rapidly to a value close to 0. This fall-off phase is much shorter than t_n. Reference [20] describes the shape of the curve around the critical time t_n.

Definition 2.4.3 *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain.*

1. *One says that $\mathcal{F}$ presents a cutoff in total variation with critical time $(t_n)_1^\infty$ if $t_n \to \infty$ and*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|H^x_{n,(1-\varepsilon)t_n} - \pi_n\|_{\mathrm{TV}} = 1$$
and
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|H^x_{n,(1+\varepsilon)t_n} - \pi_n\|_{\mathrm{TV}} = 0.$$

2. *Let $(t_n, b_n)_1^\infty$ such that $t_n, b_n \geq 0$, $t_n \to \infty$, $b_n/t_n \to 0$. One says that $\mathcal{F}$ presents a cutoff of type $(t_n, b_n)_1^\infty$ in total variation if for all real c*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|H^x_{n,t_n+b_nc} - \pi_n\|_{\mathrm{TV}} = f(c)$$
with $f(c) \to 1$ when $c \to -\infty$ and $f(c) \to 0$ when $c \to \infty$.

Clearly, $2 \Rightarrow 1$. The ultimate cutoff result consists in a precise description of the function f. In Theorem 2.4.2 there is in fact a (t_n, b_m)-cutoff with $t_n = \frac{1}{4}n \log n$ and $b_n = n$. See [20].

In practical terms, the cutoff phenomenon means the following: in order to approximate the stationary distribution π_n one should not stop the chain $H_{n,t}$ before $t = t_n$ and it is essentially useless to run the chain for more than t_n. It seems that the cutoff phenomenon is widespread among natural examples. See [4, 18]. Nevertheless it is rather difficult to verify that a given family of chains satisfy one or the other of the above two definitions. This motivates the following weaker definition.

Definition 2.4.4 *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Fix $1 \leq p \leq \infty$.*

1. *One says that $\mathcal{F}$ presents a weak ℓ^p-cutoff with critical time $(t_n)_1^\infty$ if $t_n \to \infty$ and*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,t_n} - 1\|_{\ell^p(\pi_n)} > 0 \quad \text{and} \quad \lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,(1+\varepsilon)t_n} - 1\|_{\ell^p(\pi_n)} = 0.$$

2. *Let $(t_n, b_n)_1^\infty$ such that $t_n, b_n \geq 0$, $t_n \to \infty$, $b_n/t_n \to 0$. One says that $\mathcal{F}$ presents a weak ℓ^p-cutoff of type $(t_n, b_n)_1^\infty$ if for all $c \geq 0$,*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,t_n+cb_n} - 1\|_{\ell^p(\pi_n)} = f(c)$$
with $f(0) > 0$ and $f(c) \to 0$ when $c \to \infty$.

The notion of weak cutoff extends readily to Hellinger distance or entropy. The advantage of this definition is that it captures some of the spirit of the cutoff phenomenon without requiring a too precise understanding of what happens at relatively small times.

Observe that a cutoff of type $(t_n, b_n)_1^\infty$ is equivalent to a cutoff of type $(t_n, ab_n)_1^\infty$ with $a > 0$ but that t_n can not always be replaced by s_n even if $t_n \sim s_n$.

Note also that if $(t_n)_1^\infty$ and $(s_n)_1^\infty$ are critical times for a family $\mathcal{F}$ (the same for t_n and s_n) then $\lim_{n\to\infty} t_n/s_n = 1$. Indeed, for any $\epsilon > 0$, we must have $(1+\epsilon)t_n > s_n$ and $(1+\epsilon)s_n > t_n$ for n large enough.

Definition 2.4.5 *Let (K,π) be a finite irreducible Markov chain. For $1 \le p \le \infty$ and $\varepsilon > 0$, define the parameter $T_p(K,\varepsilon) = T_p(\varepsilon)$ by*

$$T_p(\varepsilon) = \inf\{t > 0 : \max_x \|h_t^x - 1\|_p \le \varepsilon\}$$

where $H_t = e^{-t(I-K)}$ is the associated continuous time chain.

The next lemma shows that for reversible chains and $1 < p \le \infty$ the different T_p's cannot be too different.

Lemma 2.4.6 *Let (K,π) be a finite irreducible reversible Markov chain. Then, for $2 \le p \le +\infty$ and $\varepsilon > 0$, we have*

$$T_2(K,\varepsilon) \le T_p(K,\varepsilon) \le T_\infty(K,\varepsilon) \le 2\,T_2(K,\varepsilon^{1/2}).$$

Furthermore, for $1 < p \le 2$ and $m_p = 1 + \lceil (2-p)/[2(p-1)] \rceil$,

$$T_p(K,\varepsilon) \le T_2(K,\varepsilon) \le m_p\,T_p(K,\varepsilon^{1/m_p}).$$

PROOF: The first assertion is easy and left as an exercise. For the second we need to use the fact that

$$\max_x \|h_{u+v}^x - 1\|_q \le \left(\max_x \|h_u^x - 1\|_r\right)\left(\max_x \|h_v^x - 1\|_s\right) \tag{2.4.7}$$

for all $u, v > 0$ and $1 \le q, r, s \le +\infty$ related by $1 + 1/q = 1/r + 1/s$. Fix $1 < p < 2$ and an integer j. Set, for $i = 1, \ldots, j-1$, $p_1 = p$, $1 + 1/p_{i+1} = 1/p_i + 1/p$, and $u_i = it/j$, $v_i = t/j$. Applying (2.4.7) $j-1$ times with $q = p_{i+1}$, $r = p_i$, $s = p$, $u = u_i$, $v = v_j$, we get

$$\max_{\mathcal{X}} \|h_t^x - 1\|_{p_j} \le \left(\max_{\mathcal{X}} \|h_{t/j}^x - 1\|_p\right)^j.$$

Now, $p_j = 1/p - (j-1)(1-1/p)$. Thus $p_j \ge 2$ for

$$j \ge 1 + (2-p)/[2(p-1)].$$

The desired result follows.

Theorem 2.4.7 *Fix $1 < p < \infty$ and $\varepsilon > 0$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Let λ_n be the spectral gap of K_n and set $t_n = T_p(K_n, \varepsilon)$. Assume that*

$$\lim_{n\to\infty} \lambda_n t_n = \infty.$$

Then the family $\mathcal{F}$ presents a weak ℓ^p-cutoff of type $(t_n, 1/\lambda_n)_1^\infty$.

PROOF: By definition $\max_{\mathcal{X}_n} \|h^x_{n,t_n} - 1\|_p = \varepsilon > 0$. To obtain an upper bound write

$$\begin{aligned} \|h^x_{n,t_n+s} - 1\|_p &= \|(H^*_{n,s} - \pi_n)(h^x_{n,t_n} - 1)\|_p \\ &\leq \|h^x_{n,t_n} - 1\|_p \, \|H^*_{n,s} - \pi_n\|_{p\to p} \\ &\leq \varepsilon \, \|H^*_{n,s} - \pi_n\|_{p\to p}. \end{aligned}$$

By Theorem 2.1.4

$$\|H^*_{n,s} - \pi_n\|_{2\to 2} \leq e^{-s\lambda_n}.$$

Also, $\|H^*_{n,s} - \pi_n\|_{1\to 1} \leq 2$ and $\|H^*_{n,s} - \pi_n\|_{\infty\to\infty} \leq 2$. Hence, by interpolation, (see Theorem 1.2.8)

$$\|H^*_{n,s} - \pi_n\|_{p\to p} \leq 4^{|1/2-1/p|} e^{-s\lambda_n(1-2|1/2-1/p|)}.$$

It follows that

$$\|h^x_{n,t_n+c/\lambda_n} - 1\|_p \leq \varepsilon 4^{|1/2-1/p|} e^{-c(1-2|1/2-1/p|)}.$$

This proves the desired result since $1 - 2|1/2 - 1/p| > 0$ when $1 < p < \infty$. This also proves the following auxilliary result.

Lemma 2.4.8 *Fix $1 < p < \infty$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let λ_n be the spectral gap of K_n. If*

$$\lim_{n\to\infty} \lambda_n T_p(K_n, \varepsilon) \to \infty$$

for some fixed $\varepsilon > 0$, then

$$\lim_{n\to\infty} \frac{T_p(K_n, \varepsilon)}{T_p(K_n, \eta)} = 1.$$

for all $\eta > 0$.

For reversible chain we obtain a necessary and sufficient condition for weak ℓ^2-cutoff.

Theorem 2.4.9 *Fix $\varepsilon > 0$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of reversible finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Let λ_n be the spectral gap of K_n and set $t_n = T_2(K_n, \varepsilon)$. A necessary and sufficient condition for $\mathcal{F}$ to present a weak ℓ^2-cutoff with critical time t_n is that*

$$\lim_{n\to\infty} \lambda_n t_n = \infty. \tag{2.4.8}$$

Furthermore, if (2.4.8) is satisfied then

1. *$\mathcal{F}$ presents a weak ℓ^∞-cutoff of type $(2t_n, 1/\lambda_n)_1^\infty$.*

2. *For each $1 < p \leq \infty$ and each $\eta > 0$, $\mathcal{F}$ presents a weak ℓ^p-cutoff of type $(T_p(K_n, \eta), 1/\lambda_n)_1^\infty$.*

PROOF: We already now that (2.4.8) is sufficient to have a weak ℓ^2-cutoff. Conversely, if (2.4.8) does not hold there exists $a > 0$ and a subsequence $n(i)$ such that $\lambda_{n(i)} t_{n(i)} \leq a$. To simplify notation assume that this hold for all n. Let ϕ_n be an eigenfunction of K_n such that $\|\phi_n\|_\infty = 1$ and $(I - K_n)\phi_n = \lambda_n \phi_n$. Then

$$\max_{\mathcal{X}_n} \|h^x_{n,t} - 1\|_2 \geq \|(H^x_{n,t} - \pi_n)\phi_n\|_2 = e^{-t\lambda_n}.$$

If follows that, for any $\eta > 0$,

$$\max_{\mathcal{X}_n} \|h^x_{n,(1+\eta)t_n} - 1\|_2 \geq e^{-(1+\eta)t_n\lambda_n} \geq e^{-(1+\eta)a}.$$

Hence

$$\lim_{n \to \infty} \max_{\mathcal{X}_n} \|h^x_{n,(1+\eta)t_n} - 1\|_2 \not\to 0$$

which shows that there is no weak ℓ^2-cutoff.

To prove the assertion concerning the weak ℓ^∞-cutoff simply observe that

$$\max_{\mathcal{X}_n} \|h^x_{n,t} - 1\|_\infty = \max_{\mathcal{X}_n} \|h^x_{n,t/2} - 1\|_2^2.$$

Hence a weak ℓ^2-cutoff of type $(t_n, b_n)_1^\infty$ is equivalent to a weak ℓ^∞-cutoff of type $(2t_n, b_n)$.

For the last assertion use Lemmas 2.4.6 and 2.4.8 to see that (2.4.8) implies $\lambda_n T_p(K, \eta) \to \infty$ for any fixed $\eta > 0$. Then apply Theorem 2.4.7.

The following theorem is based on strong hypotheses that are difficult to check. Nevertheless, it sheds some new light on the cutoff phenomenon.

Theorem 2.4.10 *Fix $\varepsilon > 0$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of reversible finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Let λ_n be the spectral gap of K_n and set $t_n = T_2(K_n, \varepsilon)$. Let α_n be the log-Sobolev constant of (K_n, π_n). Set*

$$A_n = \max\{\|\phi\|_\infty : \|\phi\|_2 = 1, K_n\phi = (1 - \lambda_n)\phi\}.$$

Assume that the following conditions are satisfied.

(1) $t_n \lambda_n \to \infty$.

(2) $\inf_n \{\alpha_n / \lambda_n\} = c_1 > 0$.

(3) $\inf_n \{A_n e^{-\lambda_n t_n}\} = c_2 > 0$.

Then the family $\mathcal{F}$ presents a weak ℓ^p-cutoff with critical time $(t_n)_1^\infty$ for any $1 \leq p < \infty$ and also in Hellinger distance.

PROOF: By Theorem 2.4.9 condition (1) implies a weak ℓ^p-cutoff of type

$$(T_p(K_n, \eta), \lambda_n)$$

for each $1 < p < \infty$ and $\eta > 0$. The novelty in Theorem 2.4.10 is that it covers the case $p = 1$ (and Hellinger distance) and that the critical time $(t_n)_0^\infty$ **does not** depend on $1 \le p < \infty$. For the case $p > 2$, it suffices to prove that $T_p(K_n, \varepsilon) \le t_n + c(p)/\lambda_n$. Using symmetry, (2.2.2) and hypothesis (2), we get

$$\|h^x_{n,t_n+s_n} - 1\|_p \le \|H_{n,s_n}\|_{2\to p}\|h^x_{n,t_n} - 1\|_2 \le \varepsilon$$

with $s_n = [\log(p-1)]/(4\alpha_n) \le [\log(p-1)]/(4c_1\lambda_n)$, which yields the desired inequality. Observe that condition (3) has not been used to treat the case $2 < p < \infty$.

We now turn to the proof of the weak ℓ^1-cutoff. Since

$$\|h_{n,t} - 1\|_1 \le \|h_{n,t} - 1\|_2$$

it suffices to prove that

$$\liminf_{n\to\infty} \|h_{n,t_n} - 1\|_1 > 0.$$

To prove this, we use the lower bound in (2.4.2) and condition (3) above. Indeed, for each n there exists a normalized eigenfunction ϕ_n and $x_n \in \mathcal{X}_n$ such that $K_n\phi_n = (1-\lambda_n)\phi_n$ and $\|\phi_n\|_\infty = \phi_n(x_n) = A_n$. It follows that

$$\begin{aligned}
\|h^{x_n}_{n,t_n+s} - 1\|_2 &= \sup_{\|\psi\|_2\le 1} \{\|(H_{n,t_n+s} - \pi_n)\psi\|_\infty\} \\
&\ge A_n\, e^{-\lambda_n(t_n+s)} \ge c_2\, e^{-\lambda_n s}.
\end{aligned}$$

Also, for $\sigma_n = (\log 2)/(4\alpha_n)$, we have

$$\begin{aligned}
\|h^x_{n,t_n+\sigma_n+s} - 1\|_3 &\le \|h^x_{n,t_n+s} - 1\psi\|_2 \\
&\le \|h^x_{n,t_n} - 1\psi\|_2\|H_{n,s} - \pi_n\|_{2\to 2} \\
&\le \varepsilon\, e^{-\lambda_n s}.
\end{aligned}$$

Hence, since $\lambda_n\sigma_n \le [\log 2]/4c_1$,

$$\begin{aligned}
\|h^{x_n}_{n,t_n+\sigma_n+s} - 1\|_1 &\ge \|h^{x_n}_{n,t_n+\sigma_n+s} - 1\|_2^2 - \|h^{x_n}_{n,t_n+\sigma_n+s} - 1\|_3^3 \\
&\ge c_2^2\, e^{-2\lambda_n(\sigma_n+s)} - \varepsilon^3\, e^{-3\lambda_n s} \\
&\ge \left(c_2^2\, e^{-2\lambda_n\sigma_n} - \varepsilon^3\, e^{-\lambda_n s}\right) e^{-2\lambda_n s} \\
&\ge \left(c_3 - \varepsilon^3\, e^{-\lambda_n s}\right) e^{-2\lambda_n s}
\end{aligned}$$

where $c_3 = c_2^2 2^{-1/4c_1}$. For each fixed n, we now pick $s = s_n = \lambda_n^{-1}\log(c_3/(2\varepsilon^3))$. Hence

$$\|h^{x_n}_{n,t_n} - 1\|_1 \ge \|h^{x_n}_{n,t_n+\sigma_n+s_n} - 1\|_1 \ge c_3/2.$$

The weak cutoff in Hellinger distance is proved the same way using (2.4.3) or (2.4.4). Finally the case $1 < p < 2$ follows from the results obtained for $p = 2$ and $p = 1$.

Chapter 3

Geometric tools

This chapter uses adapted graph structures to study finite Markov chains. It shows how paths on graphs and their combinatorics can be used to prove Poincaré and Nash inequalities. Isoperimetric techniques are also considered. Path techniques have been introduced by M. Jerrum and A. Sinclair in their study of a stochastic algorithm that counts perfect matchings in a graph. See [72]. Paths are also used in [79] in a somewhat different context (random walk on finitely generated groups). They are used in [35] to prove Poincaré inequalities. The underlying idea is classical in analysis and geometry. The simplest instance of it is the following proof of a Poincaré inequality for the unit interval $[0,1]$:

$$\int_0^1 |f(s)-m|^2 ds \leq \frac{1}{8}\int_0^1 |f'(s)|^2 ds$$

where m is the mean of f. Write $f(s)-f(t)=\int_t^s f'(u)du$ for any $0 \leq t < s \leq 1$. Hence, using the Cauchy-Schwarz inequality, $|f(s)-f(t)|^2 \leq (s-t)\int_t^s |f'(u)|^2 du$. It follows that

$$\begin{aligned}
\int_0^1 |f(s)-m|^2 ds &\leq \int_0^1\int_0^1 |f(s)-f(t)|^2 dt ds \\
&\leq \int_0^1 |f'(u)|^2 \left\{\int_0^1\int_0^1 (s-t)\mathbf{1}_{t\leq u\leq s}(u) dt ds\right\} du \\
&= \int_0^1 |f'(u)|^2 \left\{\frac{u(1-u)}{2}\right\} du \\
&\leq \frac{1}{8}\int_0^1 |f'(u)|^2 du.
\end{aligned}$$

The constant $1/8$ obtained by this argument must be compared with the best possible constant which is $1/\pi^2$.

This chapter develops and illustrates several versions of this technique in the context of finite graphs.

3.1 Adapted edge sets

Definition 3.1.1 *Let K be an irreducible Markov chain on a finite set $\mathcal{X}$. An edge set $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ is say to be adapted to K if $\mathcal{A}$ is symmetric (that is $(x,y) \in \mathcal{A} \Rightarrow (y,x) \in \mathcal{A}$), $(\mathcal{X}, \mathcal{A})$ is connected, and*

$$(x,y) \in \mathcal{A} \Rightarrow K(x,y) + K(y,x) > 0.$$

In this case we also say that the graph $(\mathcal{X}, \mathcal{A})$ is adapted.

Let K be an irreducible Markov kernel on $\mathcal{X}$ with stationary measure π. It is convenient to introduce the following notation. For any $e = (x,y) \in \mathcal{X} \times \mathcal{X}$, set

$$df(e) = f(y) - f(x)$$

and define

$$Q(e) = \frac{1}{2}\left(K(x,y)\pi(x) + K(y,x)\pi(y)\right).$$

We will sometimes view Q as a probability measure on $\mathcal{X} \times \mathcal{X}$. Observe that, by Definition 2.1.1 and (2.1.1), the Dirichlet form $\mathcal{E}$ of (K, π) satisfies

$$\mathcal{E}(f,f) = \frac{1}{2} \sum_{e \in \mathcal{X} \times \mathcal{X}} |df(e)|^2 Q(e).$$

Let $\mathcal{A}$ be an adapted edge set. A path γ in $(\mathcal{X}, \mathcal{A})$ is a sequence of vertices $\gamma = (x_0, \ldots, x_k)$ such that $(x_{i-1}, x_i) \in \mathcal{A}$, $i = 1, \ldots, k$. Equivalently, γ can be viewed as a sequence of edges $\gamma = (e_1, \ldots, e_k)$ with $e_i = (x_{i-1}, x_i) \in \mathcal{A}$, $i = 1, \ldots, k$. The length of such a path γ is $|\gamma| = k$. Let Γ be the set of all paths γ in $(\mathcal{X}, \mathcal{A})$ which have no repeated edges (that is, such that $e_i \neq e_j$ if $i \neq j$). For each pair $(x,y) \in \mathcal{X} \times \mathcal{X}$, set

$$\Gamma(x,y) = \{\gamma = (x_0, \ldots, x_k) \in \Gamma : x = x_0,\ y = x_k\}.$$

3.2 Poincaré inequality

A Poincaré inequality is an inequality of the type

$$\forall f, \quad \mathrm{Var}_\pi(f) \leq C\mathcal{E}(f,f).$$

It follows from the definition 2.1.3 of the spectral gap λ that such an inequality is equivalent to $\lambda \geq 1/C$. In other words, the smallest constant C for which the Poincaré inequality above holds is $1/\lambda$. This section uses Poincaré inequality and path combinatorics to bound λ from below. We start with the simplest result of this type.

Theorem 3.2.1 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set. For each $(x,y) \in \mathcal{X} \times \mathcal{X}$ choose exactly one path $\gamma(x,y)$ in $\Gamma(x,y)$. Then $\lambda \geq 1/A$ where*

$$A = \max_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y\in\mathcal{X}:\\ \gamma(x,y)\ni e}} |\gamma(x,y)|\pi(x)\pi(y) \right\}.$$

PROOF: For each $(x,y) \in \mathcal{X} \times \mathcal{X}$, write

$$f(y) - f(x) = \sum_{e\in\gamma(x,y)} df(e)$$

and, using Cauchy-Schwarz,

$$|f(y) - f(x)|^2 \leq |\gamma(x,y)| \sum_{e\in\gamma(x,y)} |df(e)|^2.$$

Multiply by $\frac{1}{2}\pi(x)\pi(y)$ and sum over all x, y to obtain

$$\frac{1}{2}\sum_{x,y} |f(y) - f(x)|^2\pi(x)\pi(y) \leq \frac{1}{2}\sum_{x,y} |\gamma(x,y)| \sum_{e\in\gamma(x,y)} |df(e)|^2\pi(x)\pi(y).$$

The left-hand side is equal to $\mathrm{Var}_\pi(f)$ whereas the right-hand side becomes

$$\frac{1}{2}\sum_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y:\\ \gamma(x,y)\ni e}} |\gamma(x,y)|\pi(x)\pi(y) \right\} |df(e)|^2 Q(e)$$

which is bounded by

$$\max_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y:\\ \gamma(x,y)\ni e}} |\gamma(x,y)|\pi(x)\pi(y) \right\} \mathcal{E}(f,f).$$

This proves the Poincaré inequality

$$\forall f, \quad \mathrm{Var}_\pi(f) \leq A\mathcal{E}(f,f)$$

hence $\lambda \geq 1/A$.

EXAMPLE 3.2.1: Let $\mathcal{X} = \{0,1\}^n$, $\pi \equiv 2^{-n}$ and $K(x,y) = 0$ unless $|x-y| = 1$ in which case $K(x,y) = 1/n$. Consider the obvious adapted edge set $\mathcal{A} = \{(x,y) : |x-y| = 1\}$. To define a path $\gamma(x,y)$ from x to y, view x, y as binary vectors and change the coordinates of x one at a time from left to right to match the coordinates of y. These paths have length at most n. Since $1/Q(e) = n\,2^n$ we obtain in this case

$$\begin{aligned} A &\leq n^2\,2^{-n} \max_{e\in\mathcal{A}} \left\{ \sum_{\substack{x,y:\\ \gamma(x,y)\ni e}} 1 \right\} \\ &= n^2\,2^{-n} \max_{e\in\mathcal{A}} \#\{(x,y) : \gamma(x,y) \ni e\}. \end{aligned}$$

Hence every thing boils down to count, for each edge $e \in \mathcal{A}$, how many paths $\gamma(x,y)$ use that edge. Let $e=(u,v)$. Since $e \in \mathcal{A}$, there exists a unique i such that $u_i \neq v_i$. Furthermore, by construction, if $\gamma(x,y) \ni e$ we must have

$$\begin{aligned} x &= (x_1,\ldots,x_{i-1},u_i,u_{i+1},\ldots,u_n)\\ y &= (v_1,\ldots,v_{i-1},v_i,\,y_{i+1},\ldots,y_n). \end{aligned}$$

It follows that $i-1$ coordinates of x and $n-i$ coordinates of y are unknown. That is, $\#\{(x,y) : \gamma(x,y) \ni e\} = 2^{n-1}$. Hence $A \le n^2/2$ and Theorem 3.2.1 yields $\lambda \ge 2/n^2$. The right answer is $\lambda = 2/n$. The above computation is quite typical of what has to be done to use Theorem 3.2.1. Observe in particular the non trivial cancellation of the exponential factors.

EXAMPLE 3.2.2: Keep $\mathcal{X} = \{0,1\}^n$ and consider the following moves: $x \to \tau(x)$ where $\tau(x)_i = x_{i-1}$ and $x \to \sigma(x)$ where $\sigma(x) = x + (1,0,\ldots,0)$. Let $K(x,y) = 1/2$ if $y=\tau(x)$ or $y = \sigma(x)$ and $K(x,y)=0$ otherwise. This chain has $\pi \equiv 2^{-n}$ as stationary distribution. It is not reversible. Define $\gamma(x,y)$ as follows. Use τ to turn the coordinates around from right to left. Use σ to ajust x_i to y_i if necessary as it passes in position 1. These paths have length at most $2n$. Let $e=(u,v)$ be an edge, say $v=\sigma(u)$. Pick an integer j, $0 \le j \le n-1$. Then, if we assume that τ as been used exactly j times before e, then $x_i = u_{i-j}$ for $j < i \le n$, $y_i = v_{n-j+i}$ for $1 \le i \le j$ and $y_{j+1} = v_1$. Hence, there are 2^{n-1} ordered pair (x,y) such that $e \in \gamma(x,y)$ appears after exactly j uses of τ. Since there are n possible values of j, this shows that the constant A of Theorem 3.2.1 is bounded by $A \le 4n^2$ and thus $\lambda \ge 1/(4n^2)$.

EXAMPLE 3.2.3: Let again $\mathcal{X} = \{0,1\}^n$. Let τ,σ be as in the preceding example. Consider the chain with kernel $K(x,y) = 1/n$ if either $y = \tau^j(x)$ for some $0 \le j \le n-1$ or $y=\sigma(x)$, and $K(x,y)=0$ otherwise. This chain is reversible with respect to the uniform distribution. Without further idea, it seems difficult to do any thing much better than using the same paths and the same analysis as in the previous example. This yields $A \le n^3$ and $\lambda \ge 1/n^3$. Clearly, a better analysis is desirable in this case because we have not taken advantage of all the moves at our disposal. A better bound will be obtained in Section 4.2.

EXAMPLE 3.2.4: It is instructive to work out what Theorem 3.2.1 says for simple random walk on a graph $(\mathcal{X},\mathcal{A})$ where $\mathcal{A}$ is a symmetric set of oriented edges. Set $d(x) = \#\{y \in \mathcal{X} : (x,y) \in \mathcal{A}\}$ and recall that the simple random walk on $(\mathcal{X},\mathcal{A})$ has kernel

$$K(x,y) = \begin{cases} 0 & \text{if } (x,y) \notin \mathcal{A} \\ 1/d(x) & \text{if } (x,y) \in \mathcal{A}. \end{cases}$$

This gives a reversible chain with respect to the measure $\pi(x) = d(x)/|\mathcal{A}|$. For each $(x,y) \in \mathcal{X}^2$ choose a path $\gamma(x,y)$ with no repeated edge. Set

$$d_* = \max_{x\in\mathcal{X}} d(x),\quad \gamma_* = \max_{x,y\in\mathcal{X}} |\gamma(x,y)|,\quad \eta_* = \max_{e\in\mathcal{A}} \#\{(x,y) \in \mathcal{X}^2 : \gamma(x,y) \ni e\}.$$

Then Theorem 3.2.1 gives $\lambda \geq 1/A$ with

$$A \leq \frac{d_*^2 \gamma_* \eta_*}{|\mathcal{A}|}.$$

The quantity η_* can be interpreted as a measure of bottle necks in the graph $(\mathcal{X}, \mathcal{A})$. The quantity γ_* as an obvious interpretation as an upper bound on the diameter of the graph.

We now turn to more sophisticated (but still useful) versions of Theorem 3.2.1.

Definition 3.2.2 *A weight function w is a positive function*

$$w : \mathcal{A} \to (0, \infty).$$

The w-length of a path γ in Γ is

$$|\gamma|_w = \sum_{e \in \gamma} \frac{1}{w(e)}.$$

Theorem 3.2.3 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set and w be a weight function. For each $(x, y) \in \mathcal{X} \times \mathcal{X}$ choose exactly one path $\gamma(x, y)$ in $\Gamma(x, y)$. Then $\lambda \geq 1/A(w)$ where*

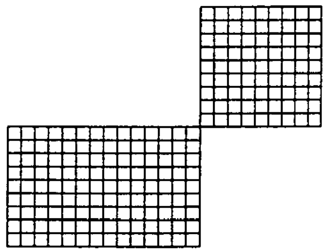
$$A(w) = \max_{e \in \mathcal{A}} \left\{ \frac{w(e)}{Q(e)} \sum_{\substack{(x,y): \\ \gamma(x,y) \ni e}} |\gamma(x, y)|_w \pi(x) \pi(y) \right\}.$$

PROOF: Start as in the proof of Theorem 3.2.1 but introduce the weight w when using Cauchy-Schwarz to get

$$\begin{aligned} |f(y) - f(x)|^2 &\leq \left(\sum_{e \in \gamma(x,y)} w(e)^{-1} \right) \left(\sum_{e \in \gamma(x,y)} |df(e)|^2 w(e) \right) \\ &= |\gamma(x, y)|_w \sum_{e \in \gamma(x,y)} |df(e)|^2 w(e). \end{aligned}$$

¿From here, complete the proof by following step by step the proof of Theorem 3.2.1. A subtle discussion of this result can be found in [55] which also contains interesting examples.

EXAMPLE 3.2.5: What is the spectral gap of the dog? (for simplicity, the dog below has no ears or legs or tail).

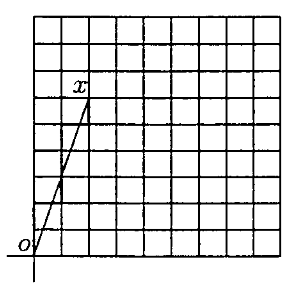


For a while, Diaconis and I puzzled over finding the order of magnitude of the spectral gap for simple random walk on the planar graph made from two square grids, say of side length n, attached together by one of their corners. This example became known to us as "the dog". It turns out that the dog is quite an interesting example. Thus, let $\mathcal{X}$ be the vertex set of two $n \times n$ square grids $\{0, \ldots, n\}^2$ and $\{-n, \ldots, 0\}^2$ attached by identifying the two corners $o = (0,0) \in \mathcal{X}$ so that $|\mathcal{X}| = 2(n+1)^2 - 1$. Consider the markov kernel

$$K(x,y) = \begin{cases} 0 & \text{if } |x-y| > 1 \\ 1/4 & \text{if } |x-y| = 1 \\ 0 & \text{if } x = y \text{ is inside or } x = y = 0 \\ 1/4 & \text{if } x = y \text{ is on the boundary but not a corner} \\ 1/2 & \text{if } x = y \text{ is a corner.} \end{cases}$$

This is a symmetric kernel with uniform stationary measure $\pi \equiv (2(n+1)^2 - 1)^{-1}$ and $1/Q(e) = 4(2(n+1)^2 - 1)$ if $e \in \mathcal{A}$. We will refer to this example as the n-dog.

We now have to choose paths. The graph structure on $\mathcal{X}$ induces a distance $d(x,y)$ between vertices. Also, we have the Euclidean distance $|x-y|$. First we define paths from any $x \in \mathcal{X}$ to o. For definitness, we work in the square lying in the first quadrant. Let $\gamma(x,o)$ be one of the geodesic paths from x to o such that, for any $z \in \gamma(x,o)$, the Euclidean distance between z and the straight line segment $[x,o]$ is at most $1/\sqrt{2}$.



Let $e = (u,v)$ be an edge with $d(o,v) = i$, $d(o,u) = i+1$. We claim that

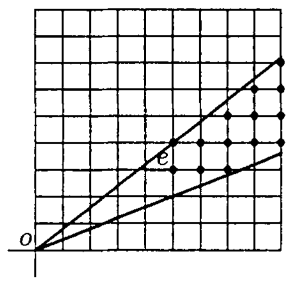
$$\#\{x : \gamma(x,o) \ni e\} \leq \frac{4(n+1)^2}{i+1}.$$

By symmetry, we can assume that $u = (u_1, u_2)$ with $u_1 \geq u_2$. This implies that $u_1 \geq (i+1)/2$. Let I be the vertical segment of length 2 centred at u. Set

$$\{x : \gamma(x, o) \ni e\} = Z(e).$$

If $z \in Z(e)$ then the straight line segment $[o, z]$ is at Euclidean distance at most $1/\sqrt{2}$ from u. This implies that $Z(e)$ is contained in the half cone $\mathcal{C}(u)$ with vertex o and base I (because $(u_1 \geq u_2)$. Thus

$$Z(e) \subset \{(z_1, z_2) \in \{0, \ldots, n\}^2 : z_1 \geq u_1, z_2 \geq u_2\} \cap \mathcal{C}(u).$$



Let $\ell(j)$ be the length of the intersection of the vertical line $U(j)$ passing through $(j, 0)$ with $\mathcal{C}$. Then $\ell(j)/j = \ell(k)/k$ for all j, k. Clearly $\ell(u_1) = 3$. Hence $\ell(j) \leq 3j/u_1$. This means that there are at most $1 + 3j/u_1$ vertices in $U_j \cap Z(e)$. Summing over all $u_1 \leq j \leq n$ we obtain

$$\#Z(e) \leq n + \frac{3n(n+1)}{2u_1} \leq \frac{4n(n+1)}{i+1}.$$

which is the claimed inequality.

Now, if x, y are any two vertices in $\mathcal{X}$, we join them by going through o using the paths $\gamma(x, o), \gamma(y, o)$ in the obvious way. This defines $\gamma(x, y)$. Furthermore, we consider the weight function w on edges defined by $w(e) = i + 1$ if e is at graph distance i from o. Observe that the length of any of the paths $\gamma(x, y)$ is at most

$$2 \sum_{0}^{2n-1} \frac{1}{i+1} \leq 2 \log(2n+1).$$

Also, the number of times a given edge e at distance i from o is used can be bounded as follows.

$$\begin{aligned} \#\{(x, y) : \gamma(x, y) \ni e\} &\leq (2(n+1)^2 - 1) \times \#\{z : \gamma(z, o) \ni e\} \\ &\leq 4(n+1)^2 (2(n+1)^2 - 1)/(i+1). \end{aligned}$$

Hence, The constant A in Theorem 3.2.3 satisfies

$$\begin{aligned} A &\leq \frac{4 \max_{x,y} |\gamma(x, y|_w}{2(n+1)^2 - 1} \max_e \{w(e) \#\{(x, y) : \gamma(x, y) \ni e\}\} \\ &\leq 16(n+1)^2 \log(2n+1). \end{aligned}$$

This yields $\lambda \geq (16(n+1)^2 \log(2n+1))^{-1}$. To see that this is the right order of magnitude, use the test function f defined by $f(x) = \text{sgn}(x) \log(1 + d(0,x))$ where $\text{sgn}(x)$ is $1, 0$ or -1 depending on whether the sum of the coordinates of x is positive 0 or negative. This function has $\pi(f) = 0$,

$$\text{Var}_\pi(f) = \|f\|_2^2 \geq \frac{n(n+1)}{2(n+1)^2 - 1}[\log(n+1)]^2$$

and

$$\begin{aligned}
\mathcal{E}(f,f) &\leq \frac{1}{2[2(n+1)^2-1]} \sum_{i=0}^{2n-1} [(i+1) \wedge (2n-i+1)] |\log(i+2) - \log(i+1)|^2 \\
&\leq \frac{1}{2(n+1)^2 - 1} \sum_{i=0}^{n-1} \frac{1}{i+1} \\
&\leq \frac{\log(n+1)}{2(n+1)^2 - 1}.
\end{aligned}$$

Hence, $\lambda \leq [n(n+1)\log(n+1)]^{-1}$. Collecting the results we see that the spectral gap of the n-dog satifies

$$\frac{1}{16(n+1)^2 \log(2n+1)} \leq \lambda \leq \frac{1}{n(n+1)\log(n+1)}.$$

One can convince oneself that there is no choice of paths such that Theorem 3.2.1 give the right order of magnitude. In fact the best that Theorem 3.2.1 gives in this case is $\lambda \geq c/n^3$. The above problem (and its solution) generalizes to any fixed dimension d. For any $d \geq 3$, the corresponding spectral gap satisfies $c_1(d)/n^d \leq \lambda \leq c_2(d)/n^d$.

In Theorems 3.2.1, 3.2.3, exactly one path $\gamma(x,y)$ is used for each pair (x,y). In certain situations it is helpful to allow the use of more than one path from x to y. To this end we introduce the notion of flow.

Definition 3.2.4 *Let (K,π) be an irreducible Markov chain on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set. A flow is non-negative function on the path set Γ,*

$$\phi : \Gamma \to [0, \infty[$$

such that

$$\forall x, y \in \mathcal{X},\ x \neq y, \quad \sum_{\gamma \in \Gamma(x,y)} \phi(\gamma) = \pi(x)\pi(y).$$

Theorem 3.2.5 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set and ϕ be a flow. Then $\lambda \geq 1/A(\phi)$ where*

$$A(\phi) = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{\gamma \in \Gamma: \\ \gamma \ni e}} |\gamma| \phi(\gamma) \right\}.$$

PROOF: This time, for each (x,y) and each $\gamma \in \Gamma(x,y)$ write

$$|f(y)-f(x)|^2 \leq |\gamma| \sum_{e\in\gamma} |df(e)|^2.$$

Then

$$|f(y)-f(x)|^2\pi(x)\pi(y) \leq \sum_{\gamma\in\Gamma(x,y)} |\gamma| \sum_{e\in\gamma} |df(e)|^2\phi(\gamma).$$

Complete the proof as for Theorem 3.2.1.

EXAMPLE 3.2.6: Consider the hypercube $\{0,1\}^n$ with the chain $K(x,y)=0$ unless $|x-y|=1$ in which case $K(x,y)=1/n$. Consider the set $\mathcal{G}(x,y)$ of all geodesic paths from x to y. Define a flow ϕ by setting

$$\phi(\gamma)=\begin{cases} [2^{2n}\#\mathcal{G}(x,y)]^{-1} & \text{if } \gamma\in\mathcal{G}(x,y) \\ 0 & \text{otherwise.}\end{cases}$$

Then $A(\phi)=\max_e A(\phi,e)$ where

$$A(\phi,e)=n2^n \sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}} |\gamma|\phi(\gamma).$$

Using the symmetries of the hypercube, we observe that $A(\phi,e)$ does not depend on e. Summing over the $n2^n$ oriented edges yields

$$\begin{aligned} A(\phi,e) &= \sum_{e\in\mathcal{A}} \sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}} |\gamma|\phi(\gamma) \\ &= \sum_{\gamma} |\gamma|^2\phi(\gamma) \leq n^2. \end{aligned}$$

This example generalizes as follows.

Corollary 3.2.6 *Assume that there is a group G which acts on $\mathcal{X}$ and such that*

$$\pi(gx)=\pi(x),\quad Q(gx,gy)=Q(x,y).$$

Let $\mathcal{A}$ be an adapted edge set such that $(x,y)\in\mathcal{A} \Rightarrow (gx,gy)\in\mathcal{A}$. Let $\mathcal{A}=\bigcup_1^k \mathcal{A}_i$, be the partition of $\mathcal{A}$ into transitive classes for this action. Then $\lambda\geq 1/A$ where

$$A=\max_{1\leq i\leq k}\left\{\frac{1}{|\mathcal{A}_i|Q_i}\sum_{x,y} d(x,y)^2\pi(x)\pi(y)\right\}.$$

Here $|\mathcal{A}_i|=\#\mathcal{A}_i$, $Q_i=Q(e_i)$ with $e_i\in\mathcal{A}_i$, and $d(x,y)$ is the graph distance between x and y.

Proof: Consider the set $\mathcal{G}(x,y)$ of all geodesic paths from x to y. Define a flow ϕ by setting

$$\phi(\gamma) = \begin{cases} \pi(x)\pi(y)/\#\mathcal{G}(x,y) & \text{if } \gamma \in \mathcal{G}(x,y) \\ 0 & \text{otherwise.} \end{cases}$$

Then $A(\phi) = \max_e A(\phi, e)$ where

$$A(\phi, e) = \frac{1}{Q(e)} \sum_{\substack{\gamma\in\Gamma: \\ \gamma\ni e}} |\gamma|\phi(\gamma).$$

By hypothesis, $A(\phi, e_i) = A_i(\phi)$ does not depend on $e_i \in \mathcal{A}_i$. Indeed, if $g\gamma$ denote the image of the path γ under the action of $g \in G$, we have $|g\gamma| = |\gamma|$, $\phi(g\gamma) = \phi(\gamma)$. Summing for each $i = 1, \dots, k$ over all the oriented edges in $\mathcal{A}_i$, we obtain

$$\begin{aligned} A(\phi, e_i) &= \frac{1}{|\mathcal{A}_i|Q_i} \sum_{e\in\mathcal{A}_i} \sum_{\substack{\gamma\in\Gamma: \\ \gamma\ni e}} |\gamma|\phi(\gamma) \\ &= \frac{1}{|\mathcal{A}_i|Q_i} \sum_{e\in\mathcal{A}_i} \sum_{x,y} \sum_{\substack{\gamma\in\mathcal{G}(x,y): \\ \gamma\ni e}} \frac{d(x,y)\pi(x)\pi(y)}{\#\mathcal{G}(x,y)} \\ &\le \frac{1}{|\mathcal{A}_i|Q_i} \sum_{x,y} N_i(x,y)d(x,y)\pi(x)\pi(y) \end{aligned}$$

where

$$N_i(x,y) = \max_{\gamma\in\mathcal{G}(x,y)} \#\{e \in \mathcal{A}_i : \gamma \ni e\}.$$

That is, $N_i(x,y)$ is the maximal number of edges of type i used in a geodesic path from x to y. In particular, $N_i(x,y) \le d(x,y)$ and the announced result follows.

Example 3.2.7: Let $\mathcal{X}$ be the set of all k-subsets of a set with n elements. Assume $k \le n/2$. Consider the graph with vertex set $\mathcal{X}$ and an edge from x to y if $\#(x \cap y) = k - 2$. This is a regular graph with degree $k(n-k)$. The simple random walk on this graph has kernel

$$K(x,y) = \begin{cases} 1/[k(n-k)] & \text{if } \#(x\cap y) = k-2 \\ 0 & \text{otherwise} \end{cases}$$

and stationary measure $\pi \equiv \binom{n}{k}^{-1}$. It is clear that the symmetric group S_n acts transitively on the edge set of this graph and preserves K and π. Here there is only one class of edges, $|\mathcal{A}| = \binom{n}{k} n(n-k)$, $Q = |\mathcal{A}|^{-1}$. Therefore Corollary 3.2.6 yields $\lambda \ge 1/A$ with

$$A = \frac{1}{|\mathcal{A}|Q} \sum_{x,y} d(x,y)^2 \pi(x)\pi(y)$$

$$\begin{aligned}
&= \frac{1}{\binom{n}{k}^2}\sum_1^k \ell^2 \binom{n}{k}\binom{k}{\ell}\binom{n-k}{\ell} \\
&= \frac{1}{\binom{n}{k}}\sum_1^k \ell^2 \binom{k}{\ell}\binom{n-k}{\ell} \\
&= \frac{k(n-k)}{\binom{n}{k}}\sum_1^k \binom{k-1}{\ell-1}\binom{n-k-1}{\ell-1} \\
&= \frac{k(n-k)}{\binom{n}{k}}\binom{n-1}{k-1} = \frac{k(n-k)^2}{n}.
\end{aligned}$$

Hence

$$\lambda \geq \frac{n}{k(n-k)^2}.$$

Here we have used the fact that the number of pair (x,y) with $d(x,y) = \ell$ is $\binom{n}{k}\binom{k}{\ell}\binom{n-k}{\ell}$ to obtain the second inequality. The true value is $n/[k(n-k)]$. See [34].

EXAMPLE 3.2.8: Let $\mathcal{X}$ be the set of all n-subsets of $\{0,\dots,2n-1\}$. Consider the graph with vertex set $\mathcal{X}$ and an edge from x to y if $\#(x \cap y) = n-2$ and $0 \in x \oplus y$ where $x \oplus y = x \cup y \setminus x \cap y$ is the symmetric difference of x and y. This is a regular graph with degree n. The simple random walk on this graph has kernel

$$K(x,y) = \begin{cases} 1/n & \text{if } \#(x \cap y) = n-2 \text{ and } 0 \in x \oplus y \\ 0 & \text{otherwise} \end{cases}$$

and stationary measure $\pi \equiv \binom{2n}{n}^{-1}$. This process can be described informally as follows: Let x be subset of $\{0,\dots,2n-1\}$ having n elements. If $0 \in x$, pick an element a uniformly at random in the complement of x and move to $y = (x \setminus \{0\}) \cup \{a\}$, that is, replace 0 by a. If $0 \notin x$, pick an element a uniformly at random in x and move to $y = (x \setminus \{a\}) \cup \{0\}$, that is, replace a by 0.

It is clear that the symmetric group S_{2n-1} which fixes 0 and acts on $\{1,\dots,2n-1\}$ also acts on this graph and preserves K and π. This action is not transitive on edges. There are two transitive classes $\mathcal{A}_1, \mathcal{A}_2$ of edges depending on whether, for an edge (x,y), $0 \in x$ or $0 \in y$. Clearly

$$|\mathcal{A}_1| = |\mathcal{A}_2| = \binom{2n}{n} n, \quad Q_1 = Q_2 = |\mathcal{A}|^{-1} = (2|\mathcal{A}_1|)^{-1}.$$

If x and y differ by exactly ℓ elements, the distance between x and y is 2ℓ if $0 \notin x \oplus y$ and $2\ell - 1$ if $0 \in x \oplus y$. Using this and a computation similar to the one in Example 3.2.7, we see that the constant A in Corollary 3.2.6 is bounded by

$$A = \frac{1}{|\mathcal{A}_1| Q_1} \sum_{x,y} d(x,y)^2 \pi(x)\pi(y)$$

$$
\begin{aligned}
&\le \frac{8}{\binom{2n}{n}^2}\sum_{1}^{n}\ell^2\binom{2n}{n}\binom{n}{\ell}^2 \\
&= 8n^2.
\end{aligned}
$$

Hence $\lambda \geq 1/(8n^2)$. This can be slightly improved if we use the $N_i(x,y)$'s introduced in the proof of Corollary 3.2.6. Indeed, this proof shows that $\lambda \geq 1/A'$ with

$$A' = \max_i \left\{ \frac{1}{|\mathcal{A}_i|Q_i} \sum_{x,y} N_i(x,y)d(x,y)\pi(x)\pi(y) \right\}$$

where $N_i(x,y)$ is the maximal number of edges of type i used in any geodesic path from x to y. In the present case, if $x \oplus y = \ell$, then the distance between x and y is atmost 2ℓ with atmost ℓ edges of each of the two types. Hence, $A' \leq 4n^2$ and $\lambda \geq 1/(4n^2)$. The true order of magnitude is $1/n$. See the end of Section 4.2.

Corollary 3.2.7 *Assume that $\mathcal{X} = G$ is a finite group with generating set $S = \{g_1, \ldots, g_s\}$. Set $K(x,y) = |S|^{-1}\mathbf{1}_S(x^{-1}y)$, $\pi \equiv 1/|G|$. Then*

$$\lambda(K) \geq \frac{1}{2|S|D^2}$$

where D is the diameter of the Cayley graph $(G, S \cup S^{-1})$. If S is symmetric, i.e., $S = S^{-1}$, then

$$\lambda(K) \geq \frac{1}{|S|D^2}.$$

PROOF: The action of the group G on its itself by left translation preserves K and π. Hence it also preserves Q. We set

$$\mathcal{A} = \left\{(x, xs) : x \in G, s \in S \cup S^{-1}\right\}.$$

There are at most $s = 2|S|$ classes of oriented edges (corresponding to the distinct elements of $S \cup S^{-1}$) and each class contains at least $|G|$ distinct edges. If S is symmetric (that is $g \in S \Rightarrow g^{-1} \in S$) then $1/Q(e) = |S||G|$ whereas if S is not symmetric, $|S||G| \leq 1/Q(e) \leq 2|S||G|$. The results now follow from Corollary 3.2.6. Slightly better bounds are derived in [24].

Corollary 3.2.8 *Assume that $\mathcal{X} = G$ is a finite group with generating set $S = \{g_1, \ldots, g_s\}$. Set $K(x,y) = |S|^{-1}\mathbf{1}_S(x^{-1}y)$, $\pi \equiv 1/|G|$. Assume that there is a subgroup H of the group of automorphisms of G which preserves S and acts transitively on S. Then*

$$\lambda(K) \geq \frac{1}{2D^2}$$

where D is the diameter of the Cayley graph $(G, S \cup S^{-1})$. If S is symmetric, i.e., $S = S^{-1}$, or if H acts transitively on $S \cup S^{-1}$, then

$$\lambda(K) \geq \frac{1}{D^2}.$$

These results apply in particular when S is a conjugacy class.

PROOF: Let $e_i = (x_i, x_i s_i) \in \mathcal{A}$, $x_i \in G$, $s_i \in S \cup S^{-1}$, $i = 1, 2$ be two edges. If $s_1, s_2 \in S$, there exists $\sigma \in H$ such that $\sigma(s_1) = s_2$. Set $\sigma(x_1) = y_1$. Then $z \to x_2 y_1^{-1} \sigma(z)$ is an automorphism of G which send x_1 to x_2 and $x_1 s_1$ to $x_2 s_2$. A similar reasoning applies if $s_1, s_2 \in S^{-1}$. Hence there are atmost two transitive classes of edges. If there are two classes, $(x, xs) \to (x, xs^{-1})$ establishes a bijection between them. Hence $|\mathcal{A}_1| = |\mathcal{A}_2| = |\mathcal{A}|/2$. Hence the desired results follow from Corollary 3.2.6.

EXAMPLE 3.2.9: Let $\mathcal{X} = S_n$ be the symmetric group on n objects. Let $K(x, y) = 0$ unless $y = x\sigma_i$ with $\sigma_i = (1, i)$ and $i = \{2, \dots, n\}$, in which case $K(x, y) = 1/(n-1)$. Decomposing any permutation θ in to disjoint cycles shows that θ is a product of at most n transpositions. Further more, any transposition (i, j) can be written as $(i, j) = (1, i)(1, j)(1, i)$. Hence any permutation is a product of at most $3n$ σ_i's and Corollary 3.2.7 yields $\lambda \geq 9n^3$. However, the subgroup $S_{n-1}(1) \subset S_n$ of the permutations that fixe 1 acts by conjugaison on S_n. Set $\psi_h : x \to hxh^{-1}$, $h \in S_{n-1}(1)$ and $H = \{\psi_h : S_n \to S_n :\ h \in S_{n-1}(1)\}$. This group of automorphisms of S_n acts transitively on $S = \{\sigma_i : i \in \{2, \dots, n\}\}$. Indeed, for $2 \leq i, j \leq n$, $h = (i, j) \in S_{n-1}(1)$ satisfies $\psi_h(\sigma_i) = \sigma_j$. Hence Corollary 3.2.8 gives the improved bound $\lambda \geq 9n^2$. The right answer is that $\lambda = 1/n$ by Fourier analysis [42].

To conclude this section we observe that there is no reason why we should choose between using a weight function as in Theorem 3.2.3 or using a flow as in Theorem 3.2.5. Furthermore we can consider more general weight functions

$$w : \Gamma \times \mathcal{A} \to (0, \infty)$$

where the weight $w(\gamma, e)$ of an edge also depends on which path γ we are considering. Again, we set $|\gamma|_w = \sum_{e \in \gamma} w(\gamma, e)^{-1}$. Then we have

Theorem 3.2.9 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set, w a generalized weight function and ϕ a flow. Then $\lambda \geq 1/A(w, \phi)$ where*

$$A(w, \phi) = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{\gamma \in \Gamma: \\ \gamma \ni e}} w(\gamma, e) |\gamma|_w \phi(\gamma) \right\}.$$

3.3 Isoperimetry

3.3.1 Isoperimetry and spectral gap

It is well known that spectral gap bounds can be obtained through isoperimetric inequalities via the so-called Cheeger's inequality introduced in a different context in Cheeger [12]. See Alon [5], Alon and Milman [6], Sinclair [71, 72], Diaconis and Stroock [35], Kannan [56], and the earlier references given there. See also [58]. This section presents this technique. It emphasizes the fact that isoperimetric inequalities are simply ℓ^1 version of Poincaré inequalities. It follows that in most circumstances it is possible and preferable to work directly with Poincaré inequalities if the ultimate goal is to bound the spectral gap. Diaconis and Stroock [35] compare bounds using Theorems 3.2.1, 3.2.3, and bounds using Cheeger's inequality. They find that, most of the time, bounds using Cheeger's inequality can be tightned by appealing directly to a Poincaré inequality.

Definition 3.3.1 *The "boundary" ∂A of a set $A \subset \mathcal{X}$ is the set*

$$\partial A = \{e = (x,y) \in \mathcal{X} \times \mathcal{X} : \ x \in A, y \in A^c \ or \ x \in A^c, y \in A\}.$$

Thus, the boundary is the set of all pairs connecting A and A^c.

Given a Markov chain (K,π), the measure of the boundary ∂A of $A \subset X$ is

$$Q(\partial A) = \frac{1}{2} \sum_{x \in A, y \in A^c} (K(x,y)\pi(x) + K(y,x)\pi(y)).$$

The "boundary" ∂A is a rather large boundary and does not depend on the chain (K,π) under consideration. However, only the portion of ∂A that has positive Q-measure will be of interest to us so that we could as well have required that the edges in ∂A satisfy $Q(e) > 0$.

Definition 3.3.2 *The isoperimetric constant of the chain (K,π) is defined by*

$$I = I(K,\pi) = \min_{\substack{A \subset \mathcal{X}: \\ \pi(A) \leq 1/2}} \left\{ \frac{Q(\partial A)}{\pi(A)} \right\}. \tag{3.3.1}$$

Let us specialize this definition to the case where (K,π) is the simple random walk on an r-regular graph $(\mathcal{X},\mathcal{A})$. Then, $K(x,y) = 1/r$ if x,y are neighbors and $\pi(x) \equiv 1/|\mathcal{X}|$. Hence $Q(e) = 1/(r|\mathcal{X}|)$ if $e \in \mathcal{A}$. Define the geometric boundary of a set A to be

$$\partial_* A = \{(x,y) \in \mathcal{A} : x \in A, y \in A^c\}.$$

Then

$$I = \min_{\substack{A \subset \mathcal{X}: \\ \pi(A) \leq 1/2}} \left\{ \frac{Q(\partial A)}{\pi(A)} \right\} = \frac{2}{r} \min_{\substack{A \subset \mathcal{X}: \\ \#A \leq \#\mathcal{X}/2}} \left\{ \frac{\#\partial_* A}{\#A} \right\}.$$

Lemma 3.3.3 *The constant I satisfies*

$$I = \min_f \left\{ \frac{\sum_e |df(e)|Q(e)}{\min_\alpha \sum_x |f(x) - \alpha|\pi(x)} \right\}.$$

Here the minimum is over all non-constant fonctions f.

It is well known and not too hard to prove that

$$\min_\alpha \sum_x |f(x) - \alpha|\pi(x) = \sum_x |f(x) - \alpha_0|\pi(x)$$

if and only if α_0 satisfies

$$\pi(f > \alpha_0) \leq 1/2 \text{ and } \pi(f < \alpha_0) \leq 1/2$$

i.e., if and only if α_0 is a median.

PROOF: Let J be the right-hand side in the equality above. To prove that $I \geq J$ it is enough to take $f = 1_A$ in the definition of J. Indeed,

$$\sum_e |d1_A(e)|Q(e) = Q(\partial A), \quad \sum_x 1_A(x)\pi(x) = \pi(A).$$

We turn to the proof of $J \geq I$. For any non-negative function f, set $F_t = \{f \geq t\}$ and $f_t = 1_{F_t}$. Then observe that $f(x) = \int_0^\infty f_t(x)dt$,

$$\pi(f) = \int_0^\infty \pi(F_t)dt$$

and

$$\sum_e |df(e)|Q(e) = \int_0^\infty Q(\partial F_t)dt. \tag{3.3.2}$$

This is a discrete version of the so-called co-area formula of geometric measure theory. The proof is simple. Write

$$\begin{aligned}
\sum_e |df(e)|Q(e) &= 2 \sum_{\substack{e=(x,y) \\ f(y)>f(x)}} (f(y) - f(x))Q(e) \\
&= 2 \sum_{\substack{e=(x,y) \\ f(y)>f(x)}} \int_{f(x)}^{f(y)} Q(e)dt \\
&= 2 \int_0^\infty \sum_{\substack{e=(x,y) \\ f(y)\geq t>f(x)}} Q(e)dt \\
&= \int_0^\infty Q(\partial F_t)dt.
\end{aligned}$$

Given a function f, let α be such that $\pi(f > \alpha) \leq 1/2$, $\pi(f < \alpha) \leq 1/2$ and set $f_+ = (f - \alpha) \vee 0$, $f_- = -[(f - \alpha) \wedge 0]$. Then, $f_+ + f_- = |f - \alpha|$ and $|df(e)| = |df_+(e)| + |df_-(e)|$. Setting $F_{\pm,t} = \{x : f_\pm(x) \geq t\}$, using (3.3.2) and the definition of I, we get

$$\begin{aligned}
\sum_e |df(e)|Q(e) &= \sum_e |df_+(e)|Q(e) + \sum_e |df_-(e)|Q(e) \\
&= \int_0^\infty Q(\partial F_{+,t})dt + \int_0^\infty Q(\partial F_{-,t})dt \\
&\geq I \int_0^\infty (\pi(F_{+,t}) + \pi(F_{-,t}))dt
\end{aligned}$$

$$
\begin{aligned}
&= I\sum_x (f_+(x)+f_-(x))\pi(x) \\
&= I\sum_x |f(x)-\alpha|\pi(x).
\end{aligned}
$$

This proves that $J \geq I$.

There is an alternative notion of isoperimetric constant that is sometimes used in the literature.

Definition 3.3.4 *Define the isoperimetric constant I' of the chain (K,π) by*

$$
I' = I'(K,\pi) = \min_{A\subset \mathcal{X}} \left\{ \frac{Q(\partial A)}{2\pi(A)(1-\pi(A))} \right\}. \tag{3.3.3}
$$

Observe that $I/2 \leq I' \leq I$.

Lemma 3.3.5 *The constant I' is also given by*

$$
I' = \min_f \left\{ \frac{\sum_e |df(e)|Q(e)}{\sum_x |f(x)-\pi(f)|\pi(x)} \right\}
$$

where the minimum is taken over all non-constant functions f.

PROOF: Setting $f = 1_A$ in the ratio appearing above shows that the left-hand side is not smaller than the right-hand side. To prove the converse, set $f_+ = f\vee 0$, and $F_t = \{x : f_+(x) \geq t\}$. As in the proof of Lemma 3.3.3, we obtain

$$
\sum_e |df_+(e)|Q(e) \geq 2I' \int_0^\infty \pi(F_t)(1-\pi(F_t))dt.
$$

Now,

$$
\begin{aligned}
2\pi(F_t)(1-\pi(F_t)) &= \sum_x |1_{F_t}(x) - \pi(1_{F_t})|\pi(x) \\
&= \max_{\substack{g;\pi(g)=0 \\ \min_\alpha |g-\alpha|\leq 1}} \sum_x 1_{F_t}(x)\, g(x)\pi(x).
\end{aligned}
$$

Here, we have used the fact that, for any function u,

$$
\sum_x |u(x)-\pi(u)|\pi(x) = \max_{\substack{g;\pi(g)=0 \\ \min_\alpha |g-\alpha|\leq 1}} \sum_x u(x)\, g(x)\pi(x).
$$

See [68]. Thus, for any g satifying $\pi(g) = 0$ and $\min_\alpha |g-\alpha| \leq 1$,

$$
\begin{aligned}
\sum_e |df_+(e)|Q(e) &\geq I' \sum_x \left(\int_0^\infty 1_{F_t}(x)\, dt \right) g(x)\pi(x) \\
&\geq I' \sum_x f_+(x) g(x)\pi(x).
\end{aligned}
$$

The same reasoning applies to $f_- = -[f \wedge 0]$ so that, for all g as above,

$$\sum_e |df_-(e)|Q(e) \geq I' \sum_x f_-(x)g(x)\pi(x).$$

Adding the two inequalities, and taking the supremum over all allowable g, we get

$$\sum_e |df(e)|Q(e) \geq I' \sum_x |f(x) - \pi(f)|\pi(x)$$

which is the desired inequality.

Lemmas 3.3.3 and 3.3.5 shows that the argument used in the proof of Theorem 3.2.1 can be used to bound I and I' from below.

Theorem 3.3.6 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set. For each $(x,y) \in \mathcal{X} \times \mathcal{X}$ choose exactly one path $\gamma(x,y)$ in $\Gamma(x,y)$. Then $I \geq I' \geq 1/B$ where*

$$B = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y \in \mathcal{X}: \\ \gamma(x,y) \ni e}} \pi(x)\pi(y) \right\}.$$

PROOF: For each $(x,y) \in \mathcal{X} \times \mathcal{X}$, write $f(y) - f(x) = \sum_{e \in \gamma(x,y)} df(e)$ and

$$|f(y) - f(x)| \leq \sum_{e \in \gamma(x,y)} |df(e)|.$$

Multiply by $\pi(x)\pi(y)$ and sum over all x, y to obtain

$$\sum_{x,y} |f(y) - f(x)|\pi(x)\pi(y) \leq \sum_{x,y} \sum_{e \in \gamma(x,y)} |df(e)|\pi(x)\pi(y).$$

This yields

$$\sum_x |f(x) - \pi(f)|\pi(x) \leq B \sum_e |df(e)|Q(e)$$

which implies the desired conclusion. There is also a version of this result using flows as in Theorem 3.2.5.

Lemma 3.3.7 (Cheeger's inequality) *The spectral gap λ and the isoperimetric constant I, I' defined at* (3.3.1), (3.3.3) *are related by*

$$\frac{I'^2}{8} \leq \frac{I^2}{8} \leq \lambda \leq I' \leq I.$$

Compare with [35], Section 3.C. There, it is proved by a slightly different argument that $h^2/2 \leq \lambda < 2h$ where $h = I/2$. This is the same as $I^2/8 \leq \lambda \leq I$.

PROOF: For the upper bound use the test functions $f = 1_A$ in the definition of λ. For the lower bound, apply

$$\sum_e |df(e)|Q(e) \geq I \min_\alpha \sum_x |f(x) - \alpha|\pi(x)$$

to the function $f = |g-c|^2 \mathrm{sgn}(g-c)$ where g is an arbitrary function and $c = c(g)$ is a median of g so that $\sum_x |f(x) - \alpha|\pi(x)$ is minimum for $\alpha = 0$. Then, for $e = (x, y)$,

$$|df(e)| \leq |dg(e)|(|g(x) - c| + |g(y) - c|)$$

because $|a^2 - b^2| = |a - b|(|a| + |b|)$ if $ab \geq 0$ and $a^2 + b^2 \leq |a - b|(|a| + |b|)$ if $ab < 0$. Hence

$$\begin{aligned}
\sum_e |df(e)|Q(e) &\leq \sum_{e=(x,y)} |dg(e)|(|g(x) - c| + |g(y) - c|)Q(e) \\
&\leq \left(\sum_e |dg(e)|^2 Q(e)\right)^{1/2} \times \\
&\quad \left(2\sum_{x,y}(|g(x) - c|^2 + |g(y) - c|^2)\pi(x)K(x,y)\right)^{1/2} \\
&= (8\mathcal{E}(g,g))^{1/2}\left(\sum_x |g(x) - c|^2\pi(x)\right)^{1/2}.
\end{aligned}$$

Hence

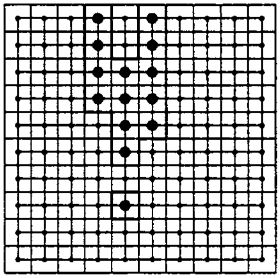
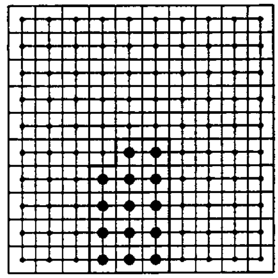
$$\begin{aligned}
I\sum_x |g(x) - c|^2\pi(x) &= I \min_\alpha \sum_x |f(x) - \alpha|^2\pi(x) \\
&\leq \sum_e |df(e)|Q(e) \\
&\leq (8\mathcal{E}(g,g))^{1/2}\left(\sum_x |g(x) - c|^2\pi(x)\right)^{1/2}.
\end{aligned}$$

and

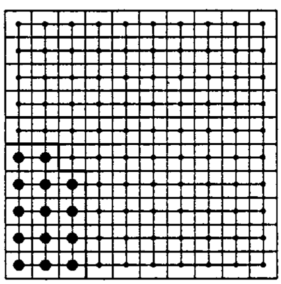
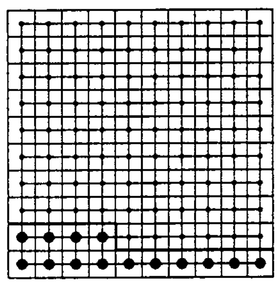
$$I^2 \mathrm{Var}_\pi(g) \leq I^2 \sum_x |g(x) - c|^2\pi(x) \leq 8\mathcal{E}(g,g).$$

for all functions g. This proves the desired lower bound.

EXAMPLE 3.3.1: Let $\mathcal{X} = \{0, \ldots, n\}^2$ be the vertex set of a square grid of side n. Hence, the edge set $\mathcal{A}$ is given by $\mathcal{A} = \{(x,y) \in \mathcal{X}^2 : |x-y| = 1\}$ where $|x-y|$ denote either the Euclidian distance or simply $\sum_i |x_i - y_i|$ (it does not matter which). Define $K(x,y)$ to be zero if $|x - y| \geq 1$, $K(x,y) = 1/4$ if $|x - y| = 1$, and $K(x,x) = 0$, $1/4$ or $1/2$ depending on whether x is interior, on a side, or a corner of $\mathcal{X}$. The uniform distribution $\pi \equiv 1/(n+1)^2$ is the reversible measure of K. To have a more geometric interpretation of the boundary, we view each vertex in $\mathcal{X}$ as the center of a unit square as in the figure below.

Then, for any subset $A \subset \mathcal{X}$, $\pi(A)$ is proportional to the surface of those unit squares with center in A. Call $\mathbf{A}$ the union of those squares (viewed as a subset of the plane). Now $Q(\partial A)$ is proportional to the length of the interior part of the boundary of $\mathbf{A}$. It is not hard to see that pushing all squares in each column down to the bottom leads to a set $\mathbf{A}^{\downarrow}$ with the same area and smaller boundary.

Similarly, we can push things left. Then consider the upper left most unit square. It is easy to see that moving it down to the left bottom most free space does not increase the boundary. Repeating this operation as many times as possible shows that, given a number N of unit squares, the smallest boundary is obtained for the set formed with $[N/(n+1)]$ bottom raws and the $N-(n+1)[N/(n+1)]$ left most squares of the $([N/(n+1)]+1)^{th}$ raw. Hence, we have

$$\frac{Q(\partial A)}{\pi(A)} = \begin{cases} \frac{N+1}{4N} & \text{if } \#A = N \leq n+1 \\ \frac{n+2}{4N} & \text{if } n+1 \leq \#A = N \text{ and } \#A \text{ does not divide } n+1 \\ \frac{n+1}{4N} & \text{if } \#A = N = k(n+1). \end{cases}$$

Theorem 3.3.8 *For the natural walk on the square grid $\mathcal{X} = \{0,\ldots,n\}^2$ the isoperimetric constants I, I' are given by*

$$I = \begin{cases} \frac{1}{2(n+1)} & \textit{if } n+1 \textit{ is even} \\ \frac{1}{2n} & \textit{if } n+1 \textit{ is odd.} \end{cases} \qquad I' = \begin{cases} \frac{1}{2(n+1)} & \textit{if } n+1 \textit{ is even} \\ \frac{1}{2n(1+(n+1)^{-2})} & \textit{if } n+1 \textit{ is odd.} \end{cases}$$

Using Cheeger's inequality yields

$$\lambda \geq \frac{1}{32(n+1)^2}.$$

This is of the right order of magnitude.

EXAMPLE 3.3.2: For comparison, consider the example of the "n-dog". That is, two square grids as above with one corner o identified. In this case, it is clear that the ratio $Q(\partial A)/\pi(A)$ (with $\pi(A) \leq 1/2$) is smallest for A one of the two squares minus o. Hence

$$I(n\text{-dog}) = \frac{1}{2[(n+1)^2-1]}.$$

In this case Cheeger's inequality yields

$$\lambda(n\text{-dog}) \geq \frac{1}{32(n+1)^4}.$$

This is far off from the right order of magnitude $1/(n^2 \log n)$ which was found using Theorem 3.2.3.

The proof of Theorem 3.3.8 works as well in higher dimension and for rectangular boxes.

Theorem 3.3.9 *For the natural walk on the parallelepiped*

$$\mathcal{X} = \{0,\dots,n_1\} \times \dots \times \{0,\dots,n_d\}$$

with $n_1 = \max n_i$, the isoperimetric constants I, I' satisfy

$$I \geq I' \geq \frac{1}{d(n_1+1)}.$$

In this case, Cheeger's inequality yields a bound which is off by a factor of $1/d$.

The above examples must not lead the reader to believe that, generaly speaking, isoperimetric inequalities are easy to prove or at least easier to prove than Poincaré inequalities. It is the case in some examples as the ones above whose geometry is really simple. There are other examples where the spectral gap is known exactly (e.g., by using Fourier analysis) but where even the order of magnitude of the isoperimetric constant I is not known. One such example is provided by the walk on the symmetric group S_n with $K(x,y) = 2/n(n-1)$ if x and y differ by a transposition and $K(x,y) = 0$ otherwise. For this walk $\lambda = 2/(n-1)$ and, by Cheeger's inequality, $2/(n-1) \leq I \leq 4/(n-1)^{1/2}$.

3.3.2 Isoperimetry and Nash inequalities

The goal of this section is to prove the following result.

Theorem 3.3.10 *Assume that (K,π) satisfies*

$$\pi(A)^{(d-1)/d} \leq S\left(Q(\partial A) + \frac{1}{R}\pi(A)\right) \tag{3.3.4}$$

for all $A \subset \mathcal{X}$ and some constants $d \geq 1$, $S, R > 0$. Then

$$\forall\, g,\ \ \|g\|_{d/(d-1)} \leq S\left(\sum_e |dg(e)|Q(e) + \frac{1}{R}\|g\|_1\right) \tag{3.3.5}$$

and

$$\forall\, g,\ \ \|g\|_2^{2(1+2/d)} \leq 16\, S^2\left(\mathcal{E}(g,g) + \frac{1}{8R^2}\|g\|_2^2\right)\|g\|_1^{4/d}. \tag{3.3.6}$$

PROOF: Since $|d|g|(e)| \leq |dg(e)|$ it suffices to prove the result for $g \geq 0$. Write $g = \int_0^\infty g_t dt$ where $g_t = 1_{G_t}$, $G_t = \{g \geq t\}$, and set $q = d/(d-1)$. Then

$$\begin{aligned}
\|g\|_q &\leq \int_0^\infty \|g_t\|_q dt = \int_0^\infty \pi(G_t)^{1/q} dt \\
&\leq S\int_0^\infty \left(Q(\partial G_t) + \frac{1}{R}\pi(G_t)\right) dt \\
&= S\left(\sum_e |dg(e)|Q(e) + \frac{1}{R}\|g\|_1\right).
\end{aligned}$$

The first inequality uses Minkowski's inequality. The second inequality uses (3.3.4). The last inequality uses the co-area formula (3.3.2). This proves (3.3.5). It is easy to see that (3.3.5) is in fact equivalent to (3.3.4) (take $g = 1_A$).

To prove (3.3.6), we observe that

$$\sum_e |dg^2(e)|Q(e) \leq [8\mathcal{E}(g,g)]^{1/2}\|g\|_2.$$

Indeed,

$$\begin{aligned}
\sum_e |dg^2(e)|Q(e) &= \sum_{e=(x,y)} |dg(e)||g(x)+g(y)|Q(e) \\
&\leq \left(\sum_e |dg(e)|^2 Q(e)\right)^{1/2} \times \\
&\quad \left(2\sum_{x,y}(|g(x)|^2 + |g(y)|^2)\pi(x)K(x,y)\right)^{1/2} \\
&= (8\mathcal{E}(g,g))^{1/2}\left(\sum_x |g(x)|^2\pi(x)\right)^{1/2}.
\end{aligned}$$

Thus, (3.3.5) applied to g^2 yields

$$\|g\|_{2q}^2 \leq S\left([8\mathcal{E}(g,g)]^{1/2}\|g\|_2 + \frac{1}{R}\|g\|_2^2\right)$$

with $q = d/(d-1)$. The Hölder inequality

$$\|g\|_2 \leq \|g\|_1^{1/(1+d)}\|g\|_{2q}^{d/(1+d)}$$

and the last inequality let us bound $\|g\|_2$ by

$$\left(S\left([8\mathcal{E}(g,g)]^{1/2}\|g\|_2+\frac{1}{R}\|g\|_2^2\right)\right)^{1/[2(1+d)]}\|g\|_1^{1/(1+d)}$$

We raise this to the power $2(1+d)/d$ and divide by $\|g\|_2$ to get

$$\|g\|_2^{(1+2/d)}\leq S\left([8\mathcal{E}(g,g)]^{1/2}+\frac{1}{R}\|g\|_2\right)\|g\|_1^{2/d}.$$

This yields the desired result.

There is a companion result related to Theorem 2.3.1 and Nash inequalities of type (2.3.1) versus (2.3.3).

Theorem 3.3.11 *Assume that* (K,π) *satisfies*

$$\pi(A)^{(d-1)/d}\leq SQ(\partial A) \tag{3.3.7}$$

for all $A\subset\mathcal{X}$ *such that* $\pi(A)\leq 1/2$. *Then*

$$\forall\, g\in\ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)}\leq 8S^2\mathcal{E}(g,g)\|f\|_1^{4/d}.$$

Before proving this theorem, let us introduce the isoperimetric constant associated with inequality (3.3.7).

Definition 3.3.12 *The d-dimensional isoperimetric constant of a finite chain* (K,π) *is defined by*

$$I_d=I_d(K,\pi)=\min_{\substack{A\subset\mathcal{X}:\\ \pi(A)\leq 1/2}}\frac{Q(\partial A)}{\pi(A)^{1/q}}$$

where $q=d/(d-1)$.

Observe that $I\geq I_d$ with I the isoperimetric constant defined at (3.3.1) (in fact $I\geq 2^{1/d}I_d$). It may be helpful to specialize this definition to the case where (K,π) is the simple random walk on a r-regular connected symmetric graph $(\mathcal{X},\mathcal{A})$. Then $Q(e)=1/|\mathcal{A}|=1/(r|\mathcal{X}|)$, $\pi\equiv 1/|\mathcal{X}|$ and

$$I_d=\frac{2}{r|\mathcal{X}|^{1/d}}\min_{\substack{A\subset\mathcal{X}:\\ \#A\leq\#\mathcal{X}/2}}\frac{\#\partial_* A}{[\#A]^{1/q}}$$

where $\partial_* A=\{(x,y)\in\mathcal{A}:x\in A,\ y\notin A\}$.

Lemma 3.3.13 *The isoperimetric constant* $I_d(K,\pi)$ *is also given by*

$$I_d(K,\pi)=\inf\left\{\frac{\sum_e|df(e)|Q(e)}{\|f-c(f)\|_q}:f\ \textit{non-constant}\right\}$$

where $q=d/(d-1)$ *and* $c(f)$ *denote the smallest median of* f.

PROOF: For $f = \mathbf{1}_A$ with $\pi(A) \leq 1/2$, $c(f) = 0$ is the smallest median of f. Hence

$$\frac{\sum_e |df(e)|Q(e)}{\|f - c(f)\|_q} = \frac{Q(\partial A)}{\pi(A)^{1/q}}.$$

It follows that

$$\min_f \left\{ \frac{\sum_e |df(e)|Q(e)}{\|f - c(f)\|_q} \right\} \leq I_d(K, \pi).$$

To prove the converse, fix a function f and let c be such that $\pi(f > c) \leq 1/2$, $\pi(f < c) \leq 1/2$. Set $f_+ = (f-c) \vee 0$, $f_- = -[(f-c) \wedge 0]$. Then $f_+ + f_- = |f-c|$ and $|df(e)| = |df_+(e)| + |df_-(e)|$. Setting $F_{\pm,t} = \{x : f_\pm(x) \geq t\}$ and using (3.3.2) we obtain

$$\begin{aligned}
\sum_e |df(e)|Q(e) &\geq \sum_e |df_+(e)|Q(e) + \sum_e |df_-(e)|Q(e) \\
&= \int_0^\infty Q(\partial F_{+,t})dt + \int_0^\infty Q(\partial F_{-,t})dt \\
&\geq I_d \int_0^\infty \left(\pi(F_{+,t})^{1/q} + \pi(F_{-,t})^{1/q} \right) dt.
\end{aligned}$$

Now

$$\pi(F_{\pm,t})^{1/q} = \|\mathbf{1}_{F_{\pm,t}}\|_q = \max_{\|g\|_r \leq 1} \langle \mathbf{1}_{F_{\pm,t}}, g \rangle$$

where $1/r + 1/q = 1$. Hence, for any g such that $\|g\|_r \leq 1$,

$$\begin{aligned}
\sum_e |df(e)|Q(e) &\geq I_d \int_0^\infty \left(\langle \mathbf{1}_{F_{+,t}}, g \rangle + \langle \mathbf{1}_{F_{-,t}}, g \rangle \right) \\
&= I_d \left(\langle f_+, g \rangle + \langle f_-, g \rangle \right) \\
&= I_d \, \langle |f - c|, g \rangle.
\end{aligned}$$

Taking the supremum over all g with $\|g\|_r \leq 1$ we get

$$\sum_e |df(e)|Q(e) \geq I_d \|f - c\|_q. \tag{3.3.8}$$

The desired inequality follows. Observe that in (3.3.8) c is a median of f.

PROOF OF THEOREM 3.3.11: Fix g and set $f = \mathrm{sgn}(g - c)|g - c|^2$ where c is a median of g, hence 0 is a median of f. The hypothesis of Theorem 3.3.11 implies that $I_d \geq 1/S$. Inequality (3.3.8) then shows that

$$\|g - c\|_{2q}^2 = \|f\|_q \leq S \sum_e |df(e)|Q(e)|.$$

As in the proof of Lemma 3.3.7 we have

$$\sum_e |df(e)|Q(e) \leq [8\,\mathcal{E}(g, g)]^{1/2} \|g - c\|_2.$$

Hence

$$\|g-c\|_{2q}^2 \le [8S^2\,\mathcal{E}(g,g)]^{1/2}\|g-c\|_2.$$

Now, the Hölder inequality $\|h\|_2 \le \|h\|_1^{1/(1+d)}\|h\|_{2q}^{d/(1+d)}$ yields

$$\|g-c\|_2 \le \left([8S^2\,\mathcal{E}(f,f)]^{1/2}\|g-c\|_2\right)^{d/2(1+d)}\|g-c\|_1^{1/(1+d)}.$$

Thus

$$\|g-c\|_2^{2(1+2/d)} \le 8S^2\mathcal{E}(f,f)\|g-c\|_1^{4/d}.$$

Since c is a median of g, it follows that

$$\mathrm{Var}_\pi(g)^{1+2/d} \le 8S^2\,\mathcal{E}(f,f)\|g\|_1^{4/d}.$$

This is the desired result.

EXAMPLE 3.3.3: Consider a square grid $\mathcal{X}=\{0,\ldots,n\}^2$ as in Theorem 3.3.8. The argument developed for Theorem 3.3.8 also yields the following result.

Theorem 3.3.14 *For the natural walk on the square grid $\mathcal{X}=\{0,\ldots,n\}^2$ the isoperimetric constant I_2 (i.e., $d=2$) is given by*

$$I_2 = \begin{cases} \frac{1}{2^{3/2}(n+1)} & \text{if } n+1 \text{ is even} \\ \frac{(n+2)^{1/2}}{2^{3/2}n^{1/2}(n+1)} & \text{if } n+1 \text{ is odd.} \end{cases}$$

By Theorem 3.3.11 *it follows that, for all $f \in \ell^2(\pi)$,*

$$\mathrm{Var}_\pi(f)^2 \le 64(n+1)^2\mathcal{E}(f,f)\|f\|_1^2.$$

By Theorem 2.3.2 *this yields*

$$\|h_t^x - 1\|_2 \le \min\left\{2^{3/2}(n+1)/t^{1/2}, e^{-[t/64(n+1)^2]+1/2}\right\}.$$

This is a very good bound which is of the right order of magnitude **for all** $t>0$.

EXAMPLE 3.3.4: We can also compute I_d for a paralellepiped in d-dimensions.

Theorem 3.3.15 *For the natural walk on the parallelepiped*

$$\mathcal{X}=\{0,\ldots,n_1\}\times\ldots\times\{0,\ldots,n_d\}$$

with $n_i \le n_1$, the isoperimetric constant I_d satisfies

$$I_d \ge \frac{1}{d2^{1-1/d}(n_1+1)}$$

with equality if n_1+1 is even. It follows that

$$\mathrm{Var}_\pi(f)^{1+2/d} \le 82^{2(1-1/d)}\,d^2\,(n_1+1)^2\,\mathcal{E}(f,f)\|f\|_1^{4/d}.$$

In [28] a somewhat better Nash inequality

$$\|f\|_2^{1+2/d} \leq 64\, d\,(n+1)^2 \left(\mathcal{E}(f,f) + \frac{8}{d(n+1)^2}\|f\|_2^2\right) \|f\|_1^{4/d}$$

is proved (in the case $n_1 = \ldots = n_d = n$) by a different argument.

EXAMPLE 3.3.5: We now return to the "n-dog". The Nash inequality in Theorem 3.3.14 yields

$$\begin{aligned}\|f\|_2^2 &\leq \left(64(n+1)^2\mathcal{E}(f,f)\|f\|_1^2\right)^{1/2} + \pi(f)^2 \\ &\leq \left(64(n+1)^2\left(\mathcal{E}(f,f) + \frac{1}{64(n+1)^2}\|f\|_2^2\right)\|f\|_1^2\right)^{1/2}.\end{aligned}$$

for all functions f on a square grid $\{0,\ldots,n\}^2$. Now the n-dog is simply two square grids with one corner in common. Hence, applying the above inequality on each square grid, we obtain (the constant factor between the uniform distribution on one grid and the uniform distribution on the n-dog cancel)

$$\|f\|_2^4 \leq 128(n+1)^2\left(\mathcal{E}(f,f) + \frac{1}{32(n+1)^2}\|f\|_2^2\right)\|f\|_1^2.$$

The change by a factor 2 in the numerical constants is due to the fact that the common corner o appears in each square grid. Recall that using Theorem 3.2.3 we have proved that the spectral gap of the dog is bounded below by

$$\lambda \geq \frac{1}{8(n+1)^2\log(2n+1)}.$$

Applying Theorem 2.3.5 and Corollary 2.3.5, we obtain the following result.

Theorem 3.3.16 *For the n-dog made of two square grids $\{0,\ldots,n\}^2$ with the corners $o = o_1 = o_2 = (0,0)$ identified, the natural chain satisfies*

$$\forall\, t \leq 32(n+1)^2, \quad \|h_t^x\|_2 \leq 8e(n+1)/t^{1/2}.$$

Also, for all $c > 0$ and $t = 8(n+1)^2(5 + c\log(2n+1))$

$$\|h_t^x - 1\|_2 \leq e^{1-c}.$$

This shows that a time of order $n^2\log n$ suffices to reach stationarity on the n-dog. Furthermore, the upper bound on λ that we obtained earlier shows that this is optimal since $\max_x \|h_t^x - 1\|_1 \geq e^{-t\lambda} \geq e^{-at/(n^2\log n)}$.

Consider now all the eigenvalues $1 = \lambda_0 < \lambda_1 \leq \ldots \leq \lambda_{|\mathcal{X}|-1}$ of this chain. Corrolary 2.3.9 and Theorem 3.3.16 show that

$$\lambda_i \geq 10^{-4}(i+1)n^{-2}$$

for all $i \geq 10^4$. This is a good estimate except for the numerical constant 10^4. However, it leaves open the following natural question. We know that $\lambda = \lambda_1$ is of order $1/(n^2 \log n)$. How many eigenvalues are there such that $n^2\lambda_i$ tends to zero as n tends to infinity? Interestingly enough the answer is that λ_1 is the only such eigenvalue. Namely, there exists a constant $c > 0$ such that, for $i \geq 2$, $\lambda_i \geq cn^{-2}$. We now prove this fact. Consider the squares

$$\mathcal{X}_- = \{-n, \ldots, 0\}^2, \quad \mathcal{X}_+ = \{0, \cdots, n\}^2$$

and set

$$\psi_\pm(x) = \mathbf{1}_{\mathcal{X}_\pm}(x), \quad x \in \mathcal{X}.$$

These functions span a two-dimensional vector space $E \subset \ell^2(\mathcal{X})$. On each of the two squares $\mathcal{X}_-, \mathcal{X}_+$, we have the Poincaré inequality

$$\sum_{x \in \mathcal{X}_\pm} |f(x)|^2 \leq \frac{1}{4}(n+1)^2 \sum_e |df(e)|^2 \tag{3.3.9}$$

for all function f on $\mathcal{X}_\pm$ satisfying $\sum_{x \in \mathcal{X}_\pm} f(x) = 0$. In this inequality, the right most sum runs over all edge e of the grid $\mathcal{X}_\pm$. There are many ways to prove this inequality. For instance, one can use Theorem 3.2.1 (with paths having only one turn), or the fact that the spectral gap is exactly $1 - \cos(\pi/(n+1))$ for the square grid.

Now, if f is a function in $\ell^2(\mathcal{X})$ which is orthogonal to E (i.e., to ψ_- and ψ_+), we can apply (3.3.9) to the restrictions f_+, f_- of f to $\mathcal{X}_+$, $\mathcal{X}_-$. Adding up the two inequalities so obtained we get

$$\forall \ f \in E^\perp, \quad \sum_{x \in \mathcal{X}} |f(x)|^2 \pi(x) \leq 2(n+1)^2 \mathcal{E}(f,f).$$

By the min-max principle (1.2.13), this shows that

$$\lambda_2 \geq \frac{1}{2(n+1)^2}.$$

Let ψ_1 denote the normalized eigenfunction associated to the spectral gap λ. For each n, let $a_n < b_n$ be such that

$$\lim_{n\to\infty} a_n n^{-2} = +\infty, \quad \lim_{n\to\infty} b_n [n^2 \log n]^{-1} = 0, \quad \lim_{n\to\infty} (b_n - a_n) = +\infty$$

and set $I_n = [a_n, b_n]$. Using the estimates obtained above for λ_1 and λ_2 together with Lemma 1.3.3 we conclude that for $t \in I_n$ and n large enough the density $h_t(x, y)$ of the semigroup H_t on the n-dog is close to

$$1 + \psi_1(x)\psi_1(y).$$

In words, the n-dog presents a sort of metastability phenomenon.

We finish this subsection by stating a bound on higher eigenvalues in terms of isoperimetry. It follows readily from Theorems 3.3.11 and 2.3.9.

Theorem 3.3.17 *Assume that* (K, π) *is reversible and satisfies* (3.3.7), *that is,*

$$\pi(A)^{(d-1)/d} \leq SQ(\partial A)$$

for all $A \subset \mathcal{X}$ *such that* $\pi(A) \leq 1/2$. *Then the eigenvalues* λ_i *satisfy*

$$\lambda_i \geq \frac{i^{2/d}}{8e^{2/d}dS^2}.$$

Compare with [14].

3.3.3 Isoperimetry and the log-Sobolev constant

Theorem 2.3.6 can be used, together with theorems 3.3.10, 3.3.11, to bound the log-Sobolev constant α from below in terms of isoperimetry. This yields the following results.

Theorem 3.3.18 *Let* (K, π) *be a finite reversible Markov chain.*

1. *Assume* (K, π) *satisfies* (3.3.7), *that is,*

$$\pi(A)^{(d-1)/d} \leq SQ(\partial A)$$

for all $A \subset \mathcal{X}$ *such that* $\pi(A) \leq 1/2$. *Then the log-Sobolev constant* α *is bounded below by*

$$\alpha \geq \frac{1}{4dS^2}.$$

2. *Assume instead that* (K, π) *satisfies* (3.3.4), *that is,*

$$\pi(A)^{(d-1)/d} \leq S\left(Q(\partial A) + \frac{1}{R}\pi(A)\right),$$

for all set $A \subset \mathcal{X}$. *Then*

$$\alpha \geq \frac{\lambda}{2\left[1 + 8R^2\lambda + \frac{d}{4}\log\left(\frac{dS^2}{2R^2}\right)\right]}.$$

EXAMPLE 3.3.6: Theorem 3.3.18 and Theorems 3.3.14, 3.3.16 prove that the two-dimensional square grid $\mathcal{X} = \{0, \ldots, n\}^2$ or the two-dimensional n-dog have $\alpha \simeq \lambda$. Namely, for the two-dimensional n-grid, α and λ are of order $1/n^2$ whereas, for the n-dog, α and λ are of order $1/[n^2 \log n]$.

EXAMPLE 3.3.7: For the d-dimensional square grid $\mathcal{X} = \{0, \ldots, n\}^d$, applying Theorems 3.3.18 and 3.3.15 we obtain

$$\alpha \geq \frac{2}{d^3(n+1)^2}$$

whereas Lemma 2.2.11 can be used to show that α is of order $1/[dn^2]$ in this case.

3.4 Moderate growth

This section presents geometric conditions that implies that a Nash inequality holds. More details and many examples can be found in [25, 26, 28]. Let us emphasize that the notions of **moderate growth** and of **local Poincaré inequality** presented briefly below are really instrumental in proving useful Nash inequalities in explicit examples. See [28].

Definition 3.4.1 *Let (K,π) be an irreducible Markov chain on a finite state space $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set according to Definition* 3.1.1. *Let $d(x,y)$ denote the distance between x and y in $(\mathcal{X},\mathcal{A})$ and $\gamma = \max_{x,y} d(x,y)$ be the diameter. Define*

$$V(x,r) = \pi(\{y : d(x,y) \le r\}).$$

(1) We say the (K,π) has (M,d)-moderate growth if

$$V(x,r) \ge \frac{1}{M}\left(\frac{r+1}{\gamma}\right)^d \quad \text{for all } x \in \mathcal{X} \text{ and all } r \le \gamma.$$

(2) We say that (K,π) satisfies a local Poincaré inequality with constant $a > 0$ if

$$\|f - f_r\|_2^2 \le ar^2\mathcal{E}(f,f) \quad \text{for all functions } f \text{ and all } r \le \gamma$$

where

$$f_r(x) = \frac{1}{V(x,r)} \sum_{y:d(x,y)\le r} f(y)\pi(y).$$

Moderate growth is a purely geometric condition. On one hand it implies (take $r=0$) that $\pi_* \ge M^{-1}\gamma^{-d}$. If π is uniform, this says $|\mathcal{X}| \le M\gamma^d$. On the other hand, it implies that the volume of a ball of radius r grows at least like r^d.

The local Poincaré inequality implies in particular (take $r=\gamma$) that $\mathrm{Var}_\pi(f) \le a\gamma^2\mathcal{E}(f,f)$, that is $\lambda \ge 1/(a\gamma^2)$. It can sometimes be checked using the following lemma.

Lemma 3.4.2 *For each $(x,y) \in \mathcal{X}^2$, $x \ne y$, fix a path $\gamma(x,y)$ in $\Gamma(x,y)$. Then*

$$\|f - f_r\|_2^2 \le \eta(r)\mathcal{E}(f,f)$$

where

$$\eta(r) = \max_{e\in\mathcal{A}} \left\{ \frac{2}{Q(e)} \sum_{\substack{x,y:d(x,y)\le r,\\ \gamma(x,y)\ni e}} |\gamma(x,y)| \frac{\pi(x)\pi(y)}{V(x,r)} \right\}.$$

See [28], Lemma 5.1.

Definition 3.4.1 is justified by the following theorem.

Theorem 3.4.3 *Assume that (K,π) has (M,d) moderate growth and satisfies a local Poincaré inequality with constant $a>0$. Then $\lambda \geq 1/a\gamma^2$ and (K,π) satisfies the Nash inequality*

$$\|f\|_2^{2(1+2/d)} \leq C\left(\mathcal{E}(f,f)+\frac{1}{a\gamma^2}\|f\|_2^2\right)\|f\|_1^{4/d}$$

with $C=(1+1/d)^2(1+d)^{2/d}M^{2/d}a\gamma^2$. It follows that

$$\|h_t^x-1\|_2 \leq Be^{-c} \quad \textit{for } t=a\gamma^2(1+c),\;\; c>0$$

with $B=(e(1+d)M)^{1/2}(2+d)^{d/4}$. Also, the log-Sobolev constant satisfies $\alpha \geq \varepsilon/\gamma^2$ with $\varepsilon^{-1}=2a(2+\log B)$.

Futhermore, there exist constants c_i, $i=1,\dots,6$, depending only on M,d,a and such that $\lambda \leq c_1/\gamma^2$, $\alpha \leq c_2/\gamma^2$ and, if (K,π) is reversible,

$$c_3e^{-c_4t/\gamma^2} \leq \max_x \|h_t^x-1\|_1 \leq c_5e^{-c_6t/\gamma^2}.$$

See [28], Theorems 5.2, 5.3 and [29], Theorem 4.1.

One can also state the following result for higher eigenvalues of reversible Markov chains.

Theorem 3.4.4 *Assume that (K,π) is reversible, has (M,d) moderate growth and satisfies a local Poincaré inequality with constant $a>0$. Then there exists a constant $c=c(M,d,a)>0$ such that $\lambda_i \geq ci^{2/d}\gamma^{-2}$.*

Chapter 4

Comparison techniques

This chapter develops the idea of comparison between two finite chains K, K'. Typically we are interested in studying a certain chain K on $\mathcal{X}$. We consider an auxilliary chain K' on $\mathcal{X}$ or even on a different but related state space $\mathcal{X}'$. This auxilliary chain is assumed to be well-known, and the chain K is not too different from K'. Comparison techniques allow us to transfer information from K to K'. We have already encounter this idea several times. It is emphasized and presented in detail in this chapter. The main references for this chapter are [23, 24, 30].

4.1 Using comparison inequalities

This section collects a number of results that are the keys of comparison techniques. Most of these results have already been proved in previous chapters, sometimes under less restrictive hypoheses.

Theorem 4.1.1 *Let (K,π), (K',π') be two irreducible finite chains defined on two state spaces $\mathcal{X}$, $\mathcal{X}'$ with $\mathcal{X} \subset \mathcal{X}'$. Assume that there exists an extention map $f \to \tilde{f}$ that associates a function $\tilde{f} : \mathcal{X} \to \mathbb{R}$ to any function $\tilde{f} : \mathcal{X}' \to \mathbb{R}$ and such that $\tilde{f}(x) = f(x)$ if $x \in \mathcal{X}$. Assume further that there exist $a, A > 0$ such that*

$$\forall f : \mathcal{X} \to \mathbb{R},\quad \mathcal{E}'(\tilde{f},\tilde{f}) \le A\mathcal{E}(f,f) \quad \textit{and} \quad \forall x \in \mathcal{X},\ a\pi(x) \le \pi'(x).$$

Then

(1) The spectral gaps λ, λ' and the log-Sobolev constants α, α' satisfy

$$\lambda \ge a\lambda'/A\,, \quad \alpha \ge a\alpha'/A.$$

In particular

$$\|h_t^x - 1\|_2 \le e^{1-c} \quad \textit{for all} \quad t = \frac{Ac}{a\lambda'} + \frac{A}{2a\alpha'}\log_+\log\frac{1}{\pi(x)} \ \textit{with}\ c > 0.$$

(2) If (K,π) and (K',π') are reversible chains, and $|\mathcal{X}|=n$, $|\mathcal{X}'|=n'$,

$$\forall i=1,\dots,n-1,\quad \lambda_i \ge a\lambda_i'/A$$

where $(\lambda_i)_0^{n-1}$ (resp $(\lambda_i')_0^{n'-1}$) are the eigenvalues of $I-K$ (resp. $I-K'$) in nondecreasing order. In particular, for all $t>0$,

$$\|h_t - 1\|^2 \;\le\; \|h'_{at/A} - 1\|^2 = \sum_1^{n'-1} e^{-2at\lambda_i'/A}$$

where

$$\|h_t - 1\|^2 = \sum_{x,y} |h_t(x,y)-1|^2\pi(x)\pi(y) = \sum_x \|h_t^x - 1\|_2^2 \pi(x).$$

(3) If (K,π) and (K',π') are reversible chains and that there exists a group G that acts transitively on $\mathcal{X}$ with $K(gx,gy)=K(x,y)$ and $\pi(gx)=\pi(x)$ then

$$\forall x\in\mathcal{X},\quad \|h_t^x - 1\|_2^2 \le \sum_1^{n'-1} e^{-2at\lambda_i'/A}.$$

(4) If (K,π) and (K',π') are invariant under transitive group actions then

$$\forall x\in\mathcal{X},\ x'\in\mathcal{X}',\quad \|h_t^x - 1\|_2 \le \|h'^{x'}_{at/A} - 1\|_2.$$

PROOF: The first assertion follows from Lemma 2.2.12 and Corollary 2.2.4. The second uses Theorem 1.2.11 and (1.2.12). The last statement simply follows from *(2)* and the fact that $\|h_t^x - 1\|_2$ does not depend on x under the hypotheses of *(3)*. Observe that the theorem applies when $\mathcal{X}=\mathcal{X}'$. In this case the extention map $f\to\tilde{f}=f$ is the identity map on functions.

These results shows how the comparison of the Dirichlet forms $\mathcal{E}$, $\mathcal{E}'$ allows us to bound the convergence of h_t towards π in terms of certain parameters related to the chain K' which we assume we understand better. The next example illustrates well this technique.

EXAMPLE 4.1.1: Let $\mathcal{Z}=\{0,1\}^n$. Fix a nonnegative sequence $\mathbf{a}=(a_i)_1^n$ and $b\ge 0$. Set

$$\mathcal{X}(\mathbf{a},b)=\mathcal{X}=\left\{x=(x_i)_1^n\in\mathcal{Z}:\sum a_i x_i\le b\right\}.$$

On this set, consider the Markov chain with Kernel

$$K_{\mathbf{a},b}(x,y)=K(x,y)=\begin{cases} 0 & \text{if } |x-y|>1\\ 1/n & \text{if } |x-y|=1\\ (n-n(x))/n & \text{if } x=y\end{cases}$$

where $n(x)=n_{\mathbf{a},b}(x)$ is the number of $y\in\mathcal{X}$ such that $|x-y|=1$, that is, the number of neighbors of x in $\mathcal{Z}$ that are in $\mathcal{X}$. Observe that this definition makes sense for any (say connected) subset of $\mathcal{Z}$. This chains is symmetric and has the uniform distibution $\pi \equiv 1/|\mathcal{X}|$ as reversible measure.

For instance, in the simple case where $a_i = 1$ for all i,

$$\mathcal{X}(1,b) = \left\{ x \in \{0,1\}^n : \sum_i x_i \leq b \right\}$$

and

$$K_{1,b}(x,y) = \begin{cases} 0 & \text{if } |x-y| > 1 \\ 1/n & \text{if } |x-y| = 1 \\ (n-b)/n & \text{if } x = y \text{ and } |x| = b. \end{cases}$$

As mentioned in the introduction, proving that a polynomial time $t = O(n^A)$ suffices to insure convergence of this chain, uniformly over all possible choices of $\mathbf{a}, b$, is an open problem.

Here we will prove a partial result for $\mathbf{a}, b$ such that $\mathcal{X}(\mathbf{a}, b)$ is big enough. Set $|x| = \sum_1^n x_i$. Set also $x \leq y$ (resp. $<$) if $x_i \leq y_i$ (resp. $<$) for $x, y \in \mathcal{Z}$. Clearly, $y \in \mathcal{X}(\mathbf{a}, b)$ and $x \leq y$ implies that $x \in \mathcal{X}(\mathbf{a}, b)$. Furthermore, if $|x - y| = 1$, then either $x < y$ or $y < x$. Set

$$V^{\downarrow}(x) = \{y \in \mathcal{Z} : |x-y| = 1, y < x\}.$$

Now, we fix $\mathbf{a} = (a_i)_1^n$ and b. For each integer c let $\mathcal{X}_c$ be the set

$$\mathcal{X}_c = \mathcal{X} \bigcup \left\{ z \in \mathcal{Z} : \sum x_i \leq c \right\}.$$

Hence $\mathcal{X}_{c+1}$ is obtained from $\mathcal{X}_c$ by adding the points z with $\sum z_i = c + 1$. On each $\mathcal{X}_c$ we consider the natural chain defined as above. We denote by

$$\mathcal{E}_c(f,f) = \frac{1}{2n|\mathcal{X}^c|} \sum_{\substack{x,y \in \mathcal{X}^c \\ |x-y|=1}} |f(x) - f(y)|^2$$

its Dirichlet form. We will also use the notation π_c, Var_c, λ_c, α_c.

Define ℓ to be the largest integer such that $\sum_{i \in I} a_i \leq b$ for all subsets $I \subset \{1, \ldots, n\}$ with $\#I = \ell$. Observe that $\mathcal{X}_c = \mathcal{X}$ for $c \leq \ell$. Also, $\mathcal{X}_n = \mathcal{Z} = \{0,1\}^n$. We claim that the following inequalities hold between the spectral gaps and log-Sobolev constants of the natural chains on $\mathcal{X}^c$, $\mathcal{X}^{c+1}$.

$$\lambda_{c+1} \leq \left(1 + \frac{2(n-c)}{c+1}\right) \lambda_c \tag{4.1.1}$$

$$\alpha_{c+1} \leq \left(1 + \frac{2(n-c)}{c+1}\right) \alpha_c. \tag{4.1.2}$$

If we can prove these inequalities, it will follow that

$$\frac{2}{n} \leq e^{2\frac{(n-\ell)^2}{\ell+1}} \lambda(\mathbf{a}, b) \tag{4.1.3}$$

$$\frac{1}{n} \leq e^{2\frac{(n-\ell)^2}{\ell+1}} \alpha(\mathbf{a}, b) \tag{4.1.4}$$

where $\lambda(\mathbf{a}, b)$ and $\alpha(\mathbf{a}, b)$ are the spectral gap and log-Sobolev constant of the chain $K = K_{\mathbf{a},b}$ on $\mathcal{X} = \mathcal{X}_{\mathbf{a},b}$. To see this use

$$\sum_{c=\ell}^{n-1} \frac{n-c}{c+1} \le (n-\ell) \sum_{\ell}^{n-1} \frac{1}{c+1} \le \frac{(n-\ell)^2}{\ell+1}.$$

To prove (4.1.1), (4.1.2) we proceed as follows. Fix $c \ge \ell$. Given a function $f : \mathcal{X}_c \to \mathbb{R}$ we extend it to a function $\tilde{f} : \mathcal{X}_{c+1} \to \mathbb{R}$ by the formula

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \mathcal{X}^c \\ \frac{1}{c+1} \sum_{y \in V^{\downarrow}(x)} f(y) & \text{if } x \in \mathcal{X}_{c+1} \setminus \mathcal{X}_c \end{cases}$$

(observe that $\#V^{\downarrow}(x) = c+1$ if $|x| = c+1$). With this definition, we have

$$\begin{aligned} \mathrm{Var}_c(f) &\le \sum_{x \in \mathcal{X}_c} |f(x) - \pi_{c+1}(\tilde{f})|^2 \frac{1}{|\mathcal{X}_c|} \\ &\le \frac{|\mathcal{X}_{c+1}|}{|\mathcal{X}_c|} \sum_{x \in \mathcal{X}_{c+1}} |\tilde{f}(x) - \pi_{c+1}(\tilde{f})|^2 \frac{1}{|\mathcal{X}_{c+1}|} \le \frac{|\mathcal{X}_{c+1}|}{|\mathcal{X}_c|} \mathrm{Var}_{c+1}(\tilde{f}) \end{aligned}$$

and, similarly, $\mathcal{L}_c(f) \le [|\mathcal{X}_{c+1}|/|\mathcal{X}_c|]\, \mathcal{L}_{c+1}(\tilde{f})$. We can also bound $\mathcal{E}_{c+1}(\tilde{f}, \tilde{f})$ in terms of $\mathcal{E}_c(f, f)$.

$$\begin{aligned} \mathcal{E}_{c+1}(\tilde{f}, \tilde{f}) &= \frac{1}{2n|\mathcal{X}_{c+1}|} \sum_{\substack{x,y \in \mathcal{X}_{c+1}: \\ |x-y|=1}} |\tilde{f}(x) - \tilde{f}(y)|^2 \\ &\le \frac{|\mathcal{X}_c|}{|\mathcal{X}_{c+1}|} \left(\frac{1}{2n|\mathcal{X}_c|} \sum_{\substack{x,y \in \mathcal{X}_c: \\ |x-y|=1}} |f(x) - f(y)|^2 \right. \\ &\qquad \left. + \frac{1}{n|\mathcal{X}_c|} \sum_{x:|x|=c+1} \sum_{y \in V^{\downarrow}(x)} |\tilde{f}(x) - f(y)|^2 \right) \\ &= \frac{|\mathcal{X}_c|}{|\mathcal{X}_{c+1}|} \left(\mathcal{E}_c(f, f) + \frac{1}{n|\mathcal{X}_c|} \mathcal{R} \right). \end{aligned}$$

We now bound $\mathcal{R}$ in terms of $\mathcal{E}_c(f, f)$. If $|x - y| = 1$, let $x \wedge y$ be the unique element in $V^{\downarrow}(x) \cap V^{\downarrow}(y)$.

$$\begin{aligned} \mathcal{R} &= \sum_{x:|x|=c+1} \sum_{y \in V^{\downarrow}(x)} |\tilde{f}(x) - f(y)|^2 \\ &\le \sum_{x:|x|=c+1} \frac{1}{2(c+1)} \sum_{y,z \in V^{\downarrow}(x)} |f(z) - f(y)|^2 \\ &\le \sum_{x:|x|=c+1} \frac{1}{c+1} |f(z) - f(z \wedge y)|^2 + |f(z \wedge y) - f(y)|^2 \end{aligned}$$

$$\leq \sum_{x:|x|=c+1} \frac{2}{c+1} \sum_{\substack{v\in V^{\downarrow}(x)\\ u\in V^{\downarrow}(v)}} |f(v)-f(u)|^2$$

$$\leq \frac{2n(n-c)|\mathcal{X}_c|}{c+1}\mathcal{E}_c(f,f).$$

Hence

$$\mathcal{E}_{c+1}(\tilde{f},\tilde{f}) \leq \frac{|\mathcal{X}_c|}{|\mathcal{X}_{c+1}|}\left(1+\frac{2(n-c)}{c+1}\right)\mathcal{E}_c(f,f).$$

Now, Lemmma 2.2.12 yields the claimed inequalities (4.1.1) and (4.1.2). We have proved the following result.

Theorem 4.1.2 *Assume that* $\mathbf{a} = (a_i)_1^n$, *b and* ℓ *are such that* $a_i, b \geq 0$ *and* $\sum_{i\in I} a_i \leq b$ *for all* $I \subset \{1,\dots,n\}$ *satisfying* $\#I \leq n - n^{1/2}$. *Then the chain* $K_{\mathbf{a},b}$ *on*

$$\mathcal{X}(\mathbf{a},b) = \left\{x = (x_i)_1^n \in \{0,1\}^n : \sum_i a_i x_i \leq b\right\}$$

satisfies

$$\lambda(\mathbf{a},b) \geq \frac{2\epsilon}{n}, \quad \alpha(\mathbf{a},b) \geq \frac{\epsilon}{n} \quad \textit{with} \quad \epsilon = e^{-4}$$

The associated semigroup $H_t = H_{\mathbf{a},b,t} = e^{-t(I-K_{\mathbf{a},b})}$ *satisfies*

$$\|h_t^x - 1\|_2 \leq e^{1-c} \textit{ for } t = (4\epsilon)^{-1} n\,(\log n\ + 2c)\,.$$

These are good estimates and I believe it would be difficult to prove similar bounds for $\|h_t^x - 1\|_2$ without using the notion of log-Sobolev constant (coupling is a possible candidate but if it works, it would only give a bound in ℓ^1).

In the case where $a_i = 1$ for all i and $b \geq n/2$, we can use the test function $f(x) = \sum_{i<n/2}(x_i - 1/2) - \sum_{i>n/2}(x_i - 1/2)$ to bound $\lambda(\mathbf{1},b)$ and $\alpha(\mathbf{1},b)$ from above. Indeed, this function satisfies $\pi_{\mathbf{1},b}(f) = \pi_{\mathcal{Z}}(f) = 0$ (use the symmetry that switches $i < n/2$ and $i > n/2$) and $\mathrm{Var}_{\mathbf{1},b}(f,f) \geq 2\frac{|\mathcal{Z}|}{|\mathcal{X}(\mathbf{a},b)|}\mathrm{Var}_{\mathcal{Z}}(f,f)$ (use the symmetry $x \to x + \mathbf{1} \bmod (2)$). Also $\mathcal{E}_{\mathbf{a},b} \leq \frac{|\mathcal{X}(\mathbf{a},b)|}{|\mathcal{Z}|}\mathcal{E}_{\mathcal{Z}}$. Hence $\lambda(\mathbf{a},b) \leq 4/n$, $\alpha(\mathbf{a},b) \leq 2/n$ in this particular case.

4.2 Comparison of Dirichlet forms using paths

The path technique of Section 3.1 can be used to compare two Dirichlet forms on a same state space $\mathcal{X}$. Together with Theorem 4.1.1 this provides a powerful tool to study finite Markov chains that are not too different from a given well-known chain. The results presented below can be seen as extentions of Theorems 3.2.1, 3.2.5. Indeed, what has been done in these theorems is nothing else than comparing the chain (K,π) of interest to the "trivial" chain with kernel $K'(x,y) = \pi(y)$ which has the same stationary distribution π. This chain K' has Dirichlet form

$\mathcal{E}'(f,f) = \mathrm{Var}_\pi(f)$ and is indeed well-known: It has eigenvalue 1 with multiplicity 1 and all the other eigenvalues vanish. Its log-Sobolev constant is given in Theorem 2.2.9. Once the Theorems of Section 3.2 have been interpreted in this manner their generalization presented below is straight-forward.

We will use the following notation. Let (K,π) be the unknown chain of interest and

$$Q(e) = \frac{1}{2}\left(K(x,y)\pi(x) + K(y,x)\pi(y)\right) \ \text{ if } \ e = (x,y).$$

Let $\mathcal{A}$ be an adapted edge-set according to Definition 3.1.1 and let

$$\Gamma = \bigcup_{x,y} \Gamma(x,y)$$

where $\Gamma(x,y)$ be the set of all paths from x to y that have no repeated edges.

Theorem 4.2.1 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge-set for K. Let (K',π') be an auxilliary chain. For each $(x,y) \in \mathcal{X}\times\mathcal{X}$ such that $x \neq y$ and $K'(x,y) > 0$ choose exactly one path $\gamma(x,y)$ in $\Gamma(x,y)$. Then $\mathcal{E}' \leq A\mathcal{E}$ where*

$$A = \max_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y\in\mathcal{X}:\\ \gamma(x,y)\ni e}} |\gamma(x,y)| K'(x,y)\pi'(x) \right\}.$$

PROOF: For each $(x,y) \in \mathcal{X}\times\mathcal{X}$ such that $K'(x,y) > 0$, write

$$f(y) - f(x) = \sum_{e\in\gamma(x,y)} df(e)$$

and, using Cauchy-Schwarz,

$$|f(y) - f(x)|^2 \leq |\gamma(x,y)| \sum_{e\in\gamma(x,y)} |df(e)|^2.$$

Multiply by $\frac{1}{2}K'(x,y)\pi'(x)$ and sum over all x,y to obtain

$$\frac{1}{2}\sum_{x,y} |f(y) - f(x)|^2 K'(x,y)\pi'(x) \leq \frac{1}{2}\sum_{x,y} |\gamma(x,y)| \sum_{e\in\gamma(x,y)} |df(e)|^2 K'(x,y)\pi(x).$$

The left-hand side is equal to $\mathcal{E}'(f,f)$ whereas the right-hand side becomes

$$\frac{1}{2}\sum_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y:\\ \gamma(x,y)\ni e}} |\gamma(x,y)K'(x,y)|\pi'(x) \right\} |df(e)|^2 Q(e)$$

which is bounded by

$$\max_{e\in\mathcal{A}}\left\{\frac{1}{Q(e)}\sum_{\substack{x,y:\\ \gamma(x,y)\ni e}}|\gamma(x,y)|K'(x,y)\pi'(x)\right\}\mathcal{E}(f,f).$$

Hence

$$\forall f,\quad \mathcal{E}(f,f)\le A\mathcal{E}(f,f)$$

with A as in Theorem 4.2.1.

Theorems 4.1.1, 4.2.1 are helpful for two reasons. First, non-trivial informations about K' can be brought to bear in the study of K. Second, the path combinatorics that is involved in Theorem 4.2.1 is often simpler than that involved in Theorem 3.2.1 because only the pairs (x,y) such that $K'(x,y)>0$ enter in the bound. These two points are illustrated by the next example.

EXAMPLE 4.2.1: Let $\mathcal{X}=\{0,1\}^n$. Let $x\to\tau(x)$, be defined by $[\tau(x)]_i=x_{i-1}$, $1<i\le n$, $[\tau(x)]_1=x_n$. Let $x\to\sigma(x)$ be defined by $\sigma(x)=x+(1,0,\dots,0)$. Set $K(x,y)=1/n$ if either $y=\tau^j(x)$ for some $1\le j\le n$ or $y=\sigma(x)$, and $K(x,y)=0$ otherwise. This chain is reversible with respect to the uniform distribution. In Section 3.2, we have seen that $\lambda\ge 1/n^3$ by Theorem 3.2.1. Here, we compare K with the chain $K'(x,y)=1/n$ if $|x-y|=1$ and $K(x,y)=0$ otherwise. For (x,y) with $|x-y|=1$, let i be such that $x_i\ne y_i$. Let

$$\gamma(x,y)=(x,\tau^j(x),\sigma\circ\tau^j(x),\tau^{-j}\circ\sigma\circ\tau^j(x)=y)$$

where $j=i$ if $i\le n/2$ and $j=n-i$ if $i>n/2$. These paths have length 3. The constant A of Theorem 4.2.1 becomes

$$A=3\max_{e\in\mathcal{A}}\#\left\{(x,y):K'(x,y)>0,\ \gamma(x,y)\ni e\right\}.$$

If $e=(u,v)$ with $v=\tau^j(u)$, there are only two (x,y) such that $e\in\gamma(x,y)$ depending on whether σ appears after or before e. If $v=\sigma(u)$, there are n possibilities depending on the choice of $j\in\{0,1,\dots,n-1\}$. Hence $A=3n$. Since $\lambda'=2/n$ and $\alpha'=1/n$, this yields

$$\lambda\ge\frac{2}{3n^2},\qquad \alpha\ge\frac{1}{3n^2}.$$

Also it follows that

$$\max_x\|h_t^x-1\|_2\le e^{1-c}\quad\text{for}\quad t=\frac{3n^2}{4}(2c+\log n),\ \ c>0.$$

EXAMPLE 4.2.2: Consider a graph $(\mathcal{X},\mathcal{A})$ where $\mathcal{A}$ is a symmetric set of oriented edges. Set $d(x)=\#\{y\in\mathcal{X}:(x,y)\in\mathcal{A}\}$ and

$$K(x,y)=\begin{cases}0 & \text{if } (x,y)\notin\mathcal{A}\\ 1/d(x) & \text{if } (x,y)\in\mathcal{A}.\end{cases}$$

This is the kernel of the simple random walk on $(\mathcal{X}, \mathcal{A})$. It is reversible with respect to the measure $\pi(x) = d(x)/|\mathcal{A}|$. For each $(x,y) \in \mathcal{X}^2$ choose a path $\gamma(x,y)$ with no repeated edges. Set

$$d_* = \max_{x\in\mathcal{X}} d(x), \quad \gamma_* = \max_{x,y\in\mathcal{X}} |\gamma(x,y)|, \quad \eta_* = \max_{e\in\mathcal{A}} \#\{(x,y) \in \mathcal{X}^2 : \gamma(x,y) \ni e\}.$$

We now compare with the chain $K'(x,y) = 1/|\mathcal{X}|$ which has reversible measure $\pi'(x) = 1/|\mathcal{X}|$ and spectral gap $\lambda' = 1$. Theorem 4.2.1 gives $\lambda \geq a/A$ with

$$A \leq \frac{|\mathcal{A}|\gamma_*\eta_*}{|\mathcal{X}|^2} \quad \text{and} \quad a = \frac{|\mathcal{A}|}{d_*|\mathcal{X}|}.$$

This gives

Theorem 4.2.2 *For the simple random walk on a graph $(\mathcal{X}, \mathcal{A})$ the spectral gap is bounded by*

$$\lambda \geq \frac{|\mathcal{X}|}{d_*\gamma_*\eta_*}.$$

Compare with Example 3.2.4 where we used Theorem 3.2.1 instead. The present result is slightly better than the bound obtained there. It is curious that one obtains a better bound by comparing with the chain $K'(x,y) = 1/|\mathcal{X}|$ as above than by comparing with the $\widetilde{K}(x,y) = \pi(y)$ which corresponds to Theorem 3.2.1.

It is a good exercise to specialize Theorem 4.2.1 to the case of two left invariant Markov chains $K(x,y) = q(x^{-1}y)$, $K'(x,y) = q'(x^{-1}y)$ on a finite group G. To take advantage of the group invariance, write any element g of G as a product

$$g = g_1^{\epsilon_1} \cdots g_k^{\epsilon_k}$$

with $q(g_i) + q(g_i^{-1}) > 0$. View this as a path $\gamma(g)$ from the identity id of G to g. Then for each (x,y) with $q'(x^{-1}y) > 0$, write

$$x^{-1}y = g(x,y) = g_1^{\epsilon_1} \cdots g_k^{\epsilon_k}$$

(where the g_i and ϵ_i depend on (x,y)) and define

$$\gamma(x,y) = x\gamma(g) = (x, xg_1, \ldots, xg_1 \ldots g_{k-1}, xg(x,y) = y).$$

With this choice of paths Theorem 4.2.1 yields

Theorem 4.2.3 *Let K, K' be two invariant Markov chains on a group G. Set $q(g) = K(\mathrm{id}, g)$, $q'(g) = K'(\mathrm{id}, g)$. Let π denote the uniform distribution. Fix a generating set S satisfying $S = S^{-1}$ and such that $q(s) + q(s^{-1}) > 0$. for all $s \in S$. For each $g \in G$ such that $q'(g) > 0$, choose a writing of g as a product of elements of S, $g = s_1 \ldots s_k$ and set $|g| = k$. Let $N(s,g)$ be the number of times $s \in S$ is used in the chosen writing of g. Then $\mathcal{E} \leq A\mathcal{E}'$ and $\lambda \geq \lambda'/A$ with*

$$A = \max_{s\in S} \left\{ \frac{2}{q(s) + q(s^{-1})} \sum_{g\in G} |g| N(s,g) q'(g) \right\}.$$

Assume further that K, K' are reversible and let λ_i (resp. λ'_i), $i = 0, \ldots, |G|-1$ denote the eigenvalues of $I - K$ (resp. $I - K'$) in non-decreasing order. Then $\lambda_i \geq \lambda'_i/A$ for all $i \in \{1, \ldots, |G|-1\}$ and

$$\forall\, x \in G, \quad \|h_t^x - 1\|_2 \leq \|h'^x_{t/A} - 1\|_2.$$

PROOF: (cf. [23], pg 702) We use Theorem 4.2.1 with the paths described above. Fix an edge $e = (z, w)$ with $w = zs$. Observe that there is a bijection between

$$\{(g, h) \in G \times G : \gamma(g, h) \ni (z, w)\}$$

and

$$\{(g, u) \in G \times G : \exists\, i \ \text{ such that } s_i(u) = s, z = gs_1(u) \cdots s_{i-1}(u)\}$$

given by $(g, h) \to (g, g^{-1}h) = (g, u)$. For each fixed $u = g^{-1}h$, there are exactly $N(s, u)$ $g \in G$ such that (g, u) belongs to

$$\{(x, u) \in G \times G : \exists\, i \ \text{ such that } s_i(u) = s, z = xxs_1(u) \cdots s_{i-1}(u)\}.$$

Hence

$$\sum_{(g,h) \in G \times G : \gamma(g,h) \ni (z,w)} |\gamma(g, h)| = \sum_{u \in G} |u| N(s, u).$$

This proves the desired result. See also [24] for a more direct argument.

We now extend Theorem 4.2.1 to allow the use of a set of paths for each pair (x, y) with $K'(x, y) > 0$.

Definition 4.2.4 *Let (K, π), K', π' be two irreducible Markov chains on a same finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge-set for (K, π). A (K, K')-flow is non-negative function $\phi : \Gamma(K') \to [0, \infty[$ on the path set*

$$\Gamma(K') = \bigcup_{\substack{x,y: \\ K'(x,y) > 0}} \Gamma(x, y)$$

such that

$$\forall x, y \in \mathcal{X},\ x \neq y,\ K'(x, y) > 0, \qquad \sum_{\gamma \in \Gamma(x,y)} \phi(\gamma) = K'(x, y)\pi'(x).$$

Theorem 4.2.5 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge-set for (K, π). Let (K', π') be a second chain and ϕ be a (K, K')-flow. Then $\mathcal{E}' \leq A(\phi)\mathcal{E}$ where*

$$A(\phi) = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{\gamma \in \Gamma(K'): \\ \gamma \ni e}} |\gamma| \phi(\gamma) \right\}.$$

PROOF: For each (x,y) such that $K'(x,y)>0$ and each $\gamma\in\Gamma(x,y)$ write

$$|f(y)-f(x)|^2 \le |\gamma|\sum_{e\in\gamma}|df(e)|^2.$$

Then

$$|f(y)-f(x)|^2K'(x,y)\pi'(x) \le \sum_{\gamma\in\Gamma(x,y)}|\gamma|\sum_{e\in\gamma}|df(e)|^2\phi(\gamma).$$

¿From here, complete the proof as for Theorem 4.2.1.

Corollary 4.2.6 *Assume that there is a group G which acts on $\mathcal{X}$ and such that*

$$\pi(gx)=\pi(x),\ \ \pi'(gx)=\pi'(x),\ \ Q(gx,gy)=Q(x,y),\ \ Q'(gx,gy)=Q'(x,y).$$

Let $\mathcal{A}$ be an adapted edge-set for (K,π) such that $(x,y)\in\mathcal{A}\Rightarrow(gx,gy)\in\mathcal{A}$. Let $\mathcal{A}=\bigcup_1^k\mathcal{A}_i$, be the partition of $\mathcal{A}$ into transitive classes for this action. Then $\mathcal{E}'\le A\mathcal{E}$ where

$$A=\max_{1\le i\le k}\left\{\frac{1}{|\mathcal{A}_i|Q_i}\sum_{x,y}N_i(x,y)d_K(x,y)K'(x,y)\pi(x)\right\}.$$

Here $|\mathcal{A}_i|=\#\mathcal{A}_i$, $Q_i=Q(e_i)$ with $e_i\in\mathcal{A}_i$, $d_K(x,y)$ is the distance between x and y in $(\mathcal{X},\mathcal{A})$, and $N_i(x,y)$ is the maximum number of edges of type i in a geodesic path from x to y.

PROOF: Consider the set $\mathcal{G}(x,y)$ of all geodesic paths from x to y. Define a (K,K')-flow ϕ by setting

$$\phi(\gamma)=\begin{cases} K'(x,y)\pi'(x)/\#\mathcal{G}(x,y) & \text{if } \gamma\in\mathcal{G}(x,y)\\ 0 & \text{otherwise.}\end{cases}$$

Then $A(\phi)=\max_e A(\phi,e)$ where

$$A(\phi,e)=\frac{1}{Q(e)}\sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}}|\gamma|\phi(\gamma).$$

By hypothesis, $A(\phi,e_i)=A_i(\phi)$ does not depend on $e_i\in\mathcal{A}_i$. Indeed, if $g\gamma$ denote the image of the path γ under the action of $g\in G$, we have $|g\gamma|=|\gamma|$, $\phi(g\gamma)=\phi(\gamma)$. Summing for each $i=1,\dots,k$ over all edges in $\mathcal{A}_i$, we obtain

$$\begin{aligned}
A(\phi,e_i) &= \frac{1}{|\mathcal{A}_i|Q_i}\sum_{e\in\mathcal{A}_i}\sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}}|\gamma|\phi(\gamma)\\
&= \frac{1}{|\mathcal{A}_i|Q_i}\sum_{e\in\mathcal{A}_i}\sum_{x,y}\sum_{\substack{\gamma\in\mathcal{G}(x,y):\\ \gamma\ni e}}\frac{d(x,y)K'(x,y)\pi'(x)}{\#\mathcal{G}(x,y)}\\
&\le \frac{1}{|\mathcal{A}_i|Q_i}\sum_{x,y}N_i(x,y)d(x,y)K'(x,y)\pi'(x).
\end{aligned}$$

This proves the desired bound.

EXAMPLE 4.2.3: Let $\mathcal{X}$ be the set of all the n-sets of $\{0, 1, \ldots, 2n-1\}$. On this set, consider two chains. The unknown chain of interest is the chain K of Example 3.2.8:

$$K(x,y) = \begin{cases} 1/n & \text{if } \#(x \cap y) = n-2 \text{ and } 0 \in x \oplus y \\ 0 & \text{otherwise} \end{cases}$$

This is a reversible chain with respect to the uniform distribution $\pi \equiv \binom{2n}{n}^{-1}$. Let $\mathcal{A}_K = \{e = (x,y) : K(x,y) \neq 0\}$ be the obvious K-adapted edge-set.

The better known chain K' that will be used for comparison is a special case of the chain considered of Example 3.2.7:

$$K'(x,y) = \begin{cases} 1/n^2 & \text{if } \#(x \cap y) = n-2 \\ 0 & \text{otherwise} \end{cases}$$

The chain K' is studied in detail in [34] using Fourier analysis on the Gelfand pair $(S_{2n}, S_n \times S_n)$. The eigenvalues are known to be the numbers

$$\frac{i(2n-i+1)}{n^2} \text{ with multiplicity } \binom{2n}{i} - \binom{2n}{i-1}, \quad 0 \leq i \leq n.$$

In particular, the spectral gap of K' is $\lambda' = 2/n$. This chain is known as the Bernoulli-Laplace diffusion model.

As in Example 3.2.8, the symmetric group S_{2n-1} which fixes 0 acts on $\mathcal{X}$ and preserves both chains K, K'. There are two classes $\mathcal{A}_1, \mathcal{A}_2$ of K-edges for this action: those edges (x,y), $x \oplus y = 2$, with $0 \in x \oplus y$ and those with $0 \notin x \oplus y$. Hence, we have $\mathcal{E}' \leq A\mathcal{E}$ with

$$A = \frac{2}{n^2 \binom{2n}{n}} \max_{i=1,2} \left\{ \sum_{\substack{x,y \\ x \oplus y = 2}} N_i(x,y) d_K(x,y) \right\}.$$

Now, if $x \oplus y = 2$ then

$$d_K(x,y) = \begin{cases} 1 & \text{if } 0 \in x \oplus y \\ 2 & \text{if } 0 \notin x \oplus y. \end{cases}$$

Moreover, in both cases, $N_i(x,y) = 0$ or 1. This yields

$$A \leq \frac{4}{n^2 \binom{2n}{n}} \sum_{\substack{x,y \\ x \oplus y = 2}} 1 = 4.$$

Thus

$$\mathcal{E}' \leq 4\mathcal{E}.$$

This shows that

$$\lambda \geq \frac{1}{2n}$$

improving upon the bound obtained in Example 3.2.8.

In their paper [34], Diaconis and Shahshahani actually show that

$$\|h_t'^x - 1\|_2 \le be^{-c} \quad \text{for} \quad t = \frac{1}{4}n(2c + \log n).$$

Using the comparison inequality $\mathcal{E}' \le 4\mathcal{E}$ and Theorem 4.1.1(2) we deduce from Diaconis and Shahshahani result that

$$|\!|\!| h_t - 1 |\!|\!| \le be^{-c} \quad \text{for} \quad t = n(2c + \log n).$$

Furthermore, the group S_{2n-1} fixing 0 acts with two transitive classes on $\mathcal{X}$. A vertex x is in one class or the other depending on whether or not x contains 0. The two classes have the same cardinality. Since $\|h_t^x - 1\|_2$ depends only of x through its class, we have

$$|\!|\!| h_t - 1 |\!|\!|^2 = \frac{1}{2}\left(\|h_t^{x_1} - 1\|_2^2 + \|h_t^{x_2} - 1\|_2^2\right)$$

where $x_1 \ni 0$ and $x_2 \not\ni 0$ are fixed elements representing their class. Hence, we also have

$$\max_x \|h_t^x - 1\|_2 \le 2be^{-c} \quad \text{for} \quad t = n(2c + \log n).$$

This example illustrates well the strength of the idea of comparison which allows a transfer of information from one example to another.

Bibliography

[1] Aldous D. (1983) *Random walks on finite groups and rapidly mixing Markov chains.* In Séminaire de Probabilités, XVII, LNM 986, Springer, Berlin.

[2] Aldous D. (1987) *On the Markov-chain simulation method for uniform combinatorial simulation and simulated annealing.* Prob. Eng. Info. Sci., 1, 33-46.

[3] Aldous D. and Fill J. (1996) Preliminary version of a book on finite Markov chains available via homepage http://www.stat.berkeley.edu/users/aldous

[4] Aldous D. and Diaconis P. (1987) *Strong uniform times and finite random walks.* Adv. Appl. Math. 8, 69-97.

[5] Alon N. (1986) *Eigenvalues and expanders.* Combinatorica 6, p.83-96.

[6] Alon N. and Milman V. (1985) λ_1, *isoperimetric inequalities for graphs and superconcentrators.* J. Comb. Th. B, 38, 78-88.

[7] Bakry D. and Emery M. (1985) *Diffusions hypercontractive.* Séminaire de probabilité XIX, Springer LNM 1123, 179-206.

[8] Bakry D. (1994) *L'hypercontractivité et son utilisation en théorie des semigroups.* In Ecole d'été de Saint Flour 1992, Springer LNM 1581.

[9] Bakry D., Coulhon T., Ledoux M., Saloff-Coste L. (1995) *Sobolev inequalities in disguise.* Indian Univ. Math. J., 44, 1043-1074.

[10] Bonami A. (1970) *Étude des coefficients de Fourier des fonctions de $L^p(G)$.* Ann. Inst. Fourier, 20, 335-402.

[11] Carlen E., Kusuoka S. and Stroock D. (1987) *Upper bounds for symmetric Markov transition functions.* Ann. Inst. H. Poincaré, Prob. Stat. 23, 245-287.

[12] Cheeger J. (1970) *A lower bound for the smallest eigenvalue of the Laplacian.* Problems in Analysis, Synposium in Honor of S. Bochner. Princeton University Press. 195-199.

[13] Chung F. and Yau S-T. (1994) *A harnack inequality for homogeneous graphs and subgraphs.* Communication in Analysis and Geometry, 2, 627-640.

[14] Chung F. and Yau S-T. (1995) *Eigenvalues of graphs and Sobolev inequalities.* Combinatorics, Probability and Computing, 4, 11-25.

[15] Davies E.B. (1989) *Heat kernels and spectral theory.* Cambridge University Press.

[16] Deuschel J-D. and Stroock D. (1989) *Large deviations.* Academic Press, Boston.

[17] Diaconis P. (1986) *Group representations in probability and statistics.* IMS, Hayward.

[18] Diaconis P. (1996) *The cutoff phenomenon in finite Markov chains.* Proc. Natl. Acad. Sci. USA, 93, 1659-1664.

[19] Diaconis P. and Fill J. (1990) *Strong stationary times via a new form of duality.* Ann. Prob. 18, 1483-1522.

[20] Diaconis P., Graham R. and Morrison J. (1990) *Asymptotic analysis of a random walk on a hypercube with many dimensions.* Random Structures and Algorithms, 1, 51-72.

[21] Diaconis P. and Gangolli A. (1995) *Rectangular arrays with fixed margins.* In Discrete Probability and Algorithms, (Aldous et al, ed.) 15-41. The IMA volumes in Mathematics and its Applications, Vol. 72, Springer-Verlag.

[22] Diaconis P. and Holmes S. (1995) *Three Examples of Monte-Carlo Markov Chains: at the Interface between Statistical Computing, Computer Science and Statistical Mechanics.* In Discrete Probability and Algorithms, (Aldous et al, ed.) 43-56. The IMA volumes in Mathematics and its Applications, Vol. 72, Springer-Verlag.

[23] Diaconis P. and Saloff-Coste L. (1993) *Comparison theorems for reversible Markov chains.* Ann. Appl. Prob. 3, 696-730.

[24] Diaconis P. and Saloff-Coste L. (1993) *Comparison techniques for random walk on finite groups.* Ann. Prob. 21, 2131-2156.

[25] Diaconis P. and Saloff-Coste L. (1994) *Moderate growth and random walk on finite groups.* G.A.F.A., 4, 1-36.

[26] Diaconis P. and Saloff-Coste L. (1995) *An application of Harnack inequalities to random walk on nilpotent quotients.*, J. Fourier Anal. Appl., Kahane special issue, 187-207.

[27] Diaconis P. and Saloff-Coste L. (1995) *Random walks on finite groups: a survey of analytical techniques.* In Probability on groups and related structures XI, H. Heyer (ed), World Scientific.

[28] Diaconis P. and Saloff-Coste L. (1996) *Nash inequalities for finite Markov chains.*, J. Th. Prob. 9, 459-510.

[29] Diaconis P. and Saloff-Coste L. (1996) *Logarithmic Sobolev inequalities and finite Markov chains.* Ann. Appl. Prob. 6, 695-750.

[30] Diaconis P. and Saloff-Coste L. (1995) *Walks on generating sets of Abelian groups.* Prob. Th. Rel. Fields. 105, 393-421.

[31] Diaconis P. and Saloff-Coste L. (1995) *What do we know about the Metropolis algorithm.* J.C.S.S. To appear.

[32] Diaconis D. and Shahshahani M. (1981) *Generating a random permutation with random transpositions.* Z. Wahrsch. Verw. Geb., 57, 159-179.

[33] Diaconis P. and Sahshahani M. (1987) *The subgroup algorithm for generating uniform random variables.* Probl. in Engin. Info. Sci., 1, 15-32.

[34] Diaconis P. and Shahshahani M. (1987) *Time to reach statinarity in the Bernoulli-Laplace diffusion model.* SIAM Jour. Math. Anal., 18, 208-218.

[35] Diaconis P. and Stroock D. (1991) *Geometric bounds for eigenvalues for Markov chains.* Ann. Appl. Prob. 1, 36-61.

[36] Dinwoodie I. H. (1995) *A probability inequality for the occupation measure of a reversible Markov chain.* Ann. Appl. Prob., 5, 37-43.

[37] Dinwoodie I. H. (1995) *Probability inequalities for the occupation measure of a Markov chain.*

[38] Dyer M., Frieze A., Kannan R., Kapoor A., Perkovic L., and Vazirani U. (1993) *A mildly exponential time algorithm for approximating the number of solutions to a multidimensional knapsack problem.* Combinatorics, Probability and Computing, 2, 271-284.

[39] Dyer M. and Frieze A. (1991) *Computing the volume of convex bodies: a case where randomness provably helps.* Probabilistic Combinatorics and its applications, Proceedings of the AMS Symposia in Applied Mathematics 44, 123-170.

[40] Feller W. (1968) *An introduction to probability theory and its applications.* Vol. I, third edition, John Wiley & Sons, New-York.

[41] Fill J. (1991) *Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with application to the exclusion process.* Ann. Appl. Probab., 1, 62-87.

[42] Flatto L., Odlyzko A. and Wales D. (1985) *Random shuffles and group representations.* Ann. Prob. 13, 151-178.

[43] Frieze A., Kannan R. and Polson N. (1994) *Sampling from log concave distributions.* Ann. Appl. Prob. 4, 812-837.

[44] Gillman D. (1993) *Hidden Markov chains: rates of convergences and the complexity of inference.* Ph.D. thesis, Massachusets Institute of Technology, Department of mathematics.

[45] Gluck D. (1996) *Random walk and character ratios on finite groups of Lie type.* Adv. Math.

[46] Gross L. (1976) *Logarithmic Sobolev inequalities.* Amer. J. Math. 97, 1061-1083.

[47] Gross L. (1993) *Logarithmic Sobolev inequalities and contractivity properties of semigroups.* In Lecture Notes in Math. 1563. Springer.

[48] Higuchi Y. and Yoshida N. (1995) *Analytic conditions and phase transition for Ising models.* Lecture notes in Japanese.

[49] Hildebrand M. (1992) *Genrating random elements in $SL_n(F_q)$ by random transvections.* J. Alg. Combinatorics, 1, 133-150.

[50] Holley R. and Stroock D. (1987) *Logarithmic Sobolev inequalities and stochastic Ising models.* J. Stat. Phys. 46, 1159-1194.

[51] Horn R. and Johnson Ch. (1985) *Matrix analysis.* Cambridge University Press.

[52] Horn R and Johnson Ch. (1991) *Topics in Matrix analysis.* Cambridge University Press.

[53] Jerrum M. and Sinclair A. (1993)*Polynomial time approximation algorithms for the Ising model,* SIAM Journal of Computing, 22, 1087-1116.

[54] Jerrum M. and Sinclair A. (1997) *The Markov chain Monte Carlo method: an approach to approximate counting and integration.* In Approximation algorithms for NP-hard problems, D.S. Hochbaum (Ed.), PWS Publishing, Boston.

[55] Kahale N. (1995) *A semidefinite bound for mixing rates of Markov chains.* To appear in Random Structures and Algorithms.

[56] Kannan R. (1994) *Markov chains and polynomial time algorithms.* Proceedings of the 35th IEEE Symposium on Foundations of Computer Science, Computer Society Press, 656-671.

[57] Kemeny J. and Snell L. (1960) *Finite Markov chains.* Van Nostrand company, Princeton.

[58] Lawler G. and Sokal A. (1988) *Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger inequality.* Trans. AMS, 309, 557-580.

[59] Lezaud P. (1996) *Chernoff-type bound for finite Markov chains.* Ph.D. Thesis in progress, Université Paul Sabatier, Toulouse.

[60] Lovász L. and Simonovits M. (1993). *Random walks in a convex body and an improved volume algorithm.* Random Structures and Algorithms, 4, 359-412.

[61] Mann B. (1996) *Berry-Essen central limit theorems for Markov chains.* Ph.D. Thesis, Harvard University, Department of Mathematics.

[62] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A. and Teller E. (1953) *Equations of state calculations by fast computing machines.* J. Chem. Phys. 21, 1087-1092.

[63] Miclo L. (1996) *Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies.* To appear in Séminaire de Probabilité XXXI. Lecture notes in Math. Springer.

[64] Nash J. (1958) *Continuity of solutions of parabolic and elliptic equations.* Amer. J. Math. 80, 931-954.

[65] Rothaus O. (1980) *Logarithmic Sobolev inequalities and the spectrum of Sturm-Liouville operators.* J. Funct. Anal., 39, 42-56.

[66] Rothaus O. (1981) *Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators.* J. Funct. Anal., 42 , 110-120.

[67] Rothaus O. (1981) *Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities.* J. Funct. Anal., 42 , 102-109.

[68] Rothaus O. (1985) *Analytic inequalities, Isoperimetric inequalities and logarithmic Sobolev inequalities.* J. Funct. Anal., 64, 296-313.

[69] Saloff-Coste L. (1996) *Simple examples of the use Nash inequalities for finite Markov chains.* In Semstat III, Current Trends in Stochastic Geometry and its Applications. W.S. Kendall, O.E. Barndorff-Nielsen and MC. van Lieshout, Eds. Chapman & Hall.

[70] Senata E. (1981) *Non negative matrices and Markov chains* (2nd ed.) Springer.

[71] Sinclair A. (1992) *Improved bounds for mixing rates of Markov chains and multicommodity flow.* Combinatorics, Probability and Computing, 1, 351-370.

[72] Sinclair A. (1993) *Algorithms for random generation and counting: a Markov chain approach.* Birkhäuser, Boston.

[73] Stein E. and Weiss G. (1971) *Introduction to Fourier analysis in Euclidean spaces.* Princeton Univ. Press, Princeton.

[74] Stong R. (1995) *Random walks on the groups of upper triangular matrices* Ann. Prob., 23, 1939-1949.

[75] Stong R. (1995) *Eigenvalues of the natural random walk on the Burnside group* $B(3,n)$. Ann. Prob., 23, 1950-1960.

[76] Stong R. (1995) *Eigenvalues of random walks on groups.* Ann. Prob., 23, 1961-1981.

[77] Swendsen R. H. and Wang J-S. (1987) *Nonuniversal critical dynamics in Monte-Carlo simulations,* Physical review letters, 58, 86-88.

[78] Varopoulos N. (1985) *Semigroupes d'opérateurs sur les espaces* L^p. C. R. Acad. Sc. Paris. 301, Série I, 865-868.

[79] Varopoulos N. (1985) *Théorie du potentiel sur les groupes et les variétés.* C. R. Acad. Sc. Paris. 302, Série I, 203-205.

LISTE DES EXPOSES

– Melle AKIAN Marianne
Des processus stochastiques au contrôle optimal : algèbre ou grandes déviations ?
– Mr BARDINA Xavier
Une extension de la formule d'Itô pour les diffusions elliptiques
– Mr BEN HARIZ Samir
T.L.C. pour le temps local d'un processus gaussien stationnaire
– Mr BERTRAND Pierre
Estimation du coefficient de diffusion d'un processus quand ce coefficient est stochastique. Détection des sauts du coefficient de diffusion
– Mr BIENVENUE Alexis
Marches aléatoires corrélées : principe d'invariance ; récurrence
– Melle BUTUCEA Cristina
Estimation adaptative de la densité sur la classe de Sobolev-constante exacte
– Mr CARMONA Philippe
Vitesse moyenne d'un mouvement brownien dans un potentiel aléatoire de type Lévy
– Mr CEROU Frédéric
Comportement en temps long pour des problèmes de filtrage sans bruit de dynamique
– Melle COT Cécile
Vitesse de convergence du recuit simulé associé à une suite de températures décroissantes par paliers
– Mr ERMAKOV Alexei
Non-dissipation of masses
– Mme FLORIT Carmen
Diffusion approximation for hyperbolic stochastic equations
– Mr FONTENAS Eric
Constantes de Sobolev pour les opérateurs de diffusion
– Mr FRANZ Uwe
Loi de Gauss et théorème de Bernstein sur les groupes quantiques
– Mr GOBRON Thierry
Fonction de partition sur l'ensemble des courbes polygonales fermées d'un graphe
– Melle GUILLOU Armelle
Construction d'intervalles de confiance de la moyenne par la méthode du bootstrap pondéré
– Mr HEINRICH Philippe
Loi zéro-un pour les polynômes d'un vecteur gaussien centré
– Mr HOFFMANN Marc
Estimation non paramétrique du coefficient de diffusion
– Mr JUREK Zbigniew J.
Infinite divisibility revisited
– Mr LEURIDAN Christophe
Le théorème de Ray-Knight à temps fixe
– Melle LOUHICHI Sarra
Moments inequalities under weak dependence
– Mme LUDENA Carenne
Intégrales du logarithme du périodogramme
– Mr MARDIN Arif
Agrégation du modèle de Hopfield-little

- Mr MARQUEZ-CARRERAS David, Université de Barcelone
  Petites perturbations d'une EDPS hyperbolique
- Mr MARQUEZ David, Université de Marne-la-Vallée
  Sur la vitesse de convergence de la diffusion du recuit simulé
- Mr MAZET Olivier
  Classification des semi-groupes de diffusion sur $\mathbb{R}$ associés à une famille de polynômes orthogonaux
- Mr MICLO Laurent
  Remarques sur l'évolution de l'entropie des chaînes de Markov
- Mme PAN Yiyan
  Moyenne harmonique du spectre de processus arborescents
- Mr RIOS Ricardo
  Towards a nonparametric test of linearity for times series
- Melle ROUSSEL Sandrine
  Marche aléatoire engendrée par les 3 cycles
- Mr SAINT LOUBERT BIE Erwan
  Etude d'une E.D.P.S. parabolique conduite par un processus de Poisson
- Mr SCHOTT René
  Processus de diffusion et analyse d'algorithmes
- Mr SHI Zhan et YOR Marc
  Espérance et variance du mouvement brownien
- Mr TINDEL Samy
  EDPS elliptiques réfléchies
- Mr VAN CASTEREN Jan A.
  On differences of self-adjoint semigroups
- Melle VAN MOFFAERT Amélie
  Study of extended dynamical systems : coupled lattice maps
- Mr WUTHRICH Mario Valentin
  Fluctuation results for brownian motion in a poissonian potential
- Mr YCART Bernard
  Modèle de transfert de charge en architecture parallèle
- Mr ZERNER Martin P.W.
  Shape theorem, Lyapounov exponents, and large deviations for random walk in a random potential

LISTE DES AUDITEURS

Melle	AKIAN Marianne	INRIA, Domaine de Voluceau, ROCQUENCOURT
Melle	ALOS Elisa	Faculté de Mathématiques, Université de BARCELONE (Espagne)
Mr	AMGHIBECH Said	Analyse, Modèles Stochastiques, Université de ROUEN
Mr	ANDJEL Enrique	Mathématiques et Informatique, Université de Provence, AIX-MARSEILLE
Mr	AZEMA Jacques	Laboratoire de Probabilités, Université PARIS VI
Mr	BALDI Paolo	Département de Mathématiques, Université de ROME (Italie)
Mr	BARDINA Xavier	Département de Mathématiques, Université de BARCELONE (Espagne)
Mr	BEN HARIZ Samir	Mathématiques, Université Paris-Sud, ORSAY
Mr	BENASSI Albert	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	BERNARD Pierre	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	BERTRAND Pierre	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	BIENVENUE Alexis	Laboratoire de Probabilités, Université Claude Bernard, LYON I
Mr	BRUNAUD Marc	U.F.R. de Mathématiques, Université PARIS VII
Melle	BUTUCEA Cristina	CREST, Laboratoire de Statistiques, MALAKOFF
Mme	CARAMELLINO Lucia	Département de Mathématiques, Université de ROME (Italie)
Mr	CARMONA Philippe	Laboratoire de Statistique et Probabilités, Université Paul Sabatier, TOULOUSE
Mr	CEROU Frédéric	Mathématiques et Informatique, Université de Provence AIX-MARSEILLE
Mr	CHABOT Nicolas	Mathématiques et Informatique, Université de Provence AIX-MARSEILLE
Melle	COT Cécile	Ecole Normale Supérieure, Rue d'Ulm, PARIS
Mr	COULHON Thierry	Département de Mathématiques, Université de CERGY-PONTOISE
Mr	DOUKHAN Paul	Département d'Economie, Université de CERGY-PONTOISE
Mr	DUNLOP François	Département de Physique, Université de CERGY-PONTOISE
Mr	EL KADIRI Mohamed	Département de Mathématiques, Université de RABAT (Maroc)
Mr	ERMAKOV Alexei	Centrum voor Wiskunde en Informatica, AMSTERDAM (Pays-Bas)
Mr	FLEURY Gérard	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	FLORESCU Mihai	Mathématiques, Université Paris-Sud, ORSAY
Mme	FLORIT Carme	Faculté de Mathématiques, Université de BARCELONE (Espagne)
Mr	FONTENAS Eric	Laboratoire de Statistique et Probabilités, Université Paul Sabatier, TOULOUSE
Mr	FRANZ Uwe	Département de Mathématiques, Institut Elie Cartan, NANCY I
Mr	GALLARDO Léonard	Département de Mathématiques, Université de TOURS
Mr	GAMBURD Alexandre	Department of Mathematics, Princeton University, PRINCETON (USA)
Mr	GOBRON Thierry	Ecole Polytechnique, PALAISEAU
Me	GRADINARU Mihai	Département de Mathématiques, Université d'EVRY
Melle	GUILLOU Armelle	CEA - CE/Saclay, SERMA/LEPP, GIF SUR YVETTE
Mr	GUIOL Hervé	Mathématiques et Informatique, Université de Provence AIX-MARSEILLE
Mr	HEINRICH Philippe	Département de Mathématiques, Université Louis Pasteur STRASBOURG
Mr	HOFFMANN Marc	U.F.R. de Mathématiques, Université PARIS VII
Mr	JUREK Z.J.	Institut de Mathématiques, Université de WROCLAW (Pologne)
Mr	KERKYACHARIAN Gérard	Département de Mathématiques, Université d'AMIENS

Mr	LANJRIZAIDI Boureddine	Mathématiques, Université de BARCELONE (Espagne)
Mr	LAPEYRE Bernard	ENPC - CERMICS, La Courtine
Mr	LEURIDAN Christophe	Institut Fourier, Université de GRENOBLE
Melle	LOUHICHI Sarra	Mathématiques, Université Paris-Sud, ORSAY
Mme	LUDENA Carenne	Mathématiques, Université Paris-Sud (ORSAY)
Mr	MARDIN Arif	Département "Signal et Image", Institut National des Télécommunications, EVRY
Mr	MARQUEZ David	Département de Mathématiques, Université de BARCELONE (Espagne)
Mr	MARQUEZ David	Analyse et Mathématiques Appliquées Université de Marne-la-Vallée, NOISY LE GRAND
Mr	MAZET Olivier	Laboratoire de Statistique et Probablités, Université Paul Sabatier, TOULOUSE
Mr	MICLO Laurent	Laboratoire de Statistique et Probabilités, Université Paul Sabatier, TOULOUSE
Mme	PAN Yiyan	LMC, IMAG, GRENOBLE
Melle	PHILIPPE Anne	U.F.R. de Mathématiques, Université de ROUEN
Mme	PICARD Dominique	U.F.R. de Mathématiques, Université PARIS VII
Mr	RACHAD Abdelhak	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	RAMPONI Alessandro	Département de Mathématiques, Université de Rome, ROME (Italie)
Melle	REMY Elisabeth	Mathématiques et Informatique, Université de Provence, AIX-MARSEILLE
Mr	RIO Emmanuel	Mathématiques, Université Paris-Sud (ORSAY)
Mr	RIOS Ricardo	Mathématiques, Université Paris-Sud (ORSAY)
Melle	ROUSSEL Sandrine	Laboratoire de Statistique et Probabilités, Université Paul Sabatier, TOULOUSE
Mr	ROUX Daniel	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	ROVIRA Carles	Faculté de Mathématiques, Université de BARCELONE (Espagne)
Melle	SAADA Ellen	URA CNRS 1378, Mathématiques, Université de ROUEN
Mr	SAINT LOUBERT BIE Erwan	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD
Mr	SAMSON Paul-Marie	Laboratoire de Statistique et Probabilités, Université Paul Sabatier, TOULOUSE
Melle	SARRA Monica	Faculté de Mathématiques, Université de BARCELONE (Espagne)
Mr	SCHILTZ Jang	Département de Mathématiques, Université de METZ
Mr	SETHURAMAN Sunder	Institut de Mathématiques, ETH-Zentrum, ZURICH (Suisse)
Melle	SHAO Ming	Département de Mathématiques, Université de WILRIJK (Belgique)
Mr	SHI Zhan	L.S.T.A., Université de PARIS VI
Mr	SOLE I CLIVILLES Josep	Faculté de Mathématiques, Université de BARCELONE (Espagne)
Mme	TIBI Danielle	Département de Mathématiques, Université PARIS VII
Mr	TINDEL Samy	Département de Mathématiques, Université de BARCELONE (Espagne)
Mr	TRAN HUNG THAO	Institut de Mathématiques, HANOI (Vietnam)
Mr	UTZET Frédéric	Département de Mathématiques, Université de BARCELONE (Espagne)
Mr	VAN CASTEREN Jan	Département de Mathématiques, Université de LOUVAIN (Belgique)
Melle	VAN MOFFAERT Amélie	Institut de Physique Théorique, Université de LOUVAIN (Belgique)
Mr	VIENS Frederi	Département de Mathématiques, Université de Californie, IRVINE (USA)
Mr	WERNER Wendelin	Ecole Normale Supérieure, rue d'Ulm à PARIS
Mr	WU Liming	Maths Appliquées, Université Blaise Pascal, CLERMONT-FD

Mr	WUTHRICH Mario Valentin	Departement de Mathématiques, ETH Zentrum, ZURICH (Suisse)
Mr	YCART Bernard	LMC/IMAG, GRENOBLE
Mr	ZERNER Martin P.W.	Département de Mathématiques, ETH Zentrum, ZURICH (Suisse)

LIST OF PREVIOUS VOLUMES OF THE "Ecole d'Eté de Probabilités"

1971 - J.L. Bretagnolle (LNM 307)
"Processus à accroissements indépendants"
S.D. Chatterji
"Les martingales et leurs applications analytiques"
P.A. MEYER
"Présentation des processus de Markov"

1973 - P.A. MEYER (LNM 390)
"Transformation des processus de Markov"
P. PRIOURET
"Processus de diffusion et équations différentielles stochastiques"
F. SPITZER
"Introduction aux processus de Markov à paramètres dans Z_v"

1974 - X. FERNIQUE (LNM 480)
"Régularité des trajectoires des fonctions aléatoires gaussiennes"
J.P. CONZE
"Systèmes topologiques et métriques en théorie ergodique"
J. GANI
"Processus stochastiques de population"

1975 - A. BADRIKIAN (LNM 539)
"Prolégomènes au calcul des probabilités dans les Banach"
J.F.C. KINGMAN
"Subadditive processes"
J. KUELBS
"The law of the iterated logarithm and related strong convergence theorems for Banach space valued random variables"

1976 - J. HOFFMANN-JORGENSEN (LNM 598)
"Probability in Banach space"
T.M. LIGGETT
"The stochastic evolution of infinite systems of interacting particles"
J. NEVEU
"Processus ponctuels"

1977 -	D. DACUNHA-CASTELLE	(LNM 678)
	"Vitesse de convergence pour certains problèmes statistiques"	
	H. HEYER	
	"Semi-groupes de convolution sur un groupe localement compact et applications à la théorie des probabilités"	
	B. ROYNETTE	
	"Marches aléatoires sur les groupes de Lie"	
1978 -	R. AZENCOTT	(LNM 774)
	"Grandes déviations et applications"	
	Y. GUIVARC'H	
	"Quelques propriétés asymptotiques des produits de matrices aléatoires"	
	R.F. GUNDY	
	"Inégalités pour martingales à un et deux indices : l'espace H^p"	
1979 -	J.P. BICKEL	(LNM 876)
	"Quelques aspects de la statistique robuste"	
	N. EL KAROUI	
	"Les aspects probabilistes du contrôle stochastique"	
	M. YOR	
	"Sur la théorie du filtrage"	
1980 -	J.M. BISMUT	(LNM 929)
	"Mécanique aléatoire"	
	L. GROSS	
	"Thermodynamics, statistical mechanics and random fields"	
	K. KRICKEBERG	
	"Processus ponctuels en statistique"	
1981 -	X. FERNIQUE	(LNM 976)
	"Régularité de fonctions aléatoires non gaussiennes"	
	P.W. MILLAR	
	"The minimax principle in asymptotic statistical theory"	
	D.W. STROOCK	
	"Some application of stochastic calculus to partial differential equations"	
	M. WEBER	
	"Analyse infinitésimale de fonctions aléatoires"	

1982 -	R.M. DUDLEY	(LNM 1097)
	"A course on empirical processes"	
	H. KUNITA	
	"Stochastic differential equations and stochastic flow of diffeomorphisms"	
	F. LEDRAPPIER	
	"Quelques propriétés des exposants caractéristiques"	
1983 -	D.J. ALDOUS	(LNM 1117)
	"Exchangeability and related topics"	
	I.A. IBRAGIMOV	
	"Théorèmes limites pour les marches aléatoires"	
	J. JACOD	
	"Théorèmes limites pour les processus"	
1984 -	R. CARMONA	(LNM 1180)
	"Random Schrödinger operators"	
	H. KESTEN	
	"Aspects of first passage percolation"	
	J.B. WALSH	
	"An introduction to stochastic partial differential equations"	
1985-87 -	S.R.S. VARADHAN	(LNM 1362)
	"Large deviations"	
	P. DIACONIS	
	"Applications of non-commutative Fourier analysis to probability theorems"	
	H. FOLLMER	
	"Random fields and diffusion processes"	
	G.C. PAPANICOLAOU	
	"Waves in one-dimensional random media"	
	D. ELWORTHY	
	"Geometric aspects of diffusions on manifolds"	
	E. NELSON	
	"Stochastic mechanics and random fields"	
1986 -	O.E. BARNDORFF-NIELSEN	(LNS 50)
	"Parametric statistical models and likelihood"	

1988 -	A. ANCONA	(LNM 1427)
	"Théorie du potentiel sur les graphes et les variétés"	
	D. GEMAN	
	"Random fields and inverse problems in imaging"	
	N. IKEDA	
	"Probabilistic methods in the study of asymptotics"	
1989 -	D.L. BURKHOLDER	(LNM 1464)
	"Explorations in martingale theory and its applications"	
	E. PARDOUX	
	"Filtrage non linéaire et équations aux dérivées partielles stochastiques associées"	
	A.S. SZNITMAN	
	"Topics in propagation of chaos"	
1990 -	M.I. FREIDLIN	(LNM 1527)
	"Semi-linear PDE's and limit theorems for large deviations"	
	J.F. LE GALL	
	"Some properties of planar Brownian motion"	
1991 -	D.A. DAWSON	(LNM 1541)
	"Measure-valued Markov processes"	
	B. MAISONNEUVE	
	"Processus de Markov : Naissance, Retournement, Régénération"	
	J. SPENCER	
	"Nine Lectures on Random Graphs"	
1992 -	D. BAKRY	(LNM 1581)
	"L'hypercontractivité et son utilisation en théorie des semigroupes"	
	R.D. GILL	
	"Lectures on Survival Analysis"	
	S.A. MOLCHANOV	
	"Lectures on the Random Media"	
1993 -	P. BIANE	(LNM 1608)
	"Calcul stochastique non-commutatif"	
	R. DURRETT	
	"Ten Lectures on Particle Systems"	

1994 - R. DOBRUSHIN (LNM 1648)
"Perturbation methods of the theory of Gibbsian fields"
P. GROENEBOOM
"Lectures on inverse problems"
M. LEDOUX
"Isoperimetry and gaussian analysis"

1996 - E. GINE (LNM 1665)
"Decoupling anf limit theorems for U-statistics and U-processes"
"Lectures on some aspects theory of the bootstrap"
G. GRIMMETT
"Percolation and disordered systems"
L. SALOFF-COSTE
"Lectures on finite Markov chains"