Use R!

Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R

Order-Restricted Analysis of Microarray Data

Springer

Use R!

Series Editors:
Robert Gentleman Kurt Hornik Giovanni Parmigiani

For further volumes:
http://www.springer.com/series/6991

Dan Lin • Ziv Shkedy • Daniel Yekutieli
Dhammika Amaratunga • Luc Bijnens
Editors

Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R

Order-Restricted Analysis of Microarray Data

Springer

Editors
Dan Lin
Veterinary Medicine Research
& Development, Pfizer Animal Health
Zaventem, Belgium

Ziv Shkedy
Interuniversity Institute for Biostatistics
and statistical Bioinformatics (I-BioStat)
Center for Statistics (CenStat)
Hasselt University, Diepenbeek, Belgium

Daniel Yekutieli
Department of Statistics and Operations
Research, School of Mathematical Sciences,
Tel-Aviv University, Tel-Aviv, Israel

Dhammika Amaratunga
Biostatistics and Programming
Janssen Pharmaceutical Companies
of Johnson & Johnson
Raritan, NJ USA

Luc Bijnens
Biostatistics and Programming
Janssen Pharmaceutical Companies
of Johnson & Johnson
Beerse, Belgium

R programs and datasets: http:/www.ibiostat.be/software; R packages on http://www.r-project.org; http://cran.r-project.org; http://www.bioconductor.org and/or via springer.com/statistics/book/978-3-642-24006-5

ISBN 978-3-642-24006-5 ISBN 978-3-642-24007-2 (eBook)
DOI 10.1007/978-3-642-24007-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012945536

Preface

Bioinformatics and statistical bioinformatics have developed rapidly over the last 15 years. In particular, the development of microarray technology introduced the challenge of the analysis of massive datasets and the need to consider inference when thousands of genes are tested.

Microarray experiments are slowly becoming an integrated part of the pharmaceutical research and development (R&D) process. Microarray experiments offer the ability to measure, at the same time, the RNA derived from entire genomes. Microarrays specific for animal species can be used to process the information coming from in vivo animal experiments. Microarrays specific for humans can be used to test biological material coming from biopsies, blood samples, or cell line cultures. The functional genomic information coming from microarray experiments can be used both at the target identification and the target validation level of the early drug discovery process. Moreover, functional genomics can be used at many later stages of pharmaceutical research and development to screen therapeutic effects and unwanted side effects in many R&D programs.

This book is about a specific setting in which gene expression is measured at different dose levels of a drug. The main goal of the analysis of dose-response microarray experiments is to detect trends in gene expression caused by increasing doses of compound. Therefore, this book is focused on estimation, inference, and clustering under order restrictions of dose-response microarray data. The aim of these microarray experiments is to get insight into the mechanism of action and the safety profile of a drug using functional genomic data to identify pathways that are affected by the compound at hand. In this context, gene expression experiments have become important either before or parallel to the clinical testing programs.

In this book, we present a toolbox for the analysis of dose-response microarray experiments. The toolbox consists of different statistical methods for the analysis and different R packages which were developed for the analysis. The web site accompanying this book contains all the R programs and datasets used to produce the output presented in the book. It can be reached through the web site of Hasselt University:

http://www.ibiostat.be/software/IsoGeneGUI/index.html

R packages can be downloaded from either the R Project web site, R-Forge, CRAN, or the Bioconductor web site: http://www.r-project.org/, http://cran.r-project.org/, http://r-forge.r-project.org/, and http://www.bioconductor.org/, respectively.

In the first part of the book, we introduce the main concepts of estimation and inference under order constraints and dose-response modeling. In the second part of the book, we focus on the analysis of dose-response microarray experiments and address issues such as multiplicity adjustment, selective inference, single and multiple contrast tests, order-restricted clustering, pathway analysis, hierarchical Bayesian models, and model-based approach.

Bioinformatics and statistical bioinformatics are multidisciplinary areas. The materials presented in this book have been developed over the last few years by a group of biologists, biostatisticians, mathematical statisticians, and computer scientists from both academia and pharmaceutical industry. Most of the coauthors of this book are part of the CHIPS (common hour involving practical statistics) network. This network was initiated about 10 years ago as biweekly workshops to give statisticians, bioinformaticians, and life scientists the opportunity to integrate expertise on the design and analysis of microarray experiments. The network has a strong culture of sharing knowledge and expertise among the different professions.

Last, but not least, we would like to thank all our collaborators, without their work this book could never have been published: Marc Aerts, Frank Bretz, Tomasz Burzykowski, Djork-Arn Clevet, An De Bondt, Gemechis D. Dijra, Filip De Ridder, Hinrich. W.H. Göhlmann, Philippe Haldermans, Sepp Hochreiter, Ludwig Hothorn, Adetayo Kasim, Bernet Kato, Martin Otava, Setia Pramana, Pieter Peeters, Tim Perrera, Jose Pinheiro, Nandini Raghavan, Roel Straetemans, Willem Talloen, Suzy Van Sanden, and Tobias Verbeke. We would like to thank Niels Thomas and Alice Blanck from Springer, Heidelberg, and the head of the production team Ms. Ranjani Shanmugaraj for all their help and support during the preparation of this book.

Zaventem, Belgium — Dan Lin
Diepenbeek, Belgium — Ziv Shkedy
Tel-Aviv, Israel — Daniel Yekutieli
Raritan, USA — Dhammika Amaratunga
Beerse, Belgium — Luc Bijnens
July 2012

Contents

Contributors

Marc Aerts Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium

Dhammika Amaratunga Biostatistics and Programming Jansenn Pharmaceutical Companies of Johnson & Johnson Raritan, NJ USA

Luc Bijnens Biostatistics and Programming Jansenn Pharmaceutical Companies of Johnson & Johnson Beerse, Belgium

Frank Bretz Integrated Information Sciences, Novartis Pharma AG, Novartis Campus, Basel, Switzerland

Tomasz Burzykowski Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium

Djork-Arn Clevert Institute of Bioinformatics, Johannes Kepler University, Linz, Austria

An De Bondt Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Filip De Ridder Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Gemechis D. Djira Department of Mathematics and Statistics, South Dakota State University, Brookings, SD, USA

Hinrich W.H. Göhlmann Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Philippe Haldermans Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium

Sepp Hochreiter Institute of Bioinformatics, Johannes Kepler University, Linz, Austria

Ludwig A. Hothorn Institute of Biostatistics, Leibniz University Hannover, Hannover, Germany

Adetayo Kasim Wolfson Research Institute, Durham University, Stockton, UK

Bernet Kato Respiratory Epidemiology and Public Health, Imperial College London, London, UK

Dan Lin Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium

Martin Otava Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium

Pieter Peeters Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Tim Perrera Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Jose Pinheiro Janssen Pharmaceutical Companies of Johnson & Johnson, Titusville, NJ, USA

Setia Pramana Department of Medical Epidemiology and Biostatistics Karolinska Institutet, Stockholm, Sweden

Nandini Raghavan Janssen Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ, USA

Ziv Shkedy Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium

Roel Straetemans Ablynx NV, Zwijnaarde, Belgium

Willem Talloen Biostatistics and Programming Jansenn Pharmaceutical Companies of Johnson & Johnson Beerse, Belgium

Ilse Van Den Wyngaert Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Suzy Van Sanden Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium

Tobias Verbeke OpenAnalytics BVBA, Heist-op-den-Berg, Belgium

Daniel Yekutieli Department of Statistics and Operations Research, Tel-Aviv University, Tel-Aviv, Israel

R Packages

`DoseFinding`
`DoseFinding` is a CRAN package for the design and the analysis of dose-finding experiments. It provides functions for multiple contrast tests, nonlinear dose-response modeling, calculating optimal designs, and an implementation of the MCPMod methodology discussed in Pinheiro et al. (2006).

`fdrame`
`FDR-AME` is a Bioconductor package that computes FDR adjustments for p-values generated in multiple hypotheses testing of gene expression data obtained by a microarray experiment. It applies both theoretical-distribution-based and resampling-based multiple testing procedures and presents as output of the adjusted p-values and p-value plots, as described in Reiner et al. (2003).

`IsoGene`
A CRAN R package for testing monotone trends in dose-response microarray experiments. The package provides several testing procedures discussed in Lin et al. (2007). Inference is based on either the asymptotic distribution of the likelihood ratio test statistic or resampling-based inference for the t-type test statistics. Adjustment for multiplicity is based on either the BH-FDR procedure or SAM.

`IsoGeneGUI`
A graphical user interface for the `IsoGene` package that does not require an extensive knowledge of R. The package performs all the statistical tests implemented in the `IsoGene` and provides several default and user-defined graphical and numerical output. The capacity of the package is discussed in Pramana et al. (2010a,b).

`limma`
The Bioconductor `Limma` package (Smyth, 2004) fits a hybrid frequentist/eBayes linear model for the expression levels of the genes in the array. The package can be used to analyze gene expression data obtained from several microarray platforms such as two-color cDNA (including normalization function for data preprocessing) and Affymetrix.

`MLP`
A Bioconductor R package for analysis of data from a microarray experiment to determine significant sets of genes that are functionally related or in a certain biological pathway. The package performs gene set analysis using the MLP approach described in Raghavan et al. (2006). Genes are mapped into gene sets or pathways by utilizing gene annotation databases such as the Gene Ontology, KEGG, etc. The *p*-values corresponding to genes in a gene set are used to define a gene set statistic. Gene set significance is determined using a permutation procedure based on randomly reassigning *p*-values to genes.

`mratios`
The `mratios` package provides simultaneous inferences for ratios of linear combinations of coefficients in the general linear model. It includes several multiple comparison procedures as applied to ratio parameters, parallel-line and slope-ratio assays, and tests for noninferiority and superiority based on relative thresholds (Dilba et al. 2007).

`multtest`
The Bioconductor package `multtest` uses resampling-based multiple testing procedures for controlling the family-wise error rate (FWER), generalized family-wise error rate (gFWER), and false discovery rate (FDR). Single-step and stepwise methods are implemented. The results are reported in terms of adjusted *p*-values, confidence regions, and test statistic cutoffs. The procedures are directly applicable to identifying differentially expressed genes in DNA microarray experiments. The package is discussed by Dudoit and van der Laan (2008), *Multiple testing procedures with applications to genomics*.

`multcomp`
A CRAN R package for simultaneous tests and confidence intervals for general linear hypotheses in parametric models, including linear, generalized linear, linear mixed effects, and survival models. The package capacity is described in Bretz et al. (2010), *Multiple comparisons using R*.

`nlme`
A CRAN R package for fitting linear mixed models, nonlinear mixed effects models and generelaized linear models. The function `gnls()` can be used to fit nonlinear models. An elaborate discussion about the methodology is given by Pinheiro and Bates (2000), *Mixed effects models in S and S plus*.

`ORCME`
A CRAN R package for simple order-restricted clustering of dose-response microarray data. The `ORCME` package finds clusters of genes with co-regulated dose-response relationship. This package implements a variation of biclustering algorithms of Cheng and Church (2000).

`ORIClust`
An R package for order-restricted clustering of dose-response microarray data. The clustering algorithm implemented in `ORIClust`, ORICC, is a model

selection-based algorithm in which an order-restricted information criterion, proposed by Liu et al. (2009a, 2009b), is used for clustering.

`ORIOGEN` 3.0
`ORIOGEN` 3.0 (Peddada et al. 2005) is not an R package but a java-based interface which can be used to test the null hypothesis of no dose effect against order-restricted alternatives. The methodology implemented in `ROIOGEN` is discussed in Peddada et al. (2003, 2005). The package can be downloaded freely from http://dir.niehs.gov/dirbb/oriogen/index.cfm.

`pava`, `isoreg`, and `monoreg`
`pava` is an R function (from the `iso` package) which calculates a weighted isotonic regression, and `monoreg` is a CRAN R package which fits a nonparametric monotone regression functions. Both functions fit weighted isotonic regression models. Both `pava` and `monoreg` can estimate increasing (isotonic) and decreasing (antitonic) models. An elaborate discussion about the implementation of the Pool-Adjacent-Violators Algorithm (PAVA) in R can be found in De Leeuw et al. (2009).

`R2WinBUGS`
The R2WinBUGS package (Sturtz et al. 2005) consists of a set of R functions which can be used in order to call WinBUGS from R. The package automatically writes the data and scripts in a format readable by WinBUGS.

`samr`
`samr` is a CRAN package that performs the significance analysis of microarray (SAM; Tusher et al. 2001). The SAM is a statistical method for finding significant genes in a set of microarray experiments. The package allows for several types of response variables such as a two-class or multiclass grouping, a quantitative variable, or a censored survival time. The SAM is a resampling-based procedure in determining the significance of the tests and estimating the false positive rate.

References

Bretz, F., Hothorn, T., & Westfall P. (2010). *Multiple comparisons using R*. Boca Raton: CRC.

Cheng, Y., & Church, G.M. (2000). Biclustering of expression data. *Proceedings of the Conference on Intelligent Systems for Molecular Biology, 55*, 93–104.

De Leeuw, J., Hornik, K., & Mair, P. (2009). Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software, 32*(5), 1–24.

Dilba, D., Schaarschmidt, F., & Hothorn, L.A. (2007). Inferences for ratios of normal means. *R News 7, 1*, 20–23.

Dudoit, S., & van der Laan, M.J. (2008). *Multiple testing procedures with applications to genomics*. New York: Springer.

Lin, D., Shkedy, Z., Yekutieli, D., Burzykowki, T., Göhlmann, H.W.H., De Bondt, A., et al. (2007). Testing for trend in dose-response microarray experiments: Comparison of several testing procedures, multiplicity, and resampling-based inference. *Statistical Application in Genetics and Molecular Biology, 6*(1). Article 26.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009a). *Order-restricted information criterion-based clustering algorithm. Reference manual*. http://cran.r-project.org/web/packages/ORIClust/.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009b). Information criterion-based clustering with order restricted candidate profiles in short time-course microarray experiments. *BMC Bioinformatics 10*, 146.

Peddada, S., Lobenhofer, E.K., Li, L., Afshari, C.A., Weinberg, C.R., & Umbach, D.M. (2003). Gene selection and clustering for time-course and dose-response microarray experimants using order-restricted inference. *Bioinformatics, 19*(7), 834–841.

Peddada, S., Harris, S., & Harvey E. (2005). ORIOGEN: Order restricted inference for ordered gene expression data. *Bioinformatics, 21*(20), 3933–3934.

Pinheiro, J., & Bates, D. (2000). Mixed-effects models in S and S-PLUS. Springer, New York

Pinheiro, J.C., Bretz, F., & Branson, M. (2006). Analysis of dose-response studies—Modeling approaches. In N. Ting (Ed.), *Dose finding in drug development* (pp. 146–171). New York: Springer.

Pramana, S. Lin, D., Haldermans, P., Shkedy, Z., Verbeke, T., Göhlmann, H., et al. (2010a). IsoGene: an R package for analyzing dose-response studies in microarray experiments. *The R Journal, 2*(1), 5–12.

Pramana, S., Lin, D., & Shkedy, Z. (2010b). IsoGeneGUI package vignette. *Bioconductor*. http://www.bioconductor.org.

Raghavan, N., Amaratunga, D., Cabrera, J., Nie, A., Qin, J., & McMillian, M. (2006). On Methods for gene function scoring as a means of facilitating the interpretation of microarray results. *Journal of Computational Biology, 13*(3), 798–809.

Reiner, A., Yekutieli, D., & Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics, 19*(3), 368–375.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology, 3*. Article 3. http://www.bepress.com/sagmb/vol3/iss1/art

Sturtz, S., Ligges, U., & Gelman, A. (2005) R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*(3), 1–16.

Tusher, V.G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrys applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences, 98*, 5116–5121.

Acknowledgments

The work of Ludwig A. Hothorn was supported by the German Science Foundation fund HO1687/9-1 and the EC FP7 ESNATS project no. 201619.

Daniel Yekutieli gratefully acknowledged the support from the Wharton School, University of Pennsylvania, for his appointment to the Harry W. Reynolds Visiting International Professorship.

Dan Lin, Ziv Shkedy, Marc Aerts, Tomasz Burzykowski, Philippe Haldermans, Martin Otava, and Setia Pramana gratefully acknowledged the support from the Belgian IUAP/PAI network "Statistical techniques and modeling for complex substantive questions with complex data."

Chapter 1
Introduction

Dan Lin, Willem Talloen, Luc Bijnens, Hinrich W.H. Göhlmann, Dhammika Amaratunga, and Roel Straetemans

1.1 Introduction

The development of new and innovative treatments for unmet medical (Barlow et al. 1972) needs is the major challenge in biomedical research. Unfortunately, for the past decade, there has been a steady decline in the number of new therapies reaching the market, despite of the increased investments in pharmaceutical R&D (FDA 2004). One of the most critical steps in a drug discovery program is target identification and validation (Sams-Dodd 2005). Good drugs are potent and specific, that is, they must have strong effects on a specific biological pathway and minimal effects on all other pathways (Marton et al. 1998). Confirmation that a compound inhibits the intended target (drug target validation) and the identification of undesirable secondary effects are among the main challenges in developing new drugs. This is the reason why dose-response experiments are pivotal in drug discovery programs. Dose-response experiments help us to understand how the drug works and to explore whether it has the desired properties of a potential novel therapy. A compound will only move further in clinical testing when it has a side

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

W. Talloen · L. Bijnens · H.W.H. Göhlmann
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: WTALLOEN@its.jnj.com; LBIJNENS@its.jnj.com; HGOEHLMA@its.jnj.com

D. Amaratunga
Biostatistics and Programming, Janssen Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ USA
e-mail: DAMARATU@its.jnj.com

R. Straetemans
Ablynx NV, Zwijnaarde, Belgium
e-mail: Roel.Straetemans@ablynx.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_1,

effect profile that is acceptable within the dose range that demonstrates a high level of target activity.

Dose-response experiments have a simple concept. The compound of interest is administered at several doses to a biological sample (a cell line, an animal model, a human volunteer, or a patient) and the response is measured. Dose-response experiments allow researchers to assess the relationship between the dose (amount, concentration) of a drug and the response observed. Despite the conceptual simplicity, however, the practical analysis is much more complicated. First, a response can change dose-dependently in a lot of different ways, and many of dose-response relationship are complex and nonlinear. Second, it is difficult to choose an appropriate response measure. Often one may want to investigate even more than one response. This is because a treatment will generally lead to multiple biological reactions, and one needs to try to disentangle direct from indirect responses and desired from undesired effects.

1.2 Dose-Response Modeling

Dose-response models aim to describe the dependency of a specific response on dose. Dose quantifies the amount of drug the subject is exposed to. Most commonly it measures the weight of the chemical compound and is expressed in absolute terms such as milligram (mg) or grams (g) when dealing with clinical studies or in relative terms such as mg/kg when dealing with in vivo animal experiments. The dose as such, however, is not the direct cause for the response or response profile. In fact, in most cases, the molecular concentration at the site of action drives the response to a drug. This information is seldom readily available or even measurable. A precursor for this information is the drug concentration in the blood. Therefore, whenever drug concentrations in plasma are available, dose can be easily, and arguably should be, replaced by concentration. Since this information is not always at hand, dose is a good alternative. There are methods which under certain circumstances, e.g., knowledge of the pharmacokinetic profile of the compound under study, can be used to work with concentration instead of dose when no plasma samples are available (Jacqmin et al. 2007; Jacobs et al. 2010). A detailed description of these techniques is beyond the scope of this book.

Response can basically be any observation of interest that can be measured. Examples range from continuous data, e.g., body weight data, to binomial data where the presence of an event is observed (yes or no) or multinomial data where multiple levels are possible such as different pain scores or types of adverse events. Although the goal of dose-response modeling for each possible type of response is the same, finding an optimal dose, the methodology behind it will differ greatly from one type to the other. The main focus of the book is on continuous data and in particular the evolution of the mean gene expression with respect to dose.

Ruberg (1995a,b) and Chuang-Stein and Agresti (1997) formulated four main questions usually asked in dose-response studies: (1) Is there any evidence of

the drug effect? (2) For which doses is the response different from the response in the control group? (3) What is the nature of the dose-response relationship? and (4) What is the optimal dose? We can answer these questions either by testing for monotone trend of the response with respect to dose or by modeling the dose-response relationship. In both cases, our underlying assumption is that the true relationship between the dose and the response is monotone. In some applications, the underlying assumption of monotonicity is not appropriate, and other non-monotone order-restricted profiles such as the simple tree order, unimodal partial order (umbrella profiles), and cyclical patterns should be considered. For a discussion about monotonicity issues within the dose-response setting, we refer to Cooke (2009) and Louis (2009).

Dose-response modeling refers to implementing a mathematical representation of some true and unknown relationship. Dose-response models can be classified as being empirical or mechanistic in nature. An empirical model, such as the four-parameter logistic model, discussed in Chaps. 4 and 14, serves to adequately describe the observed pattern between a dose and a response without giving an understanding of the underlying biological process. In other words, the parameters present in the model do not represent biological processes. A mechanistic model on the other hand uses mechanistical pathways to explain the observed pattern. In this book, we focus on the first type of models.

1.3 Dose-Response Microarray Experiments

Now the genome of man and other species have been completely sequenced, we enter the so-called "post-genomic era" that concentrates on harvesting the fruits hidden in the genomic text (Lengauer 2001). The advent of biotechnologies such as microarrays allows us to do so by effectively measuring the activity of an entire genome at once under different conditions. The wealth of biological information of this procedure presents immense new opportunities for developing effective therapies. History has taught us that 30–40% of experimental drugs fail because an inappropriate biological target was pursued (Butcher 2003). The major impact of genomic information may therefore be to reduce this biological failure rate by earlier definition of drug targets related to disease susceptibility or progression. This becomes clearer when one reflects about what a "drug-target" actually is. A drug target is a relatively vague term referring to any number of biological molecular classes (proteins, genes, RNA, sugars, ...) that are "druggable". To be druggable, a target needs to be accessible to putative drug molecules and bind them in such a way that a beneficial biological effect is produced. With microarrays and other high-content screening tools, a wide array of target identification and validation technologies becomes available. Genomics, transcriptomics (i.e., gene expression profiling), and proteomics allow researchers to study many of these drug targets in an unprecedented high-content way. They allow researchers to monitor and discover the biological effects of a potential drug. In summary, the availability of

the human genome sequence represents an exciting advance for the development of novel treatments. When combined with high-content screening methods such as microarrays, the success rate for experimental drugs can be expected to improve (Butcher 2003). The major issue with microarray data analysis is the curse of high dimensionality. Because so much information is gathered on biological activity, it becomes a challenge to find the relevant information in the haystack of irrelevant information. One runs the risk of missing the interesting results in a mass of false positive findings.

Microarray dose-response experiments allow researchers to study the relationship between the dose of a drug and the activity of an entire genome at once. It combines the information wealth of microarrays with the benefits of dose-response studies. Their combined use yields two additional advantages. First, the proportion of false positive findings will be substantially reduced as more information on the entire dose-response profile is collected. False positive genes identified by a one-dose treatment study are easier to unmask in multiple-dose studies when the gene has an unrealistic dose-response relationship. Second, genes within the same biological sample may respond differently to drug dose. One therefore wants to investigate more than one gene. How many exactly is difficult to say, but in a discovery phase, it is typically the more the better. In early stages of drug development, one indeed tries to explore as many potential effects of the drug as possible. A microarray dose-response experiment studies the entire genome at once, and is therefore an ideal tool to elucidate variation in dose-dependency of a treatment across all genes and all known pathways.

Although analysis of gene expression data is the main focus of the book, the discussion about microarray technology is beyond the scope of this book. We refer to Amaratunga and Cabrera (2003) and Göehlmann and Talloen (2009) for an elaborate discussion about the microarray technology and topics related to the analysis of microarray data.

1.4 The Book Structure

The general structure of this book is shown in Fig. 1.1. Although the main part of this book is devoted to the specific setting of microarray dose-response experiment, we introduce the main concept of dose-response modeling in the first part of the book. Estimation under order restrictions and inference are discussed in Chaps. 2 and 3, while parametric nonlinear modeling of dose-response data is described in Chap. 4. The methodology discussed in these chapters is introduced in a general setting, and materials for these chapters are used throughout the second part of the book.

The second part of the book starts with an introduction to dose-response microarray experiments and their specific data structure in Chap. 5, in which the case studies are introduced as well. The analysis of microarray data introduces the challenge of multiple testing. A general guidance for the multiple testing problem in a microarray

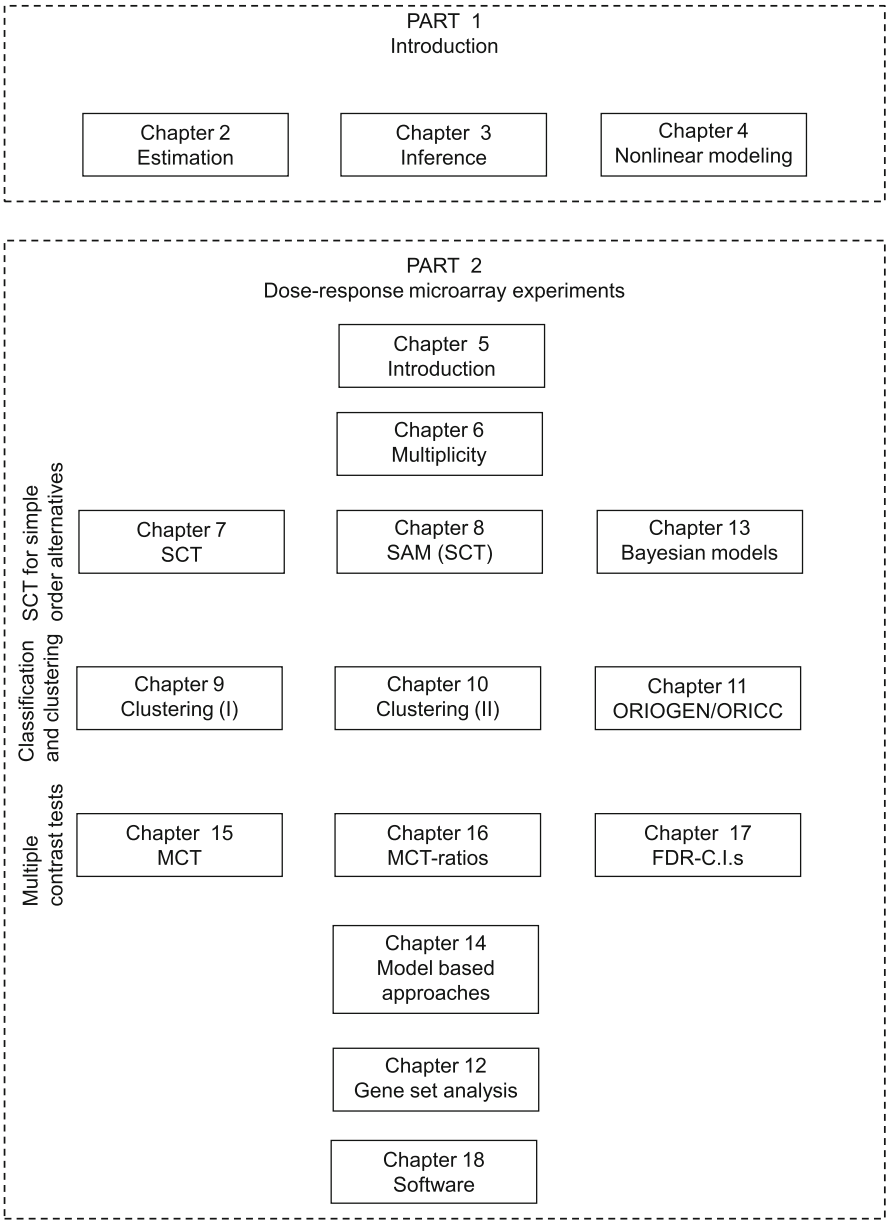

Fig. 1.1 The book structure

setting is given in Chap. 6. We discuss several methods and adjusting procedures such as Bonferoni and Holm's procedures for controlling the family-wise error rate and Benjamini and Hochbergh's procedure and the significance analysis of microarrays (SAM) for controlling the false discovery rate (FDR). Chapters 7, 8,

and 13 are devoted to order-restricted inference in the dose-response microarray setting. We discuss several inference procedures for detecting genes with monotone dose-response relationships such as permutation tests, the SAM for dose-response data, and Bayesian approaches. More advanced inferential topics within the dose-response microarray setting are given in Chaps. 15–17 in which we discuss the topics of multiple contrast tests, ratio tests, and FDR-adjusted confidence intervals. Three chapters in the book are devoted to methods which can be used in order to interpret the results obtained from the inference step. Chapters 9 and 10 present methods for classification and clustering of dose-response curves which can be applied after an initial inference step (discussed in Chaps. 7 and 8). In Chap. 11, we relax the assumption of monotonicity and discuss order-restricted dose-response relationships which are not necessarily monotone such as the unimodal partial order and simple tree order. This setting is further discussed in Chap. 15 in the context of multiple contrast tests. Chapter 12 focuses on the interpretation of the genes detected using the gene set analysis based on Gene Ontology library. As mentioned above, we discuss the general concept of parametric dose-response modeling in Chap. 4 in the first part of the book. In Chap. 14, we focus on parametric modeling of dose-response microarray data. Note that in contrast to other chapters in the second part of the book, the aim of the analysis presented in Chap. 14 is not to detect genes with significant monotone trend, but to perform a secondary analysis in which characteristics of the dose-response relationship are investigated in more details.

The analysis presented in the book is done using several R packages. The methodology discussed in Chaps. 7 and 8 is implemented in the R package `IsoGene` and its graphical user interface `IsoGeneGUI`, which are developed in line with the book. A detailed illustration of the `IsoGeneGUI` is given in Chap. 18, while the use of the `IsoGene` package is illustrated in most of the book chapters. Throughout the book, various R packages are applied for specific settings, which include `nlme`, `multtest`, `ORCM`, `MLP`, `BRUGs`, `MCPMod`, `multcomp`, and `mratios`. Our working assumption is that the readers of this book have a basic knowledge of R, and therefore, complete working examples are provided for the data analysis presented in the book. Readers with limited knowledge of R who wish to perform the analyses presented in the book can do so easily by using the `IsoGeneGUI` which does not require knowledge of R syntax.

1.5 Notation

Throughout the book, we denote d_i as the ith dose level. We use $\mu(d_i)$ and μ_i for the mean gene expression at the ith dose level. Isotonic means at the ith dose are denoted as $\hat{\mu}(d_i)^*$ and $\hat{\mu}_i^*$. Unless specified otherwise, all the models presented in the book are gene specific. In order to simplify notation, we drop the index for the gene. R code is presented in the following way:

```
> # Example of R code.
> age = c(8, 8, 8, 10, 10, 10, 12, 12, 12,14, 14)
> size = c(21,23.5,23,24,21,25,21.5,22,19,23.5,25)
```

Note that, the R code for the analysis is presented as complete as possible. Due to space limit and the length of the code, parts of the code is omitted. The complete R code can be downloaded from the website of the book at:

http://www.ibiostat.be/software/IsoGeneGUI/index.html.

References

Amaratunga & Cabrera, J. (2003). Exploration and Analysis of DNA Microarray and Protein Array Data, New York: John Wiley

Barlow, R.E., Bartholomew, D.J., Bremner, M.J., & Brunk, H.D. (1972). *Statistical inference under order restriction*. New York: Wiley.

Butcher, S. (2003). Target discovery and validation in the post-genomic era. *Neurochemical Research, 28*, 367–377.

Chuang-Stein, C., & Agresti, A. (1997). Tutorial in biostatistics: A review of tests for detecting a monotone dose-response relationship with ordinal response data. *Statistics in Medicine, 16*, 2599–2618.

Cooke, R.M. (Ed.). (2009). *Uncertainty modeling in dose-response*. New York: Wiley.

FDA, U. (2004). *Innovation or stagnation: Challenge and opportunity on the critical path to new medical products*. Silver Spring: US Food and Drug Administation. New York: Wiley

Goehlmann, H., & Talloen, W. (2009). *Gene expression studies using Affymetrix microarrays*. Boca Raton: Chapman & Hall/CRC. New York: Wiley

Jacqmin, P., Snoeck, E., van Schaick, E., Gieschke, R., Pillai, P., Steimer, J.-L., et al. (2007). Modelling response time profiles in the absence of drug concentrations: Definition and performance evaluation of the KPD model. *Journal of Pharmacokinetics and Pharmacodynamics, 34*, 57–85.

Jacobs, T., Straetmans, R., Molenberghs, G., Bouwknecht, A., & Bijnens, L. (2010). Latent pharmacokinetic time profile to model dose-response survival data. *Journal of Biopharmaceutical Statistics, 20*(4), 759–767.

Lengauer, T. (2001). Computational biology at the beginning of the post-genomic era. In *Informatics—10 years back. 10 years ahead* (Vol. 355). Heidelberg: Springer. New York: Wiley.

Louis, T.A. (2009). Math/Stat perspective in Chapter 2: agreement and disagreement, in Cooke, R.M. (Ed.). (2009). *Uncertainty modeling in dose-response*. New York: Wiley.

Marton, M., DeRisi, J., Bennett, H., Iyer, V., Meyer, M. Roberts, C., Stoughton, R., Burchard, J., Slade, D., & Dai, H. (1998). Drug target validation and idenfication of secondary drug target effects using DNA microarrays. *Nature Medicine, 4*, 1293–1301.

Ruberg, S.J. (1995a). Dose response studies. I. Some design considerations. *Journal of Biopharmaceutical Statistics, 5*(1), 1–14.

Ruberg, S. J. (1995b) Dose response studies. II. Analysis and interpretation. *Journal of Biopharmaceutical Statistics, 5*(1), 15–42.

Sams-Dodd, F. (2005). Target-based drug discovery: Is something wrong? *Drug Discovery Today, 10*, 139–147.

Part I
Dose–Response Modeling: An Introduction

Chapter 2
Estimation Under Order Restrictions

Ziv Shkedy, Dhammika Amaratunga, and Marc Aerts

2.1 Introduction

The basic setting on which we focus in the first part of this book is one in which a response variable Y is expected to increase or decrease monotonically with respect to increasing levels of a predictor variable x which in biomedical applications is usually the dose or concentration of a drug. We assume that the mean response is given by

$$E(Y|x) = \mu(x),$$

where $\mu()$ is an unknown monotone function. The case in which $\mu(x)$ is order restricted but not monotone will be discussed in Chaps. 5, 11 and 15 in the second part of the book. The main problem is that although $\mu()$ is a monotone function, unless the sample size at each design point increases to infinity, neither the observed data nor the estimated means are necessarily monotone. For illustration, consider a linear model with five discrete design points:

$$Y_{ij} \sim N(\mu_i, 1), \quad i = 1, 2, 3, 4, 5, \ j = 1, \ldots, n_i. \tag{2.1}$$

Here, μ_i is the true mean at the each design point, $\mu_i = 5$, 5.5, 6, 6.5, and 7, respectively, and n_i is the sample size at each design point. We generate 10×6 datasets according to model (2.1) with the sample size at each design point equal to 5, 10, 25, 50, 100, and 1,000, respectively. Figure 2.1a shows an example of

Z. Shkedy (✉) · M. Aerts
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat),
Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be; marc.aerts@uhasselt.be

D. Amaratunga
Biostatistics and Programming, Jansenn Pharmaceutical Companies of Johnson & Johnson,
Raritan, NJ USA
e-mail: DAMARATU@its.jnj.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_2,

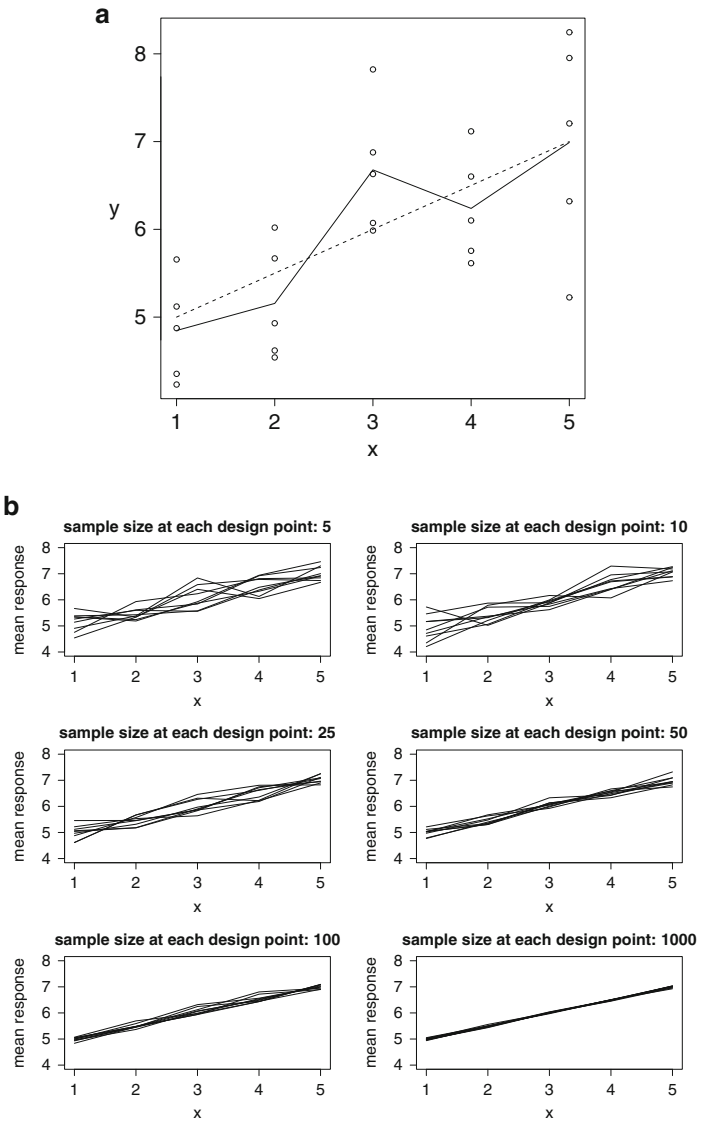

Fig. 2.1 Illustrative example with true mean at each deign point equal to 5, 5.5, 6, 6.5, and 7, respectively. (**a**) Illustrative example of a dataset with five observations at each design point. *Dashed line*: unrestricted means; *solid line*: true (monotone) means. (**b**) Sample means at each design point for ten datasets generated under model (2.1), sample sizes are equal to 5, 10, 25, 50, 100, and 1,000

one dataset with five observations at each design point while Fig. 2.1b shows the estimated means for ten datasets for an increasing number of observations at each design point. Clearly, when sample sizes are relatively small, the observed means are not monotone even though the true means are monotone.

Two questions arise now. The first is how to estimate the mean response under the assumption that the true mean is monotone with respect to x. The second is how to test whether the mean responses in the population are indeed monotone. In this chapter, we focus on the first question and discuss the estimation problem using isotonic regression while the topic of inference under order restrictions will be discussed in Chap. 3.

2.2 Isotonic Regression

According to Barlow et al. (1972), isotonic regression is the statistical theory that deals with problems in which conditional expectations are subject to order restrictions. Let $X = \{x_1, x_2, \ldots, x_n\}$, where $x_1 \leq x_2 \leq \cdots \leq x_n$ denote the finite set of n observed values for the predictor variable and $y_1, y_2, \ldots, y_n$ denote the corresponding observed values for the response variable. Let $\mu(x)$ denote the mean of the conditional distribution, $\mu(x) = E(Y|x)$. Within the framework of linear regression, estimating $\mu(x)$ is typically done by minimizing the least squares criterion in the class of arbitrary linear functions f on X. However, it might be assumed or known that $\mu(x)$ is nondecreasing in x, that is, isotonic with respect to the simple order on X. In that situation, isotonic regression refers to minimizing the least squares criterion in the class of isotonic functions f on X, i.e. f is isotonic, if $x_i \leq x_j, i \neq j$ implies that $f(x_i) \leq f(x_j), i, j = 1, \ldots, n$. Barlow et al. (1972) proposed using the "pool adjacent violators algorithm" (PAVA), where successive approximation is used to isotonize the minimizer of the least squares criterion. Robertson et al. (1988) defined isotonic regression as follows.

Let g be a given function of X. A function $g^\star$ on X is an isotonic regression of g with weight w if and only if $g^\star$ is isotonic and $g^\star$ minimizes

$$\sum_{x \in X} [g(x) - f(x)]^2 w(x), \tag{2.2}$$

in the class of all isotonic functions f on X. The PAVA was proposed by Barlow et al. (1972) and Robertson et al. (1988) in order to minimize (2.2) subject to the constraint $f(x_j) \leq f(x_i)$ for $x_j \leq x_i$.

2.2.1 The PAVA

Let us focus again on an experiment in which the predictor variable X has $K+1$ discrete levels and the response variable has n_i replicates at each level of the predictor variable. Hence, at each design point, the observed data consists of the pairs $\{(x_i, y_{ij})\}$, $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_i$. Without loss of generality, we assume $x_0 \leq \cdots \leq x_K$. Denote the maximum likelihood estimate of $\mu(x_i)$ by $\hat{\mu}(x_i)$.

Suppose i^* is the first index for which $\hat{\mu}(x_i) \geq \hat{\mu}(x_{i+1})$, i.e., the first index for which a "violation" of monotone behavior is observed. The PAVA states that these values need to be "pooled." In other words, $\hat{\mu}(x_i)$ and $\hat{\mu}(x_{i+1})$ are both replaced by the weighted mean:

$$\hat{\mu}(x_i, x_{i+1}) = \frac{n_i \hat{\mu}(x_i) + n_{i+1}\hat{\mu}(x_{i+1})}{n_i + n_{i+1}}.$$

The algorithm proceeds by recursively checking monotone behavior and by pooling if necessary and finally stops if monotonicity is achieved.

2.2.2 Example 1: Size of Pituitary Fissure (Robertson et al. 1988)

Robertson et al. (1988) discussed a dental study in which the size of the pituitary fissure was measured for groups of girls at age 8, 10, 12, and 14. The raw data are shown in Fig. 2.2a. The underlying assumption is that the size of the pituitary fissure increases with age. However, as can be seen clearly in Fig. 2.2a, the observed mean at each age group is not monotone with age since the observed mean at age 12 (20.8333) is smaller than the observed mean at age 10 (23.3333).

```
> age = c(8, 8, 8, 10, 10, 10, 12, 12, 12,14, 14)
> size = c(21,23.5,23,24,21,25,21.5,22,19,23.5,25)
> #sample means at each age group
> msize<-tapply(size,as.factor(age),mean)
> par(mfrow=c(2,2))
> plot(age,size)
> lines(unique(age),msize,lty=2)
> title("a:row data")
> msize
       8       10       12       14
22.50000 23.33333 20.83333 24.25000
```

In the first step of the PAVA, we pool together the means of the second and third age groups,

$$\hat{\mu}(10, 12) = \frac{n_{10}\hat{\mu}(10) + n_{12}\hat{\mu}(12)}{n_{10} + n_{12}}.$$

```
> msize1<-msize
> msize1[2:3]<-(3*23.33333+3*20.83333)/6
> msize1
       8       10       12       14
22.50000 22.08333 22.08333 24.25000
> plot(age,size)
> lines(unique(age),msize,lty=2)
> lines(unique(age),msize1)
> title("b:step 1")
```

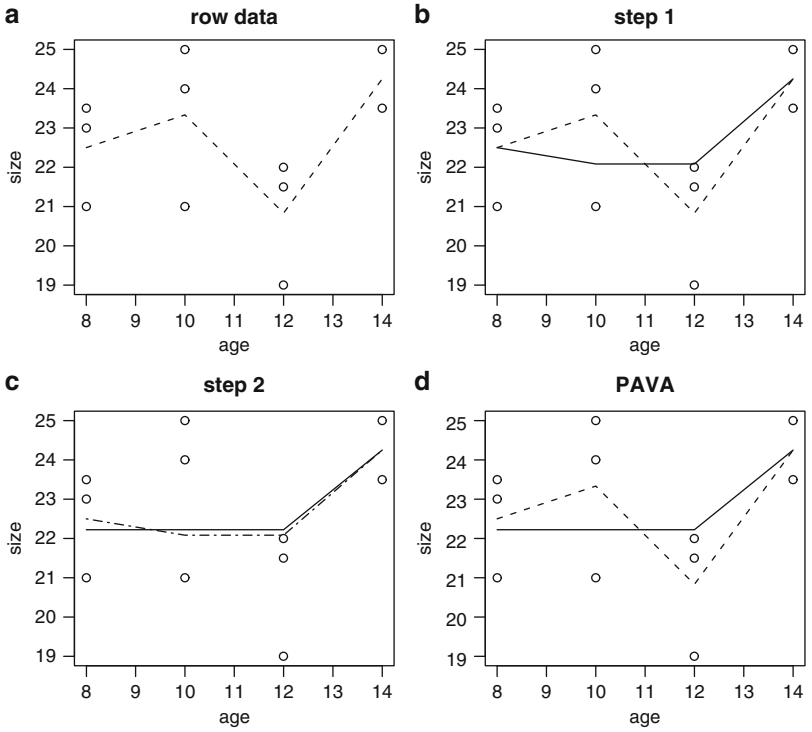

Fig. 2.2 Isotonic regression for the pituitary fissure example from Robertson et al. (1988). Panels (**a**)–(**c**): isotonic regression step by step. Panel (**d**): isotonic regression using the `pava()` function. *Dashed line*: unrestricted means

The result is presented in Fig. 2.2b, and we can see that after the first pooling, the mean in the first age group is higher than the pooled mean of the second and third groups, and therefore, a second pooling is needed,

$$\hat{\mu}(8,10,12) = \frac{n_8\hat{\mu}(8) + (n_{10}+n_{12})\hat{\mu}(10,12)}{n_8 + n_{10} + n_{12}}.$$

```
> msize2<-msize1
> msize2[1:3]<-(3*22.5+6*22.08333)/9
> msize2
       8       10       12       14
22.22222 22.22222 22.22222 24.25000
> plot(age,size)
> lines(unique(age),msize1,lty=2)
> lines(unique(age),msize2)
> title("c:step 2")
```

Figure 2.2c shows that after the second pooling, the isotonic means are monotone. Isotonic regression can be fitted using the R function `pava()`. A general call of the function in the `Iso` package has the form

```
pava(observed means,weights)
```

The first object is a vector of the unrestricted means while the second object is a vector which specifies the weight (i.e., the sample size) at each design point. For the example of the dental study, we use

```
> iso.r1<-pava(msize,w=c(3,3,3,2))
> iso.r1
       8       10       12       14
22.22222 22.22222 22.22222 24.25000
> plot(age,size)
> lines(unique(age),msize,lty=2)
> lines(unique(age),iso.r1)
> title("d:PAVA")
```

2.2.3 Example 2: Isotonic Regression for Binary Data

As a second example, we consider a sample in which the response is a binary variable indicating whether an individual has been infected by a disease or not. Let Z_l be the age of infection of the lth individual, $l = 1, 2, \ldots, N$, a_l be the current age of individual l, and $Y_l = I(a_l > Z_l) = 1$ if the individual has been infected, $Y_l = 0$ if not. Thus, the observed seroprevalence sample consists of N observations, $(a_1, Y_1), \ldots, (a_N, Y_N)$. Let $\pi(a_l) = P(Y_l = 1) = P(a_l > Z_l)$. Then the likelihood of $(y_1, \ldots, y_N)$ is

$$L = \prod_{l=1}^{N} \pi(a_l)^{y_l} (1 - \pi(a_l))^{(1-y_l)}. \tag{2.3}$$

Note that the observation unit in (2.3) is the individual. Alternatively, for data grouped into n unique age groups, $a_1 < a_2 < \cdots < a_n$, the likelihood is

$$L = \prod_{i=1}^{n} \pi(a_i)^{\sum_{j=1}^{n_i} y_{ij}} (1 - \pi(a_i))^{(n_i - \sum_{j=1}^{n_i} y_{ij})}. \tag{2.4}$$

Here, n_i is the sample size in the ith age group, $N = \sum_{i=1}^{n} n_i$ and $\hat{\pi}(a_i) = \sum_{j=1}^{n_i} y_{ij} / n_i$ is the observed prevalence at the ith age group. Figure 2.3 shows the prevalence of rubella in the UK. The maximum likelihood estimators under order constraints for $\pi(a_i)$ in (2.3) and (2.4) are identical. It is the isotonic regression of the observed prevalence, $\hat{\pi}(a_i)$, with weights n_i. The isotonic regression is a step function with respect to age with Ł the final number of sets (or the final number of steps) and $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_Ł$ the jump points. The observed prevalence and the estimated isotonic means are shown in Fig. 2.3. We can see that as long as the observed prevalence is monotone, the isotonic regression will reproduce the observed mean (for example, as in age groups 4.5–9.5). In our example, the final number of sets is equal to 19. The first violation of the order occurs in age groups 2.5 and 3.5 with observed prevalence equal to 0.2055 and 0.2024, respectively. The second violation of the order occurs at age 9.5 and 10.5 (with observed prevalence equal to $\hat{\pi}(9.5) = 0.7444$ and $\hat{\pi}(10.5) = 0.6875$, respectively). For these two age

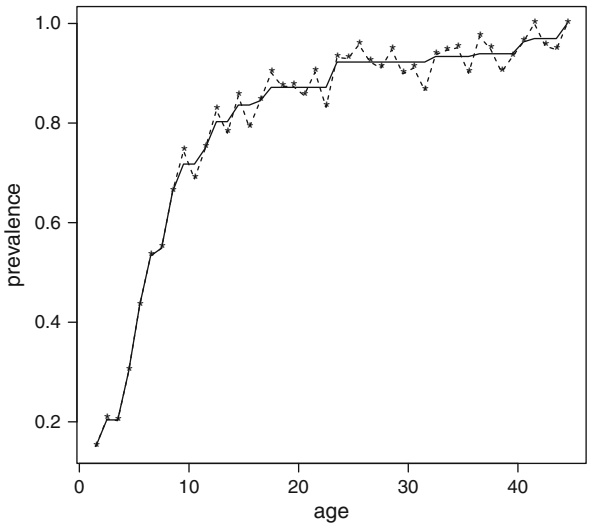

Fig. 2.3 Prevalence of Rubella in the UK. *Dotted* and *dashed line*: observed prevalence. *Solid line*: isotonic regression

groups, the PAVA pools the two means together and estimates the isotonic mean to be equal to $\hat{\pi}(9.5, 10.5) = 0.7176$.

2.3 Graphical Interpretation of the PAVA: The Greatest Convex Minorant

The PAVA discussed in the previous section is the algorithm we use in order to estimate the isotonic regression for a given dataset. Graphical interpretation of isotonic regression is discussed by Barlow et al. (1972) and Robertson et al. (1988). Let w_i be the weight at the ith design point, $G_\ell = \sum_{i=1}^{\ell} g(x_i)w_i$ and $W_\ell = \sum_{i=1}^{\ell} w_i$. The cumulative sum diagram (CSD) is the curve that joins together the two dimensional points $P_\ell = (W_\ell, G_\ell)$, $\ell = 0, \ldots, K$ with $P_0 = (0, 0)$. Figure 2.4a presents an example of isotonic regression with five design points discussed by Silvapulle and Sen (2005). Note that the first two sample means were pooled together by the PAVA.

```
# EXAMPLE from Silvapulle and Sen (2005)
> xi<-c(1,2,3,4,5,6)
> wi<-c(1,1,3,1,2,2)
> yi<-c(-1,-3,1,1,3,4)
> plot(xi,yi,pch="*",xlab="x",ylab="y")
> iso1<-pava(yi,w=wi)
> lines(xi,iso1)
> cbind(xi,wi,yi,iso1)
     xi wi yi iso1
[1,]  1  1 -1   -2
[2,]  2  1 -3   -2
[3,]  3  3  1    1
[4,]  4  1  1    1
[5,]  5  2  3    3
[6,]  6  2  4    4
```

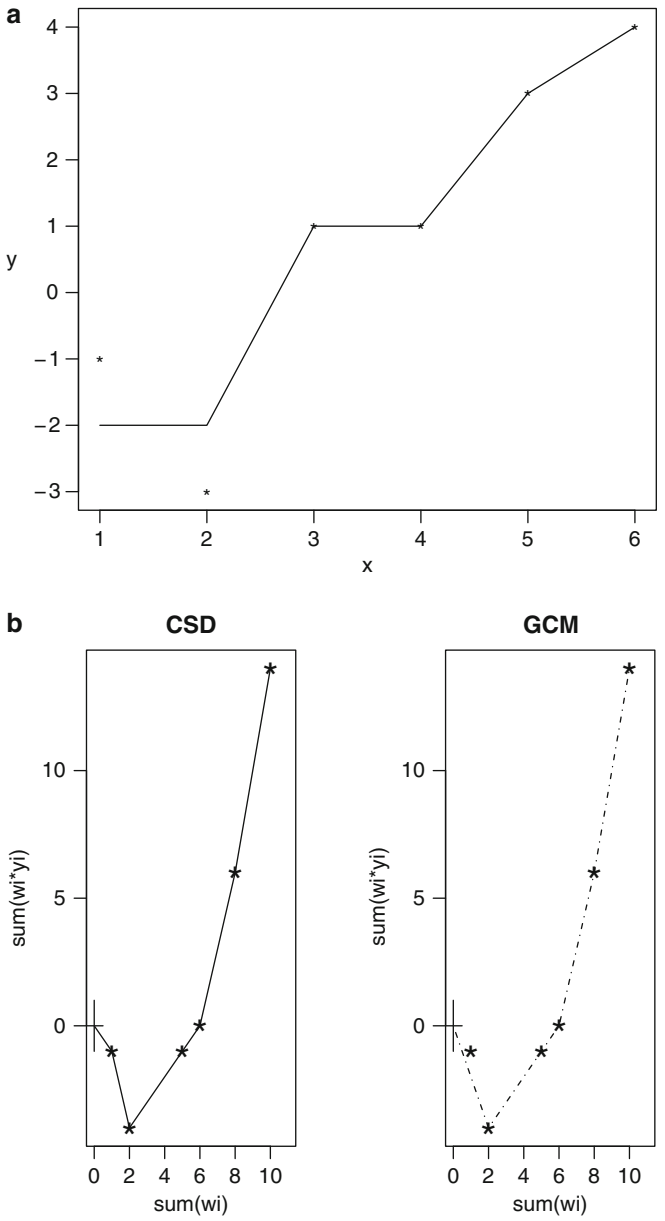

Fig. 2.4 Graphical interpretation of the PAVA. (**a**) Unrestricted (*stars*) means and isotonic regression (*solid line*) for the example in Table 2.9 of Silvapulle and Sen (2005). (**b**) *Left panel*: the cumulative sum diagram (CSD). *Right panel*: the greatest convex minorant plot (GCM)

The corresponding CSD is shown in Fig. 2.4b. Note that the unrestricted estimated of $g(x_i)$ in Fig. 2.4 are the left slopes at P_ℓ of the CSD, that is,

$$g(x_i) = \frac{G_\ell - G_{\ell-1}}{W_\ell - W_{\ell-1}}.$$

The greatest convex minorant (GCM) is defined as the graph of the supremum of all convex functions whose graphs lie below the CSD (Barlow et al. 1972). The right panel in Fig. 2.4b shows the GCM for our example. Note that the slope of the straight line that joins P_0 to P_1 is greater than the slope of the straight line that joins P_0 to P_2. As a result, the point P_1 lies above the GCM, and the first two unrestricted means $\hat{g}(x_1)$ and $\hat{g}(x_2)$ are pooled together by the PAVA.

```
> # Calculation of the cumulative sums G_j and W_j
> # Table 2.9 in Silvapulle and Sen (2005)
> wiyi<-wi*yi
> swi<-cumsum(wi)
> swiyi<-cumsum(wiyi)
> cbind(xi,wi,yi,swi,swiyi)
      xi wi yi swi swiyi
[1,]  1  1 -1   1    -1
[2,]  2  1 -3   2    -4
[3,]  3  3  1   5    -1
[4,]  4  1  1   6     0
[5,]  5  2  3   8     6
[6,]  6  2  4  10    14

# CSD and GCM plots

> par(mfrow=c(1,2))
> plot(swi,swiyi,pch=" ",xlim=c(0,11),xlab="sum(wi)"
       ,ylab="sum(wi*yi)")
> points(swi,swiyi,pch="*",cex=2)
> lines(c(-1,0.5),c(0,0))
> lines(c(0,0),c(-1,1))
> lines(c(0,swi),c(0,swiyi))
> title("CSD")
> plot(swi,swiyi,pch=" ",xlim=c(0,11),xlab="sum(wi)"
       ,ylab="sum(wi*yi)")
> points(swi,swiyi,pch="*",cex=2)
> lines(c(-1,0.5),c(0,0))
> lines(c(0,0),c(-1,1))
> l1<-c(0,2,5,6,8,10)
> l2<-c(0,-4,-1,0,6,14)
> lines(l1,l2,lty=4)
> title("GCM")
```

Let us focus now on the connection between the GCM plot and the PAVA. Figure 2.5a presents an hypothetical example with five design points and five observations at each design point. We can see in Fig. 2.5a and from the panel below that the PAVA pooled together the first two and the last three unrestricted means.

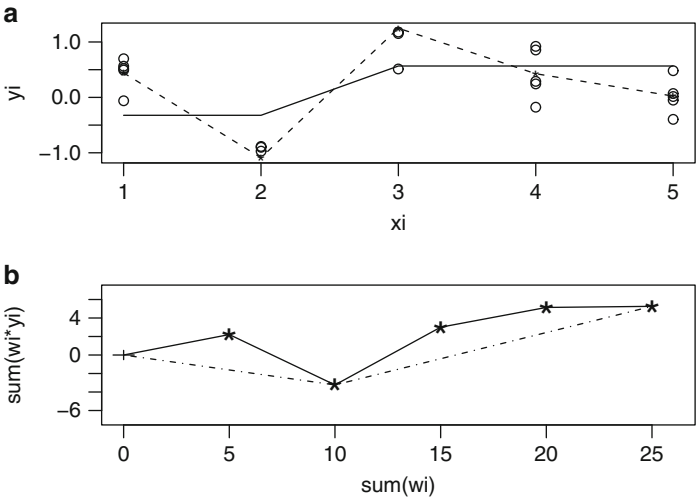

Fig. 2.5 Illustrative example of CSD and GCM (see also Fig. 1.2.3 in Robertson et al. 1988). Panel (**a**): data, observed means (*dashed line*) and isotonic means (*solid line*). Panel (**b**): CSD (*solid line*) and GCM (*dashed-dotted line*)

```
# unrestricted means: yi
# isotonic means: iso1
> cbind(xi,wi,yi,iso1)
  xi wi         yi       iso1
1  1  5  0.4424360 -0.3233958
2  2  5 -1.0892276 -0.3233958
3  3  5  1.2481568  0.5663858
4  4  5  0.4272252  0.5663858
5  5  5  0.0237755  0.5663858
```

Figure 2.5b shows the CSD and the GCM plots. The GCM is the curve which joins P_0–P_2 and P_2– P_5 with straight lines (dashed-dotted line). The slope of the straight line that joins P_0–P_1 is greater than the slope of the straight line that joins P_0–P_2. As a result, P_1 lies above the GCM, and the unrestricted means at x_1 and x_2 are pooled together. Similarly, since the points P_3, and P_4 lie above the GCM, the unrestricted means at x_3, x_4, x_5 are pooled together.

2.4 The Number of Final Sets

Let $\hat{\boldsymbol{\mu}}(x) = (\hat{\mu}(x_0), \hat{\mu}(x_1), \ldots, \hat{\mu}(x_K))$ be the estimated unrestricted means at the design points $x_0, x_1, x_2, \ldots, x_K$ (hence, the unrestricted maximum likelihood estimates) and $n_0, n_1, n_2, \ldots, n_K$ be the sample sizes at each design point. As we showed in the previous sections, the maximum likelihood estimate under order restriction is the isotonic regression of $\hat{\mu}(x_i)$ with weights n_i. We denote the isotonic regression by $\hat{\boldsymbol{\mu}}^{\star}(x) = (\hat{\mu}^{\star}(x_0), \hat{\mu}^{\star}(x_1), \ldots, \hat{\mu}^{\star}(x_K))$. Let ϕ_ℓ be the ℓth final

set, $\ell = 1, \ldots, Ł$. For all design points belonging to the same final set ϕ_ℓ, say, $x_u < x_{u+1} <, \ldots, < x_{u+v}$, the isotonic regression is the same, $\hat{\mu}^\star_{x_u} = \hat{\mu}^\star_{x_{u+1}} = \cdots = \hat{\mu}^\star_{x_{u+v}}$. Hence, given the final numbers of sets, the isotonic regression can be expressed as

$$\hat{\boldsymbol{\mu}}^\star = \mathbf{S}\hat{\boldsymbol{\mu}}, \tag{2.5}$$

where $\mathbf{S}_{K+1\times K+1}$ is a block diagonal smoothing matrix,

$$\mathbf{S} = \begin{pmatrix} S_1 & 0 & 0 & \ldots & 0 \\ 0 & S_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots, \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & S_Ł \end{pmatrix}, \tag{2.6}$$

for which the ilth entry is given by

$$[\mathbf{S}]_{il} = \begin{cases} \frac{n_i}{\sum_{i\in\phi_\ell} n_i} & x_i \text{ belong to the } \ell\text{th final set,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Suppose that design points $x_i, x_{i+1}, \ldots, x_{i+m}$ belong to the final set ϕ_ℓ. Then the rows in the corresponding sub-matrix are identical and given by

$$[\mathbf{S}_\ell]_{i.} = \left(\frac{n_i}{\sum_{i\in\phi_\ell} n_i}, \frac{n_{i+1}}{\sum_{i\in\phi_\ell} n_i}, \ldots, \frac{n_{i+m}}{\sum_{i\in\phi_\ell} n_i} \right), \tag{2.8}$$

and

$$\hat{\boldsymbol{\mu}}^\star(x_i, \ldots, x_{i+m}) = \left(\frac{n_i}{\sum_{i\in\phi_\ell} n_i}, \frac{n_{i+1}}{\sum_{i\in\phi_\ell} n_i}, \ldots, \frac{n_{i+m}}{\sum_{i\in\phi_\ell} n_i} \right) (\hat{\mu}(x_i), \ldots, \hat{\mu}(x_{i+m}))'.$$

It follows that $\sum_i [\mathbf{S}]_{\ell_{ii}} = 1$ and $\text{trace}(\mathbf{S}) = Ł$. Within the framework of non-parametric regression, the trace of the smoothing matrix is equivalent to the effective number of parameters (Hastie and Tibshirani 1990). In our setting, it is simply the final number of sets.

2.4.1 The Number of Levels and the GCM Curve

In the first example presented in Fig. 2.6a, the unrestricted means are in order. As a result, isotonic regression interpolates the unrestricted means. As can be seen in Fig. 2.6b, the GCM in this case is identical to the CSD, so the number of straight lines joining P_ℓ to $P_{\ell+1}$ is equal to $K + 1$. In the second example, presented in Fig. 2.6c, the unrestricted means at x_2 and x_3 were pooled together. This implies

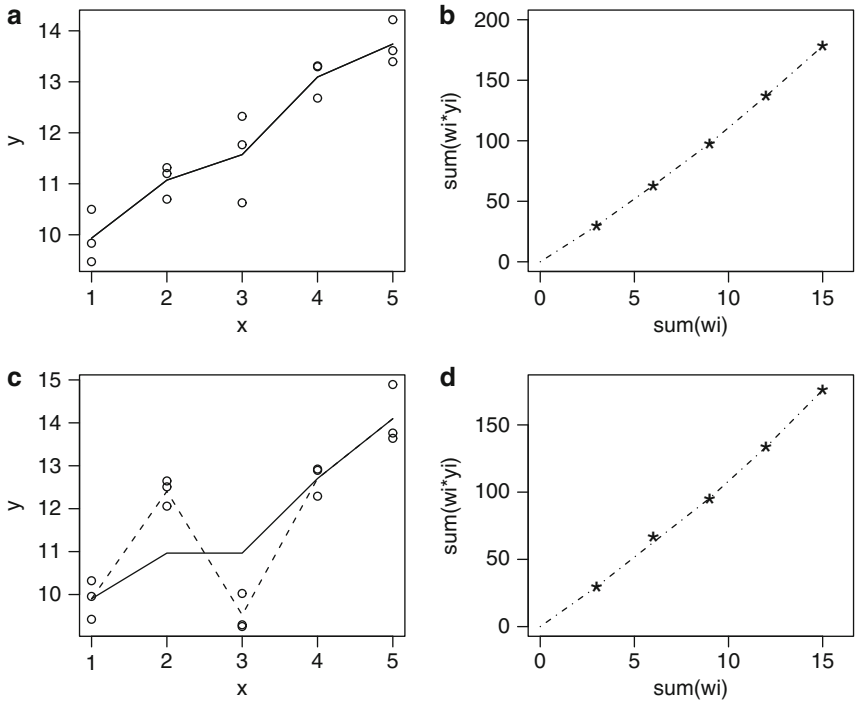

Fig. 2.6 Panel (**a**): data for example I, unrestricted means (*dotted line*) and isotonic regression (*solid line*). Panel (**b**): GCM plot (example I). Panel (**c**): data for example II, unrestricted means (*dotted line*) and isotonic regression (*solid line*). Panel (**d**): GCM plot (example II)

that the number of straight lines joining P_ℓ to $P_{\ell+1}$ in the GCM plot is equal to four. Note that the point P_2 in Fig. 2.6d lies above the GCM curve.

```
# Data for example I

> x1<-c(rnorm(3,10,0.5),rnorm(3,11,0.5),rnorm(3,12,0.5)
       ,rnorm(3,13,0.5),rnorm(3,14,0.5))
> d1<-c(rep(1,3),rep(2,3),rep(3,3),rep(4,3),rep(5,3))
> plot(d1,x1,xlab="dose",ylab="response")
> mx1<-tapply(x1,as.factor(d1),mean)
> lines(unique(d1),mx1)
> iso.r1<-pava(mx1,w=rep(3,3,3,3,3))
> lines(unique(d1),iso.r1)

# Data for example II

> x1<-c(rnorm(3,10,0.5),rnorm(3,11,0.5),rnorm(3,9,0.5)
        ,rnorm(3,13,0.5),rnorm(3,14,0.5))
> d1<-c(rep(1,3),rep(2,3),rep(3,3),rep(4,3),rep(5,3))
> plot(d1,x1,xlab="dose",ylab="response")
> mx1<-tapply(x1,as.factor(d1),mean)
> lines(unique(d1),mx1,lty=2)
> iso.r1<-pava(mx1,w=rep(3,3,3,3,3))
> lines(unique(d1),iso.r1)
```

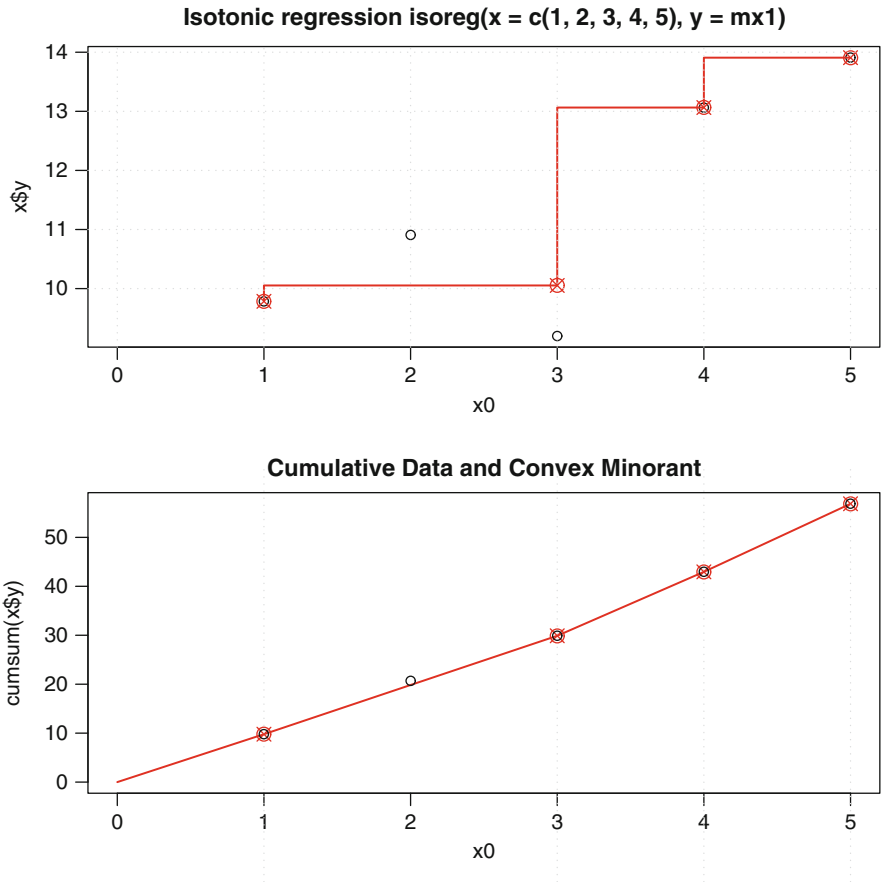

Fig. 2.7 Graphical output of the `isoreg()` function for example II

The GCM plot can be produced automatically with the R function `isoreg()`. A typical call of the function has the form of the `isoreg(x,y)`, where `y` is a vector of the sample means at each design point. For the example discussed above, we use

```
> mx1<-c(9.784394,10.907780,9.197900,13.063718,13.907857)
> iso.fit<-isoreg(c(1,2,3,4,5),mx1)
> plot(iso.fit,plot.type = "row")
```

The object `iso.fit` contains information about the observed isotonic means (`y` and `yf`, respectively) and the cumulative sum (`yc`). Figure 2.7 shows the graphical output of the function. Note that the `isoreg()` function can be used only for the case of equal weights, i.e., equal sample sizes.

```
> iso.fit
Isotonic regression from isoreg(x = c(1, 2, 3, 4, 5), y = mx1),
  with 4 knots / breaks at obs.nr. 1 3 4 5 ;
  initially ordered 'x'
  and further components List of 4
 $ x : num [1:5] 1 2 3 4 5
 $ y : num [1:5]  9.78 10.91  9.20 13.06 13.91
 $ yf: num [1:5]  9.78 10.05 10.05 13.06 13.91
 $ yc: num [1:6]  0.00  9.78 20.69 29.89 42.95 \ldots
```

2.5 Isotonic Regression and Antitonic Regression

The upper panel in Fig. 2.8 shows an example of the propulsion dose-response experiment in which the response decreases with increasing dose. An elaborate description of the experiment is given in Sect. 4.2.3. We use the R function `monoreg()` to obtain the isotonic regression fit of the data. Note that `monoreg()` automatically groups the data and calculates the observed means according to the unique values of the dose. The `monoreg()` function can be used to calculate both increasing (the default) and decreasing monotone means. The response was measured in eight different unique dose levels. At the first two and last two dose levels, there are five observations per dose, while for the other dose levels there are ten observations per dose. Hence, the weights for the isotonic regression (`wi`) are given by $w_i = (5, 5, 10, 10, 10, 10, 5, 5)$. To estimate the isotonic means we use the `monoreg()` function:

```
> library(fdrtool)
> wi <- c(5,5,10,10,10,10,5,5)
> # my1: the observed means at each dose level
> y.up <- monoreg(sort(unique(x1)),my1,w=wi,type=c("isotonic"))
```

In our example, isotonic regression produces a flat profile since all the observed means were pooled together (see the upper panel in Fig. 2.8). The object `yf` is the isotonic mean vector for which all elements are equal to 56.8087.

```
> y.up
$x
[1] -4.6051702 -3.9120230 -3.2188758 -1.8325815 -0.4620355
[6]  0.9162907  2.3025851  3.6888795

$y
[1] 97.01493 83.28822 84.17740 61.74196 43.70133 32.83132 28.22858 28.26821

$w
[1]  5  5 10 10 10 10  5  5

$yf
[1] 56.80866 56.80866 56.80866 56.80866 56.80866 56.80866 56.80866 56.80866

$type
[1] "isotonic"

$call
monoreg(x = sort(unique(x1)), y = my1, w = wi, type = c("isotonic"))

attr(,"class")
[1] "monoreg"
```

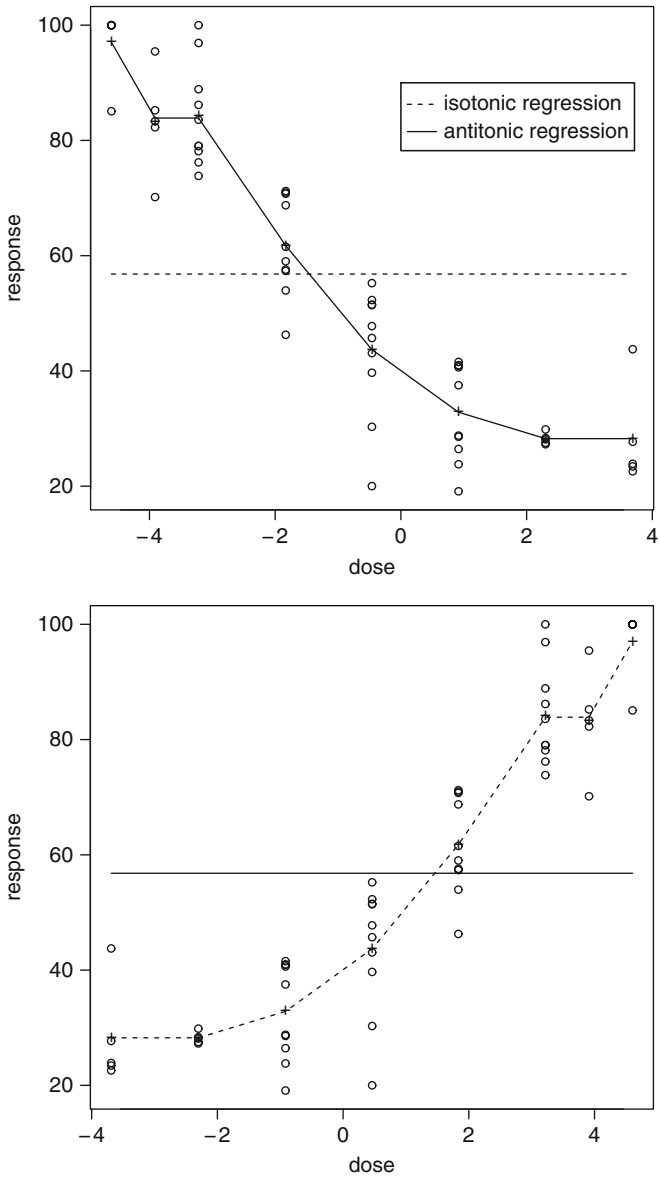

Fig. 2.8 Isotonic and antitonic regression for the propulsion dose-response experiment. *Dashed line*: isotonic regression; *solid line*: antitonic regression. *Upper panel*: data are presented in the original order (response versus log(dose)). *Lower panel*: the data are presented in the reverse order (response versus −log(dose))

In order to calculate a monotone decreasing trend (antitonic regression), we use

```
> y.down <- monoreg(sort(unique(x1)),my1,w=wi,type=c("antitonic"))
```

Note that the observed means of the second and the third dose levels were pooled together by the PAVA since the observed means at these dose levels violate the order, $\hat{\mu}_2 = 83.28822$ and $\hat{\mu}_3 = 84.17740$.

```
> y.down
$x
[1] -4.6051702 -3.9120230 -3.2188758 -1.8325815 -0.4620355
[6]  0.9162907  2.3025851  3.6888795

$y
[1] 97.01493 83.28822 84.17740 61.74196 43.70133 32.83132 28.22858 28.26821

$w
[1]  5  5 10 10 10 10  5  5

$yf
[1] 97.01493 83.88101 83.88101 61.74196 43.70133 32.83132 28.24840 28.24840

$type
[1] "antitonic"

$call
monoreg(x = sort(unique(x1)), y = my1, w = wi, type = c("antitonic"))

attr(,"class")
[1] "monoreg"
```

If we reverse the order of the dose (i.e., treat the highest dose as the lowest dose), the antitonic regression produces a flat profile for the means, while the isotonic regression estimate produces a mirror image of the antitonic means discussed above (see lower panel in Fig. 2.8).

```
> y.down
.
.
$yf
[1] 97.01493 83.88101 83.88101 61.74196 43.70133 32.83132 28.24840 28.24840

$type
[1] "antitonic"

> y.up
.
.
$yf
[1] 28.24840 28.24840 32.83132 43.70133 61.74196 83.88101 83.88101 97.01493

$type
[1] "isotonic"
```

2.6 Discussion

Isotonic functions can be fitted to data using the R functions `isoreg()`, `monoreg()`, and `iso()`. The `isoreg` function can be used to fit isotonic regression models with equal weights. For an elaborate discussion about the

implementation of the PAVA algorithm in R and R functions available for isotonic/monotonic regression we refer to De Leeuw et al. (2009). An example of isotonic regression modeling within the area of dose-response experiments can be found in Chap. 2 of Cooke (2009) where dose-response models using probabilistic inversion with isotonic regression are discussed. The order restrictions discussed in this chapter imply monotonicity, i.e., $\mu(x_1) \leq \mu(x_2), \ldots, \leq \mu(x_K)$. Other order-restricted profiles of the form $\mu(x_1) \leq \mu(x_2), \ldots, \leq \mu(x_s) \geq \mu(x_{s+1}) \geq \mu(x_K)$ will be discussed in Chaps. 5, 11, and 15.

References

Barlow, R. E., Bartholomew, D. J., Bremner, M. J., & Brunk, H. D. (1972). *Statistical inference under order restriction*. New York: Wiley.

Cooke, R. M. (Ed.). (2009). *Uncertainty modeling in dose-response*. New York: Wiley.

De Leeuw, J., Hornik, K., Mair, P. (2009). Isotone optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software*, *32*(5), 1–24.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: Order, inequality, and shape constraints*. New York: Wiley.

Hastie, T.J. and Tibshirani, R.J., (1990) Generalized Additive Models, Chapman & Hall/CRC.

Chapter 3
Testing of Equality of Means Against Ordered Alternatives

Ziv Shkedy, Dhammika Amaratunga, and Dan Lin

3.1 Introduction

In the previous chapter, we discussed the problem of estimation under order restrictions. We used isotonic regression to estimate the order-restricted means. This chapter is devoted to order-restricted inference in a dose-response setting. In Sect. 3.2, we formulate the null hypothesis and the ordered alternatives. In Sect. 3.3, we discuss the test statistics proposed by Williams (1971, 1972) and Marcus (1976); both are "t-test"-type test statistics that can be used to compare each nonzero dose level to the zero dose (i.e., the control) under order constraints. In Sect. 3.4, we discuss the likelihood ratio test (Barlow et al. 1972; Robertson et al. 1988).

3.2 Testing for Homogeneity of the Means Under Order-Restricted Alternatives

Let us consider the following one-way ANOVA model:

$$Y_{ij} = \mu(d_i) + \varepsilon_{ij}, \quad i = 0, 1, \ldots, K, \; j = 1, 2, \ldots, n_i, \tag{3.1}$$

Z. Shkedy (✉)
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be

D. Lin
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

D. Amaratunga
Biostatistics and Programming, Jansenn Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ, USA
e-mail: DAMARATU@its.jnj.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_3,

where Y_{ij} is the jth response at the ith dose level, d_i are the $K+1$ dose levels, d_0 is the zero dose level (i.e., the control group), $\mu(d_i)$ is the mean response at ith dose level, and $\varepsilon_{ij} \sim N(0, \sigma^2)$ independent of one another. The null hypothesis of no dose effect is given by

$$H_0 : \mu(d_0) = \mu(d_1) = \cdots = \mu(d_K). \tag{3.2}$$

A one-sided alternative hypothesis of a positive dose effect for at least one dose level (i.e., an increasing trend) is specified by

$$H_1^{\text{Up}} : \mu(d_0) \leq \mu(d_1) \leq \cdots \leq \mu(d_K), \tag{3.3}$$

with at least one strict inequality. When testing the effect of a drug for a positive outcome, the researcher may be able to specify a positive effect as the desirable alternative. However, without prior knowledge, it seems reasonable to assume that the response levels may increase or decrease in response to increasing dose, but with the direction of the trend not known in advance. Thus, we must also consider an additional alternative:

$$H_1^{\text{Down}} : \mu(d_0) \geq \mu(d_1) \geq \cdots \geq \mu(d_K), \tag{3.4}$$

with at least one strict inequality.

Testing H_0 against H_1^{Down} or H_1^{Up} requires estimation of the means under both the null and the alternative hypotheses. Under the null hypothesis, the estimator for the mean response $\hat{\mu}$ is the sample mean across all the data in all the dose groups. As we showed in Chap. 2, the isotonic (antitonic) regression $\hat{\boldsymbol{\mu}}^\star = (\hat{\mu}_0^\star, \hat{\mu}_1^\star, \ldots, \hat{\mu}_K^\star)$ is the maximum likelihood estimate for the means (at each dose level) under the ordered alternatives [either (3.3) or (3.4) depending on the direction of the trend].

3.3 t-Test-Type Statistics: Williams' (1971, 1972) and Marcus' (1976)

Williams' procedure defines H_0 as the null hypothesis, and either H_1^{Up} or H_1^{Down} as the one-sided alternative. Williams' (1971, 1972) test statistic was suggested for a setting in which n observations are available at each dose level. Each dose level is compared with the control level using the test statistic:

$$t_i = \frac{\hat{\mu}_i^\star - \bar{y}_0}{\sqrt{2s^2/n}}. \tag{3.5}$$

Here, $\bar{y}_0$ is the sample mean at the zero dose level (control), $\hat{\mu}_i^\star$ is the estimate of the mean at the ith dose level under the ordered alternative, and s^2 is an estimate of the variance. For the $\hat{\mu}_i^\star$, Williams (1971, 1972) used the isotonic regression estimates

of the observed means with respect to dose. In case the number of observations at dose i $n_i \neq n$, the denominator of Williams' test statistic can be adjusted as a two-sample t-test.

Williams' test procedure is a sequential procedure. In the first step, $\hat{\mu}_K^\star$ is compared to $\bar{y}_0$. If the null hypothesis is rejected (implying that there is a dose-response trend across the K groups), $\hat{\mu}_{K-1}^\star$ is compared to $\bar{y}_0$, etc. This process stops whenever a null hypothesis is not rejected. If this happens at the K'th dose level ($0 \leq K' \leq K$), then it implies that there is a significant difference between the control dose and the $(K'+1)$th dose level onwards. In certain contexts, the $(K'+1)$th dose is referred to as the "minimum effective dose" (MED). Identifying the MED (or its analog, the "no effect level" or NOEL, which would be the K'th dose level) is important in many pharmaceutical settings.

Williams (1971) discussed an example of a dose-response experiment including six dose levels and a zero control with $n = 8$ observations at each dose level. The observed means and isotonic regression for the experiment are shown in Fig. 3.1a. Note that the PAVA algorithm pools together the first three dose levels and the last two dose levels ($\hat{\mu}_0^\star = \hat{\mu}_1^\star = \hat{\mu}_2^\star = 10.1, \hat{\mu}_3^\star = 10.6, \hat{\mu}_4^\star = 11.4, \hat{\mu}_5^\star = \hat{\mu}_6^\star = 11.8$).

```
> ybar<-c(10.4,9.9,10.0,10.6,11.4,11.9,11.7)
> dose<-c(0,1,2,3,4,5,6)
> plot(dose,ybar,ylab="mean response")
> lines(dose,ybar,lty=4)
> iso.y<-pava(ybar,w=rep(1,7))
> lines(dose,iso.y)

> cbind(dose,ybar,iso.y)
     dose ybar iso.y
[1,]    0 10.4  10.1
[2,]    1  9.9  10.1
[3,]    2 10.0  10.1
[4,]    3 10.6  10.6
[5,]    4 11.4  11.4
[6,]    5 11.9  11.8
[7,]    6 11.7  11.8
```

Williams (1971) assumed that the mean square error s^2 is equal to 1.16. The test statistic to compare the last and the first dose levels is

$$t_6 = \frac{11.8 - 10.4}{\sqrt{2 \times 1.16/8}} = 2.60.$$

The distribution of the test statistic under the null hypothesis was derived by Williams, and the critical value for significance at 5% is 1.81. As shown in Fig. 3.1b, a significant difference can be detected between the last dose level by and the control.

Williams (1971) proposed a modification of the test statistic (3.5) that replaced $\bar{y}_0$ with $\hat{\mu}_0^\star$, the estimate of the first dose (control) mean under order restrictions:

$$t_i = \frac{\hat{\mu}_i^\star - \hat{\mu_0}^\star}{\sqrt{2s^2/n}}. \tag{3.6}$$

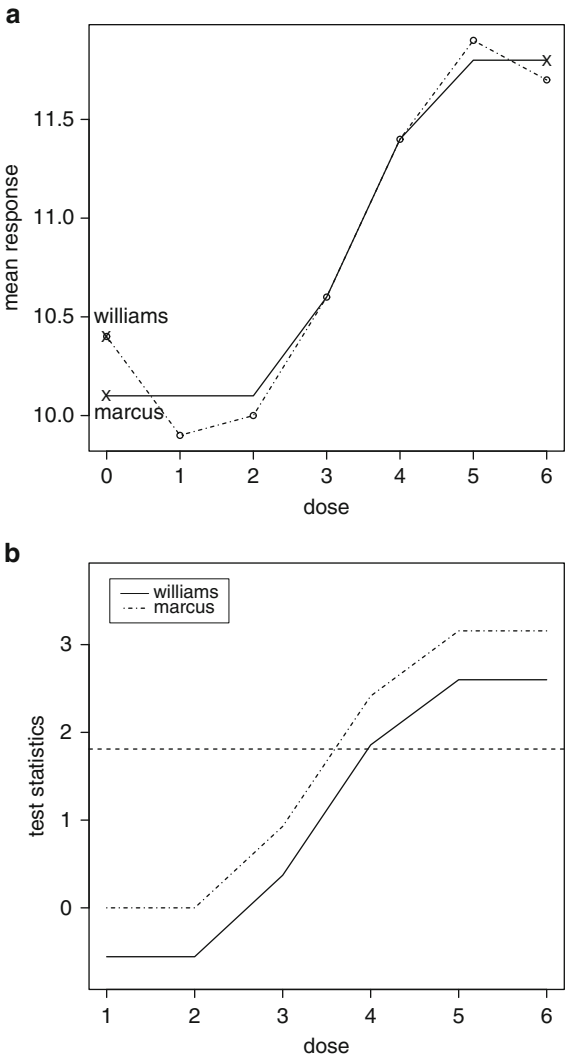

Fig. 3.1 Williams' and Marcus' test statistics. (**a**) Example of dose-response experiment with six dose levels and a control from Williams (1971). *Dashed line*: mean response; *solid line*: isotonic regression. (**b**) Williams' (*solid line*) and Marcus' (*dotted dashed line*) test statistics at each dose level. *Dashed horizontal line*: the critical value for Williams' test statistics (for $\alpha = 0.05$, $K + 1 = 7$, $n = 8$)

Marcus (1976) derived the distribution under the null hypothesis of the modified test statistic (3.6). For the example presented above, the modified test statistic for the last two dose levels is

```
#williams
> (11.8-10.4)/sqrt(2*1.16/8)
[1] 2.599735
#Marcus
> (11.8-10.1)/sqrt(2*1.16/8)
[1] 3.156821
```

It has been shown that the performance of Marcus' test is close to that of Williams' in terms of power (Marcus 1976). In certain cases, the two tests are even equivalent. For $K = 1$, Williams' and Marcus' test statistics reduce to the two-sample t-test. The two test statistics are also equal in the case that the observed means are monotone since in that case, the isotonic regression reproduces the observed means.

3.3.1 The Distribution of Williams' and Marcus' Test Statistics Under the Null Hypothesis

Although both Williams' and Marcus' test statistics are t-test-type statistics and formulated in a similar way, for a given direction, they differ in terms of the distribution of the test statistics under the null hypothesis. Figure 3.2 shows an example of a dose-response experiment with four dose levels and three observations at each dose level. We test the null hypothesis (3.2) against the one sided ordered alternative H_1^{Up}. The observed test statistics are 2.69 and 3.63 for Williams' and Marcus' tests, respectively.

```
> #y:vector of gene expression data.
> #x: vector of dose levels (1,1,1,2,2,2,3,3,3,4,4,4)
> xi<-unique(x) # xi is the dose levels
> wi<-c(3,3,3,3)
> yi.i<-tapply(y,as.factor(x),mean)
> plot(x,y,pch="*",xlab="dose",ylab="gene expression")
> lines(xi,yi.i,lty=2)
> iso.i<-pava(yi.i,w=wi)
> lines(xi,iso.i)
> mse<-anova(aov(y~as.factor(x)))[2,3]
> x0<-yi.i[1]
> m0<-iso.i[1]
> m3<-iso.i[4]
> r<-3
>#Williams test statistic
> tkw.obs<-(m3-x0)/(sqrt(2*mse/r))
>#Marcus test statistic
> tkm.obs<-(m3-m0)/(sqrt(2*mse/r))

>#Williams
> tkw.obs
2.695187

>#Marcus
> tkm.obs
3.639801
```

We use a nonparametric bootstrap approach to approximate the distribution of the test statistics under the null hypothesis. We will elaborate on the bootstrap algorithm in the next section. The bootstrap p values are equal to 0.011 and 0.008 for the Williams' and Marcus' statistics, respectively. Therefore, the null hypothesis is rejected in both tests. Figure 3.2b shows the density estimate for the distributions of the test statistics under the null and reveals a fundamental difference. While the distribution of Williams' test statistic is symmetric around zero, the distribution of Marcus' test statistic is truncated at zero.

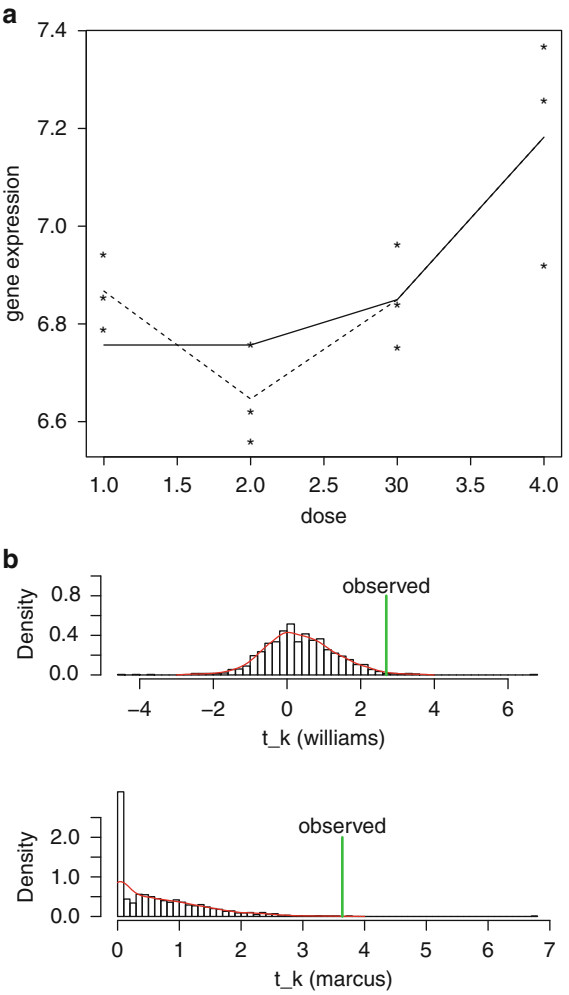

Fig. 3.2 An illustrative example of the distribution under the null hypothesis for Williams' and Marcus' test statistics. For both methods, the mean gene expression in the highest dose level is compared with the mean gene expression of the control group. (**a**) Example of a dose-response experiment. *Dashed line*: mean response; *solid line*: isotonic regression. (**b**) Bootstrap distribution of Williams' (*upper panel*) and Marcus' (*lower panel*) test statistics. *Vertical line*: observed test statistics

```
> # bootstrap for the dose-response experiment
> r<-3
> B<-1000
> tkw.boot<-tkm.boot<-c(1:B)
> for(b in 1:B){
+ index<-sample(c(1:length(y)), length(y), replace = TRUE)
+ y.boot<-y[index]
+ y.m<-tapply(y.boot,as.factor(x),mean)
+ y.is.u<-pava(y.m,w=wi)
+ x0<-y.m[1]
+ m0<-y.is.u[1]
+ m3<-y.is.u[4]
+ mse<-anova(aov(y.boot~as.factor(x)))[2,3]
+ tkw.boot[b]<-(m3-x0)/(sqrt(2*mse/r))
+ tkm.boot[b]<-(m3-m0)/(sqrt(2*mse/r))
+ }
```

```
>#  p-value Williams
> sum(tkw.boot>tkw.obs)/B
[1] 0.011
>#  p-value Marcus
> sum(tkm.boot>tkm.obs)/B
[1] 0.008
```

We illustrate the difference between the two test statistics using three hypothetical examples presented in Fig. 3.3. In Fig. 3.3a, $\bar{x}_0 = \hat{\mu}_0^\star$ and as a result, both test statistics are equal to 1.105. In Fig. 3.3b, we see $\bar{x}_0 > \hat{\mu}_0^\star$, and since $\hat{\mu}_3^\star > \bar{x}_0 > \hat{\mu}_0^\star$, Williams' test statistic is smaller than Marcus' test statistic, but both statistics are positive. Figure 3.3c shows an example in which $\bar{x}_0 > \hat{\mu}_0^\star$ and $\bar{x}_0 > \hat{\mu}_3^\star$. As a result, Williams' test statistic is negative, whereas Marcus' test statistic is positive.

3.4 Likelihood Ratio Test Statistic for Monotonicity

Williams' and Marcus' procedures are step-down procedures, i.e., the comparison between a lower dose and control is tested only if the test of a higher dose versus the control is significant. The underlying assumption is that there is a monotone dose-response relationship with a known direction. In this section, we discuss the likelihood ratio test when the direction is unknown. Generally, the likelihood ratio test presented in this section is applied once, not sequentially, as with the t-test-type of the previous section. It is worth noting, however, that when the objective is to identify the MED, the likelihood ratio test too can be applied sequentially in the same way as presented in Sect. 3.3 (Amaratunga and Ge 1998); this approach, however, is not pursued further in this book.

As a starting point, let us focus on testing the null hypothesis against an unconstrained alternative H_2, i.e.,

$$H_0 : \mu(d_0) = \mu(d_1) = \cdots = \mu(d_K), \quad \text{versus} \quad H_2 : \mu(d_i) \neq \mu(d_\ell), \tag{3.7}$$

for at least one pair of i and ℓ, $i \neq \ell$. The maximum likelihood estimator under the null hypothesis H_0 is the sample mean $\hat{\mu} = \bar{y}$; while under the alternative H_2, the maximum likelihood estimators are the sample means at each dose level $(\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_K) = (\bar{y}_0, \bar{y}_1, \ldots, \bar{y_K})$. Following Silvapulle and Sen (2005), we define the residual sum of squares under each hypothesis:

$$\begin{aligned} \text{RSS}_0 &= \textstyle\sum_{ij} \left(y_{ij} - \bar{y}\right)^2, \quad \text{residual sum of squares under } H_0, \\ \text{RSS}_2 &= \textstyle\sum_{ij} \left(y_{ij} - \bar{y}_i\right)^2, \quad \text{residual sum of squares under } H_2. \end{aligned} \tag{3.8}$$

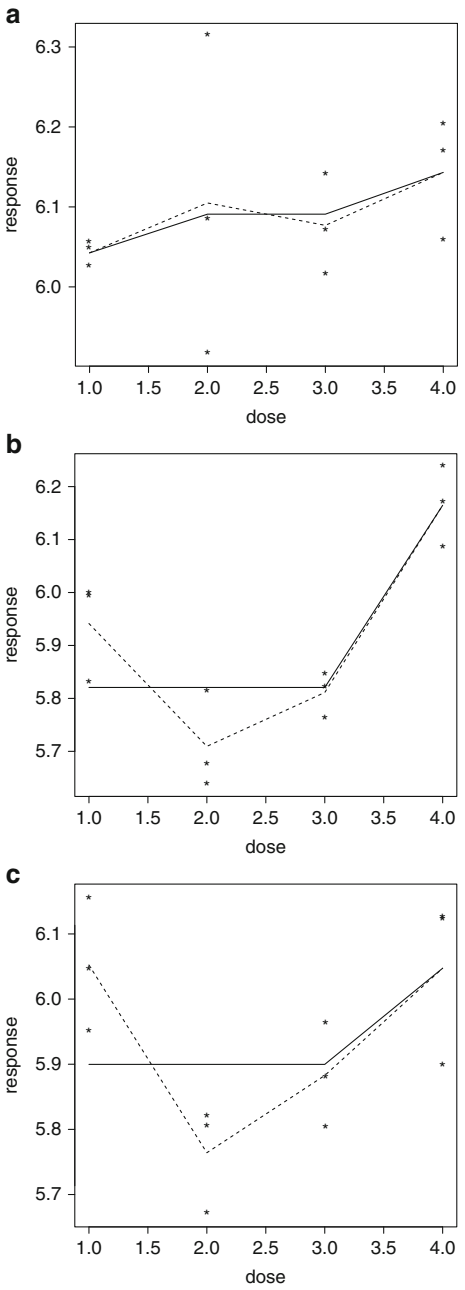

Fig. 3.3 Hypothetical examples of dose-response experiment with four dose levels with three replicates at each dose level. *Dashed line*: unconstrained mean; *solid line*: isotonic regression. Test statistics are calculated for the last dose level versus the control. (**a**) Example 1: $t_{\text{Williams}} = t_{\text{Marcus}} = 1.105$. (**b**) Example 2: $t_{\text{Williams}} = 3.44,\ t_{\text{Marcus}} = 5.303$. (**c**) Example 3: $t_{\text{Williams}} = -0.0612,\ t_{\text{Marcus}} = 1.792$

The null hypothesis will be rejected if the difference $\mathrm{RSS}_0 - \mathrm{RSS}_2$ is large, or equivalently if the value of the pseudo-F statistic given by

$$\tilde{F} = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_2}{\mathrm{RSS}_0} = 1 - \frac{\mathrm{RSS}_2}{\mathrm{RSS}_0}, \tag{3.9}$$

is large. In what follows, we define an equivalent pseudo-F statistic for the case of an ordered alternative.

Testing the equality of ordered means using a likelihood ratio test, for the case that the response is assumed to be normally distributed, was discussed by Barlow et al. (1972) and Robertson et al. (1988). The likelihood ratio test works out to be the ratio of the error variances under the null and the alternative hypotheses:

$$\Lambda_{01}^{\frac{2}{N}} = \frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2} = \frac{\sum_{ij}(y_{ij} - \hat{\mu}_i^\star)^2}{\sum_{ij}(y_{ij} - \hat{\mu})^2} = \frac{\mathrm{RSS}_1}{\mathrm{RSS}_0}, \tag{3.10}$$

where $\hat{\sigma}_{H_0}^2$ and $\hat{\sigma}_{H_1}^2$ are the ML estimates for the error variance under the null and the alternative hypothesis, respectively, and RSS_1 is the residual sum of squares under the ordered alternative. The null hypothesis is rejected for a small value of $\Lambda_{01}^{\frac{2}{N}}$. Equivalently, H_0 is rejected for large value of $\bar{E}_{01}^2$, where

$$\bar{E}_{01}^2 = 1 - \Lambda_{01}^{\frac{2}{N}} = \frac{\sum_{ij}(y_{ij} - \hat{\mu})^2 - \sum_{ij}(y_{ij} - \hat{\mu}_i^\star)^2}{\sum_{ij}(y_{ij} - \hat{\mu})^2} = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_0}. \tag{3.11}$$

Estimating the parameters $\mu(d_i)$ using isotonic regression requires knowledge of the direction of the trend. In practice, the direction of the trend may not be known in advance. In such a case, one can maximize the likelihood twice: for a monotone decreasing trend and for a monotone increasing trend, and choose the trend with the higher likelihood. In practice, we can calculate $\bar{E}_{01}^2$ for each direction and choose the higher value of $\bar{E}_{01}^2$ (Barlow et al. 1972).

3.4.1 The Distribution of $\bar{E}_{01}^2$ Under the Null Hypothesis

The null distribution of $\bar{E}_{01}^2$ is a mixture of beta distributions with mixture probabilities $P(\ell, L, \boldsymbol{w})$. The mixture probabilities are also known as the level probabilities, the probability that the number of final sets of the isotonic regression is equal to ℓ in an experiment with L possible levels. As shown by Barlow et al. (1972), the p value can be calculated by

$$P(\bar{E}_{01}^2 \geq C) = \Sigma_{\ell=1}^{L} P(\ell, L, w) P(B_{\frac{1}{2}(\ell-1), \frac{1}{2}(N-\ell)} \geq C) \tag{3.12}$$

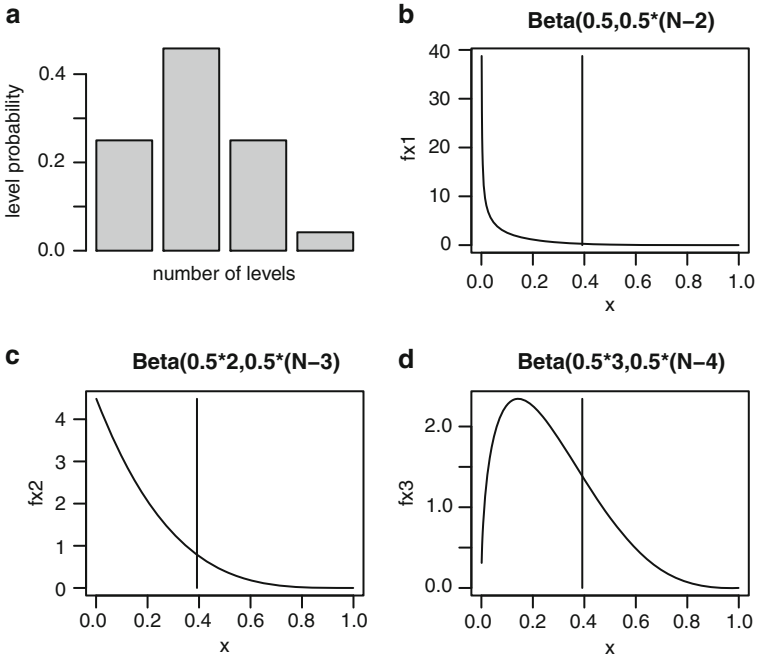

Fig. 3.4 Level probabilities and the components for the beta mixture distribution for $N = 12$ and $L = 4$

Here, N is the total number of observations, ℓ is the number of final levels, $B_{\frac{1}{2}(\ell-1),\frac{1}{2}(N-\ell)}$ denotes a beta distribution with $\alpha = 1/2(\ell-1)$ and $\beta = 1/2(N-\ell)$, and $\boldsymbol{w}^{-1} = (w_0^{-1}, \ldots, w_L^{-1})$ is the variance of the response at each dose. Figure 3.4a shows the level probabilities for the case with $N = 12$ and $L = 4$, while Fig. 3.4b–d shows the beta components for this case. The vertical line shows the critical value (0.392) for the significance level of 0.05.

```
# calculation on p value for N=12 and L=4
# critical value for N=12, L=4 and alpha=0.05 (0.392)
> 0.45833*(1-pbeta(0.392,0.5,0.5*(N-2)))+
+ 0.25*(1-pbeta(0.392,0.5*2,0.5*(N-3)))+
+ 0.04167*(1-pbeta(0.392,0.5*3,0.5*(N-4)))
[1] 0.05011451

# critical value for N=12, L=4 and alpha=0.01 (0.575)
> 0.45833*(1-pbeta(0.575,0.5,0.5*(N-2)))+
+ 0.25*(1-pbeta(0.575,0.5*2,0.5*(N-3)))+
+ 0.04167*(1-pbeta(0.575,0.5*3,0.5*(N-4)))
[1] 0.009983637
```

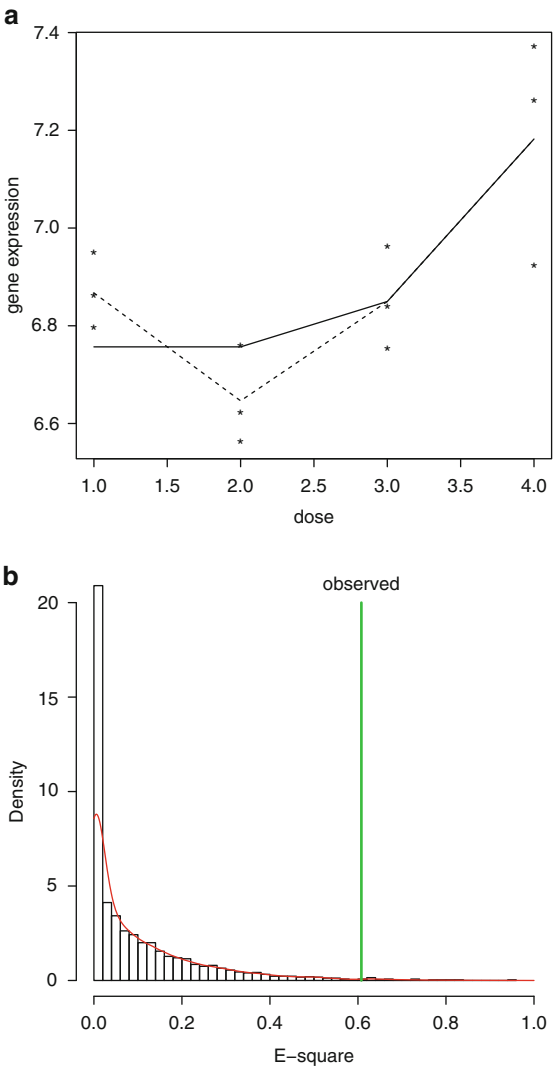

Fig. 3.5 A dose-response microarray experiment with four dose levels with three replicates at each dose level. (**a**) Gene expression versus dose. *Solid line*: isotonic regression. *Dashed line*: observed means. (**b**) Histogram and density estimate for the bootstrap replicates of $\bar{E}_{01}^2$

3.4.2 Example: Dose-Response Microarray Experiment

We consider an example of a dose-response microarray experiment consisting of four dose levels (the first is the control) and three samples at each dose level. Figure 3.5a shows an example of one gene which we use for illustration in this chapter.

```
# x: dose
# y: gene expression
> cbind(x,y)
      x        y
V2    1 6.948563
V3    1 6.859254
V4    1 6.793810
V5    2 6.621029
V6    2 6.560225
V7    2 6.758055
V8    3 6.960683
V9    3 6.837769
V10   3 6.750577
V11   4 7.257968
V12   4 6.920253
V13   4 7.368275
> xi <- unique(x) # xi is the dose levels
> yi.i<-tapply(y,as.factor(x),mean)
> plot(x,y,pch="*",xlab="dose",ylab="gene expression")
> lines(xi,yi.i,lty=2)
> iso.i<-pava(yi.i,w=wi)
> lines(xi,iso.i)
```

The observed test statistic is equal to 0.6078, and the p value of 0.0069 is calculated using the distribution of $\bar{E}_{01}^2$ according to (3.12). The small p value indicates that the null hypothesis should be rejected.

```
> NN<-12
> y.m<-tapply(y,as.factor(x),mean)
> y.is.u<-pava(y.m,w=wi)
> rep.iso.u<-rep(y.is.u,wi)
> RSS0<-sum((y-mean(y))^2)
> RSS1<-sum((y-rep.iso.u)^2)
> Ebar01.obs<-1-(RSS1/RSS0)
> (RSS0-RSS1)/RSS0
[1] 0.6077756
> 0.45833*(1-pbeta(Ebar01.obs,0.5,0.5*(NN-2)))+
+0.25*(1-pbeta(Ebar01.obs,0.5*2,0.5*(NN-3)))+
+0.04167*(1-pbeta(Ebar01.obs,0.5*3,0.5*(NN-4)))
[1] 0.006993
```

The analysis presented above relies on the assumption that gene expression observations are normally distributed. An alternative approach is to perform a bootstrap test (Efron and Tibshirani 1993) which does not require any distributional assumption for validity. The basic idea is to generate bootstrap samples under the null hypothesis by resampling with replacement from the gene expression vector and keeping the dose level fixed. We used the following bootstrap algorithm:

1. For b in $1 : B$

   (a) Generate a bootstrap sample of size N drawn with replacement under the null hypothesis from the observed data:

$$(x_1, y_1^{(b)}), (x_2, y_2^{(b)}), \ldots, (x_N, y_N^{(b)}).$$

(b) For each bootstrap sample, calculate the residuals sum of squares under the null and alternative hypotheses, $\mathrm{RSS}_0^{(b)}$ and $\mathrm{RSS}_1^{(b)}$, respectively, and calculate the bootstrap replicate for the test statistic

$$\bar{E}_{01}^{2(b)} = 1 - \frac{\mathrm{RSS}_1^{(b)}}{\mathrm{RSS}_0^{(b)}}.$$

2. Calculate the p value by

$$P = \frac{\#(\bar{E}_{01}^{2(b)} > \bar{E}_{01}^{2(\mathrm{obs})})}{B}.$$

The above bootstrap algorithm can be implemented in R as follows:

```
> wi<-c(3,3,3,3)
> set.seed(100)
> B<-2000
> Ebar01.boot<-c(1:B)
>#begin the bootstrap
> for(i in 1:B)
+ {
+ index<-sample(c(1:length(y)), length(y),
                replace = TRUE)
+ y.boot<-y[index]
+ y.m<-tapply(y.boot,as.factor(x),mean)
+ y.is.u<-pava(y.m,w=wi)
+ rep.iso.u<-rep(y.is.u,wi)
+ RSS0<-sum((y.boot-mean(y.boot))^2)
+ RSS1<-sum((y.boot-rep.iso.u)^2)
+ Ebar01.boot[i]<-1-(RSS1/RSS0)
+ }
>#end of the bootstrap
> # P-VALUE
> sum(Ebar01.boot>Ebar01.obs)/B
[1] 0.01
```

The bootstrap p value is equal to 0.01, indicating that the null hypothesis should be rejected. Figure 3.5b shows the histogram of the bootstrap replicates for $\bar{E}_{01}^2$.

3.5 Discussion

In Chaps. 2 and 3, we discussed the estimation and inference for a one-way ANOVA model of the form

$$Y_{ij} = \mu(d_i) + \varepsilon_{ij}.$$

We use isotonic regression in order to estimate the mean response under order restrictions at each dose level. The testing procedures we discussed in this chapter were used either to test the global null hypothesis of no dose effect (using the

LRT test) or to compare between the mean response of two dose levels (Williams' and Marcus' test statistics). Using isotonic regression implies that the estimated dose-response curve is a step function. In the next chapter, we focus on parametric dose-response models for which

$$\mu(d_i) = f(\boldsymbol{\theta}, d_i).$$

Here, $f()$ is a smooth continuous function of the dose (linear or nonlinear), and $\boldsymbol{\theta}$ is the parameter vector to be estimated. Hence, in contrast with Chaps. 2 and 3, the focus of chapter 4 is on the dose-response curve shape which depends on different parameterization for the mean structure of f.

References

Amaratunga, D., & Ge, N. (1998). Step-down trend tests to determine a minimum effective dose. *Journal of Biopharmaceutical Statistics, 8*, 145–156.

Barlow, R. E., Bartholomew, D. J., Bremner, M. J., & Brunk, H. D. (1972). *Statistical inference under order restriction*. New York: Wiley.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Marcus, R. (1976). The powers of some tests of the quality of normal means against an ordered alternative. *Biometrika, 63*, 177–183.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: Order, inequality, and shape constraints*. New York: Wiley.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics, 27*, 103–117.

Williams, D. A. (1972). The comparision of several dose levels with a zero dose control. *Biometrics, 28*, 519–531.

Chapter 4
Nonlinear Modeling of Dose-Response Data

Roel Straetemans

4.1 Introduction

The importance of dose-response modeling within the pharmaceutical development process of a chemical substance to a medicine is perfectly summarized by a famous and well-known statement made by Paracelsus in the sixteenth century: "all substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy." Finding the efficacious dose or dose range without having unacceptable side effects nowadays is even more challenging due to increasing complexities of diseases and their molecular pathways. Therefore, knowledge of the dose-response characteristics of a compound is crucial to successfully developing a new medicine. In contrast with the previous chapters in which we discussed the LRT for an order-restricted ANOVA model, this chapter is focused on parametric dose-response modeling of a continuous response variable. The main model of interest will be the four-parameter logistic (4PL) model. Interpretation of the model parameters, as well as practical implementation using the R function `gnls()`, is given in Sect. 4.2. In Sect. 4.3, we discuss several alternatives for the 4PL model such as the 3PL and 5PL models, the Emax model, and growth models such as the Gompertz function and Richards function. Finally, as described in Sect. 1.2, whenever PK information in the form of plasma drug concentrations is available, dose response modeling should be replaced by concentration response modeling. For the models described in this chapter, dose and concentration are interchangeable

R. Straetemans (✉)
Ablynx NV, Zwijnaarde, Belgium
e-mail: Roel.Straetemans@ablynx.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_4,

4.1.1 The Goal of Dose-Response Modeling

Dose-response modeling can serve different purposes. It will be mainly used to describe the observed relationship between the dose range studied and a response of interest. Equally it can easily be used to predict dose values corresponding to certain effect levels of interest or vice versa within the available data or, when the model allows it, outside the dose range and/or effect range studied via extrapolation. A final main objective of dose-response modeling is to test for possible differences between treatments. We might, for example, be interested if two treatments have a different potency, i.e., the dose needed to achieve 50% of the maximum effect of each treatment, and the model can be used to quantify these differences and test for its statistical and biological relevance.

4.2 The 4PL Model

4.2.1 Typical Dose-Response Pattern

A common pattern in dose-response data is a symmetrical sigmoidal increasing or decreasing relation between the response variable and increasing dose values when dose is portrayed on a logarithmic scale as shown in Fig. 4.1.

This pattern can be described by a 4PL model (Straetemans and Bijnens 2010) given by

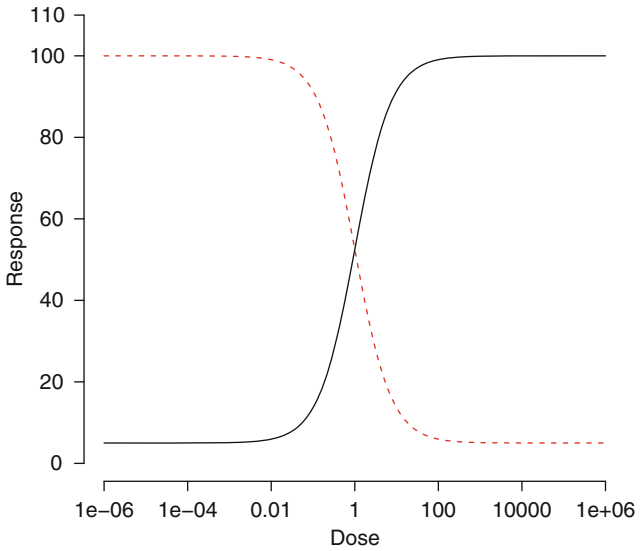

Fig. 4.1 Illustration of common sigmoidal dose-response patterns

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{1 + 10^{(x_i - \theta_2)\theta_3}} + \varepsilon_{ij}, \quad i = 0, \ldots, K, \; j = 1, 2, \ldots, n_i, \tag{4.1}$$

where Y_{ij} is the jth observation at the ith dose level, x_i is the dose i expressed as decadic logarithm (base 10 value), θ_1 and θ_4 are the response values in the asymptotic region of the curve, θ_2 is the x (dose) value for which half the effect (difference between the two asymptotes) is obtained and θ_3 is a slope parameter. It is further assumed that, $\varepsilon_{ij} \sim N(0, \sigma^2)$. The 4PL model (4.1) uses the log10 parametrization and can be specified in the $\log_e$ parametrization if wanted or deemed more suitable. The 4PL nonlinear model in (4.1) can generally be written as

$$Y_{ij} = f(\boldsymbol{\theta}, x_i) + \varepsilon_{ij}, \tag{4.2}$$

where f is the nonlinear function in (4.1) and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$, is a parameter vector to be estimated.

4.2.2 Interpretation of the Parameters of the 4PL Model

4.2.2.1 The Asymptote Parameters θ_1 and θ_4

The parameters θ_1 and θ_4 in Eq. (4.1) correspond to the response values at the plateau of the sigmoidal curve depicted in Fig. 4.2a. Which of the two parameters corresponds to the asymptote at dose zero and which to the asymptote at infinite dose depends on the value of the slope parameter θ_3? If $\theta_3 > 0$, θ_1 corresponds to the response value at infinite dose and θ_4 the response value at dose zero as illustrated in the left panel of Fig. 4.2a. If $\theta_3 < 0$, the interpretation of θ_1 and θ_4 is reversed as shown in the right panel of Fig. 4.2a.

4.2.2.2 The Slope Parameter θ_3

The parameter θ_3 is a slope parameter which determines the steepness of the curve or in other words which quantifies the sensitivity of the response to the dose range of the drug. Figure 4.2b shows three 4PL models which are identical with the exception of the value for θ_3 and illustrates how a larger absolute value for θ_3 corresponds to a steeper curve. An important consequence of a steep dose-response curve is that the response is sensitive to relative small changes in the dose range between the two plateau values.

4.2.2.3 The ED_{50} Parameter θ_2

The parameter θ_2 is commonly named as the ED_{50} dose or alternatively ID_{50}, EC_{50} or IC_{50} depending on the response (effect or inhibition) and the exposure (dose or concentration). It is the dose at which 50% of the effect is observed, where 50% of

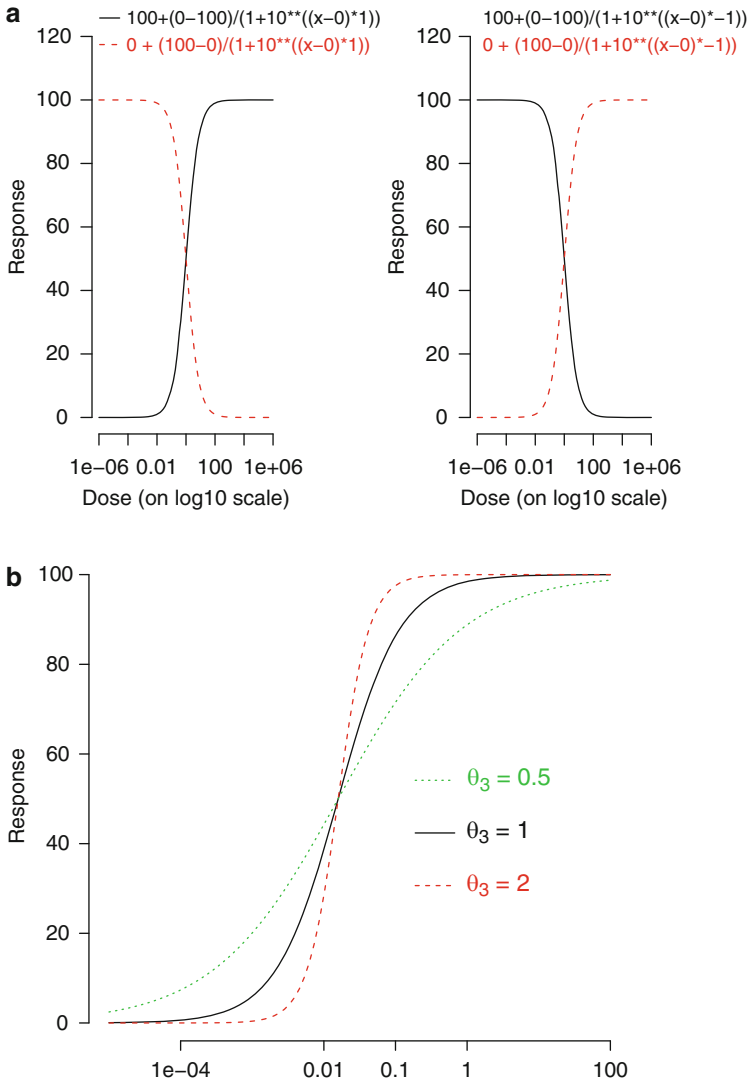

Fig. 4.2 Parameters in the 4PL model. (**a**) θ_1 and θ_4 parameters in the 4PL model (4.1). *Left panel*: $\theta_3 > 0$. *Right panel*: $\theta_3 < 0$. (**b**) The influence of θ_3 on the sigmoidal pattern of the 4PL model

the effect corresponds to the value at half the difference between the two asymptotes. Notice that θ_2 is the dose value at the inflection point of the sigmoidal curve. This is graphically illustrated in Fig. 4.3a where the dose-response profile goes from a minimum response of 20 at the lower dose to a maximum response of 100 at the highest dose, meaning that the maximum obtained dose effect is an increase of 80. The ED_{50} parameter corresponds to the dose for which the response value is 60.

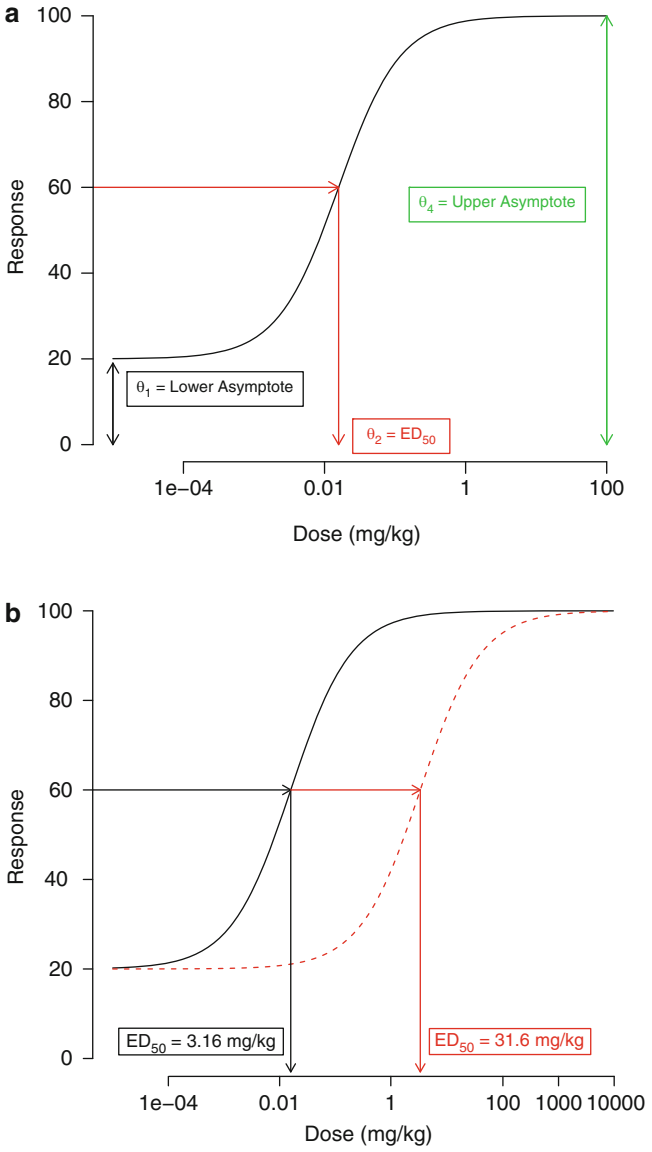

Fig. 4.3 The ED_{50} (θ_2) parameter in the 4PL model. (**a**) Illustration of influence of θ_2 on the sigmoidal pattern of the 4PL model. (**b**) Dose-response curve for two compounds following a 4PL dose-response model (4.1) where one compound is ten times as potent as the other

A common mistake is to think that, in cases where the response value is a percentage, the ED_{50} parameter is the dose for which the response value equals 50%. This is only true if the lower and upper asymptotes are 0% and 100% or in the case where half the maximum effect is given by 50%, e.g. asymptotes go from 20% to 80%.

The ED_{50} parameter is also called the *potency* of a compound, since it quantifies the dose necessary to reach 50% of the maximal effect and as such is a measure of how potent the compound is. Let us assume that at the early stages of research, an animal study is performed to compare the efficacy of two compounds. When the ED_{50} dose of one compound (Fig. 4.3b, solid line) is ten times as low as the ED_{50} dose of the other (dotted line), this compound is said to be ten times as potent.

Potency however only tells a part of the story. When two compounds have the same *efficacy window* (difference between the two asymptotes), and assuming both are equivalent with respect to factors such as toxicity and pharmacokinetic (PK) characteristics, the difference in potency is obviously very meaningful and the more potent compound will be the preferred chemical entity to go further in development. If a difference in efficacy window between two compounds exists, a difference in their potency becomes less meaningful. Imagine an example where the more potent compound has a maximal effect of 80% and the less potent compound a maximal effect of 100%. With all other relevant characteristics being comparable, the choice of preference is less obvious. Overall, however, the potency, expressed as the ED_{50} dose, is a powerful tool to summarize and compare different compounds under the assumption that the molecular weights are comparable.

4.2.3 Example: The Propulsion Dataset

4.2.3.1 The Experiment and Data

The propulsion dataset originates from an *in vivo* charcoal meal test in rats. This test was developed to evaluate effects of a test compound on bowel motility. The protocol of the study is as follows. A charcoal meal is administered by gavage to rats. Twenty minutes after feeding, the rats are sacrificed and the small intestine is surgically removed. The length traversed by the charcoal is divided by the total length of the small intestine and multiplied by 100, resulting in the percent distance traveled as the response of interest. For an elaborate discussion about the experiment, we refer to Tanila et al. (1993).

The study design involved a parallel group design with a vehicle group (dose = 0) and eight doses (0.01, 0.02, 0.04, 0.16, 0.63, 2.5, 10, or 40 mg/kg) of three chemicals. Each dose group had at least 5 animals, the vehicle group had 64 animals, and some dose groups had 10 animals. The raw data are shown in Fig. 4.4, with the individual animal data in the left panel and the averages by treatment and dose in the right panel. The upper panel shows the response plotted versus dose, and the lower panel shows the response plotted versus $\log_e$ dose, with dose 0 replaced by a small value for plotting purposes. Figure 4.4 shows that there is a decrease in ratio with

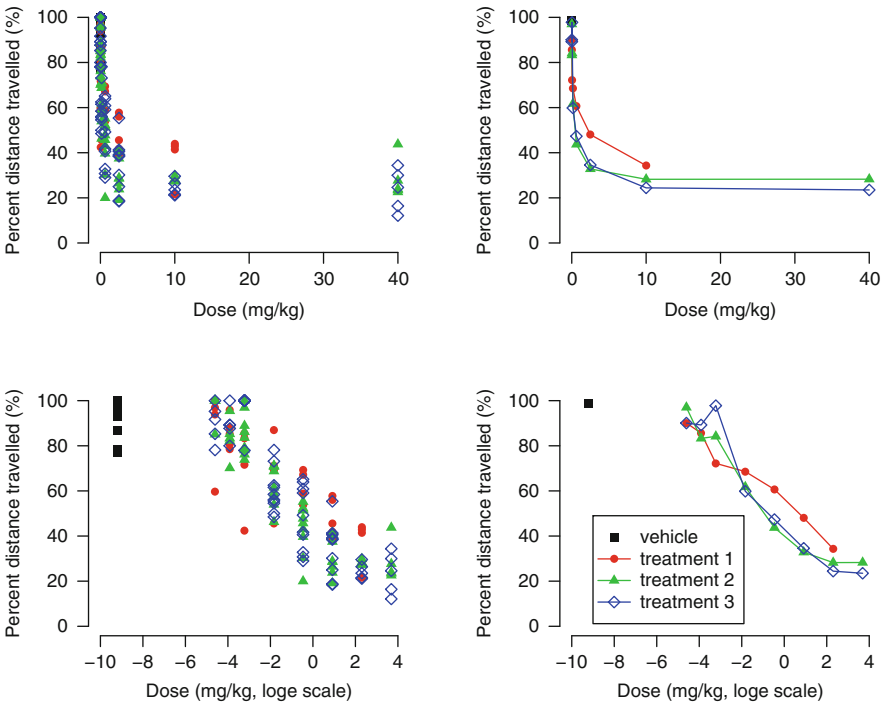

Fig. 4.4 The propulsion dataset. *Left panels*: individual data. *Right panels*: mean profiles. *Top panels*: dose plotted on the absolute scale (mg/kg). *Bottom panels*: dose plotted on the $\log_e$ scale

increasing dose for the three treatments, that the average vehicle response is close to 100%, and that for treatment 1 no responses at the maximum dose (40 mg/kg) are available. Noteworthy is the apparently less steep decline of the ratio under treatment 1 compared to the other two treatments.

4.2.3.2 Implementation of the 4PL Model Using the *R* Function `gnls()`

Our primary interest is to test for treatment effects, i.e., to test if the three treatments differ in their dose-response profile. The following 4PL model was fitted to the data:

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{1 + \mathrm{e}^{(x_i - \theta_2)\theta_3}} + \varepsilon_{ij}\,, \tag{4.3}$$

Notice that the dose in Eq. (4.3) is expressed as the natural logarithm (base e value), the reason being the doubling pattern in the dose range studied. The 4PL model can be fitted in R using the function `gnls()` (Pinheiro and Bates 2000). A general call for the function has the form:

```
gnls(response ~ mean structure, starting values, options)
```

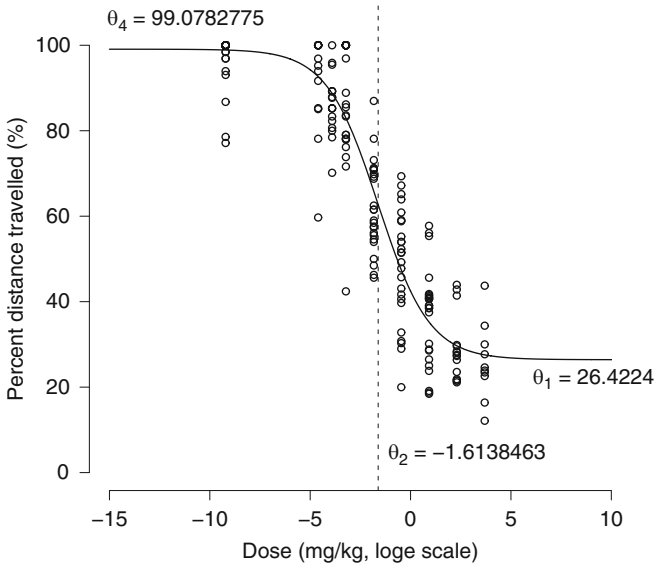

Fig. 4.5 The propulsion data and the estimated 4PL model

The 4PL model (4.3) was fitted using the following code:

```
gnls.model001<-gnls(
    ratio~(th1+(th4-th1)/(1+(exp((lmpk-th2)*th3)))),
    data=data2b,
    params=list(th1+th2+th3+th4~1),
    start=c(90,-0.2,1,28),
    control=gnlsControl(nlsTol=0.1))
```

Note that the mean structure is specified by

```
(th1+(th4-th1)/(1+(exp((lmpk-th2)*th3))))
```

Where `lmpk` is the dose on the $\log_e$ scale. The option

```
params=list(th1+th2+th3+th4~1)
```

implies that the model is fitted without taking into account possible treatment differences. Parameter estimates for $\boldsymbol{\theta}$ are shown in the panel below and in Fig. 4.5. The vertical line is the ED_{50} value, $\hat{\theta}_2 = -1.614$, while the parameter estimates for the upper ($\hat{\theta}_4$) and lower ($\hat{\theta}_1$) asymptote are equal to 99.08 and 26.42, respectively.

```
> summary(gnls.model001)
Generalized nonlinear least squares fit
  Model: ratio ~ (th1 + (th4 - th1)/(1 + (exp((lmpk - th2) * th3))))
  Data: data2b
       AIC      BIC    logLik
  1627.413 1644.358 -808.7064

Coefficients:
       Value Std.Error   t-value p-value
th1 26.42247 2.4654373  10.71715       0
th2 -1.61385 0.1534649 -10.51606       0
th3  0.78141 0.0755436  10.34384       0
th4 99.07828 1.2469491  79.45655       0

 Correlation:
    th1    th2    th3
th2 -0.728
th3  0.702 -0.408
th4 -0.197 -0.237 -0.443

Standardized residuals:
       Min         Q1        Med         Q3        Max
-4.1327652 -0.5091635  0.1135234  0.5000150  2.2931539

Residual standard error: 9.806069
Degrees of freedom: 219 total; 215 residual
```

In the next step, the 4PL model is fitted to the data, taking into account possible treatment effects. This can be done by including a dummy variable for each treatment. In the code below, the R objects `dum1`, `dum2`, `dum3`, and `dum4` are dummy variables for the vehicle data, treatment groups 1, 2, and 3, respectively, which take the value 1 for the respective group and zero for the others. This model specification allows estimation of θ_4 based on the vehicle data and estimation of a separate θ_1, θ_2, and θ_3 parameters for each of the active treatments (with different lower asymptotes, ED_{50}s, and slopes). The model was fitted using the following code:

```
gnls.model002<-gnls(
        ratio ~ th4*(dum1) +
        (th11 + (th4-th11)/(1 + (exp((lmpk-th21)*th31))))*(dum2) +
        (th12 + (th4-th12)/(1 + (exp((lmpk-th22)*th32))))*(dum3) +
        (th13 + (th4-th13)/(1 + (exp((lmpk-th23)*th33))))*(dum4),
          data = data2b,
        params=list(th4+th21+th22+th23+th31+th32+th33+
                    th11+th12+th13~1),
          start=c(90,
        -0.2, -1.3, -1.3,
        0.4, 0.9, 1,
        28, 28, 28),
          control=gnlsControl(nlsTol=0.1, apVar=TRUE))
```

Data and estimated models for each group are shown in Fig. 4.6. Parameter estimates are shown below.

```
> summary(gnls.model002)
Generalized nonlinear least squares fit
  Model: ratio ~ th4 * (dum1)
  + (th11 + (th4 - th11)/(1 + (exp((lmpk -th21) * th31)))) * (dum2)
  + (th12 + (th4 - th12)/(1 + (exp((lmpk -th22) * th32)))) * (dum3)
  + (th13 + (th4 - th13)/(1 + (exp((lmpk -th23) * th33)))) * (dum4)
```

```
  Data: data2b
       AIC      BIC    logLik
  1603.530 1640.809 -790.7648

Coefficients:
        Value Std.Error   t-value p-value
th4  98.98596   1.10624  89.47954  0.0000
th21  0.36960   2.32540   0.15894  0.8739
th22 -1.87467   0.18409 -10.18359  0.0000
th23 -1.58505   0.16093  -9.84918  0.0000
th31  0.36993   0.11659   3.17282  0.0017
th32  0.89430   0.11817   7.56773  0.0000
th33  1.03261   0.13712   7.53061  0.0000
th11  4.16074  37.10361   0.11214  0.9108
th12 27.37244   2.95050   9.27722  0.0000
th13 26.56520   2.82509   9.40332  0.0000
```

Next, a likelihood ratio test is performed to determine whether the null hypothesis of all parameters being equal, i.e., the treatments having equal dose-response profiles, can be rejected or not. This can be done using the R function `anova()` in the following way:

```
anova(gnls.model001, gnls.model002)
Model           df AIC      BIC       logLik     Test    L.Ratio   p-value
gnls.model001 1  5 1627.413 1644.358  -808.7064
gnls.model002 2 11 1603.530 1640.809  -790.7648  1 vs 2  35.88319  <.0001
```

The two models differ in six degrees of freedom since the full model `gnls.model002` has six more parameters compared to the null model `gnls.model001`, namely a separate θ_2, θ_3, and θ_4 for each treatment group. The p value for the likelihood ratio test is smaller than 0.0001, indicating that there is statistical evidence to reject the hypothesis of the three treatments having an equal dose-response curve. Parameter estimates and 95% confidence intervals are obtained using the function `intervals()`. Note that the confidence intervals for θ_{21} and θ_{11}, which were found to be nonsignificant, cover the value of zero as expected.

```
> intervals(gnls.model002)
Approximate 95% confidence intervals

 Coefficients:
           lower       est.      upper
th4   96.8051420 98.9859633 101.166785
th21  -4.2146358  0.3696039   4.953844
th22  -2.2375775 -1.8746713  -1.511765
th23  -1.9023087 -1.5850504  -1.267792
th31   0.1400810  0.3699315   0.599782
th32   0.6613390  0.8943032   1.127267
th33   0.7622878  1.0326054   1.302923
th11 -68.9845584  4.1607408  77.306040
th12  21.5558871 27.3724447  33.189002
th13  20.9958865 26.5652039  32.134521
attr(,"label")
[1] "Coefficients:"

 Residual standard error:
   lower     est.    upper
 8.36291  9.16350 10.13493
```

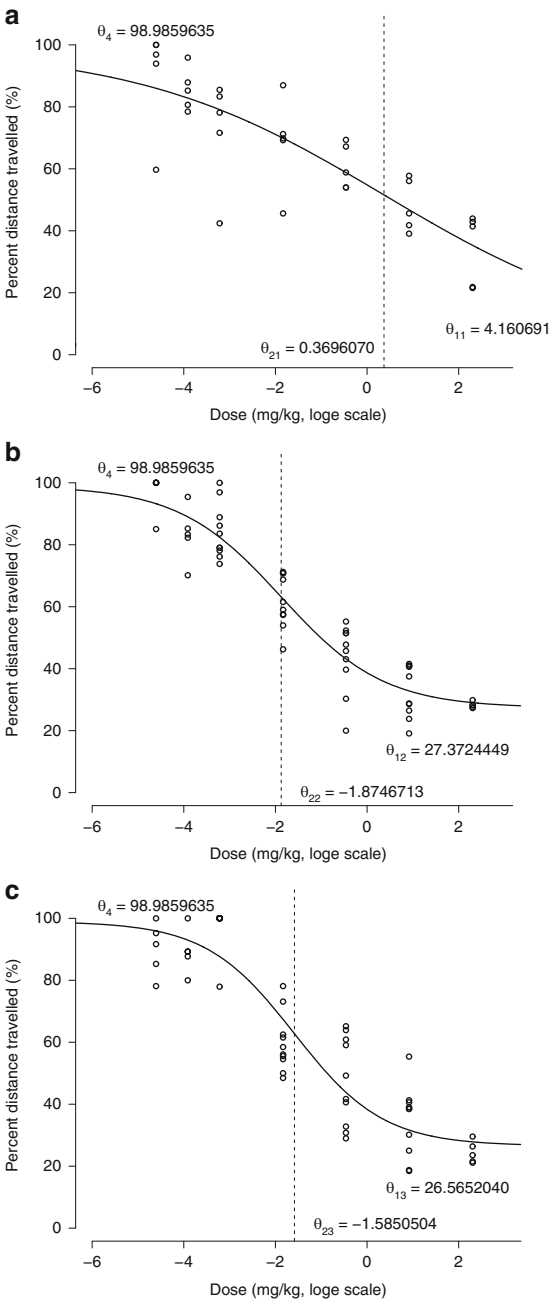

Fig. 4.6 Data and estimated 4PL models for the propulsion dataset

The broad confidence interval for θ_{21} (0.015; 141.72) obtained through exponentiating (the estimated lower and upper values in the panel above) and θ_{11} parameters are noteworthy; however, they should not be unexpected. Figure 4.4 shows that

for treatment 1, no data were available for the 40 mg/kg dose, and as a result, no information in the lower plateau of the curve is present. This lack of information is represented in the high uncertainty on the parameter estimate for θ_{11}. Note that θ_2 and θ_1 in the 4PL model tend to be correlated since the parameter θ_2 is defined as the dose for which half of the total effect is obtained where the total effect is given by the difference in the lower and upper asymptote. As a result, the missing information in the lower plateau area for treatment 1 is translated into a highly imprecise estimates for θ_{21} and θ_{11}. Although this dataset is a good example to illustrate how the 4PL dose-response model can be fitted for independent data, it also highlights a warning. These models are very flexible in nature, and care should be given to the interpretation of the obtained results.

4.3 Other Dose-Response Models for Continuous Data

4.3.1 Alternative Parameterizations of the 4PL Model

The nonlinear models formulated in (4.4) (Pinheiro and Bates 2000, p. 517) and in (4.5) (Plikaytis et al. 1991) are two examples of the 4PL model with different parametrizations. Note that for the 4PL model (4.5), the dose variable x_i is introduced on the original dose scale. Although this model will often fit the data, the disadvantage is that the ED_{50} parameter is typically log distributed. Ignoring this might lead to negative lower limits on the estimated confidence intervals.

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{1 + \mathrm{e}^{(\theta_2 - x_i)/\theta_3}} + \varepsilon_{ij}, \tag{4.4}$$

and

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{1 + (x_i/\theta_2)^{\theta_3}} + \varepsilon_{ij}. \tag{4.5}$$

The two models can be fitted using the following code:

```
#4PL model (4)
gnls.dp1<-gnls(ratio~(th1+(th4-th1)/(1+(exp((th2-lmpk )/th3)))),
    data = data2b,
    params = list(th1 + th2 + th3 + th4 ~1),
    start=c(90,-0.2,1,28),
    control=gnlsControl(nlsTol=0.1))

#4PL model (5)
gnls.dp2<-gnls(ratio~(th1+(th4-th1)/(1+((mpk/th2)^th3))),
    data = data2b,
    params = list(th1 + th2 + th3 + th4 ~1),
    start=c(99,0.2,1,28),
    control=gnlsControl(nlsTol=0.1))
```

Parameter estimates are shown in the panel below, and the fitted models are shown in Fig. 4.7. Notice how in model (4.4), the meaning of the θ_1 and θ_4 is switched due

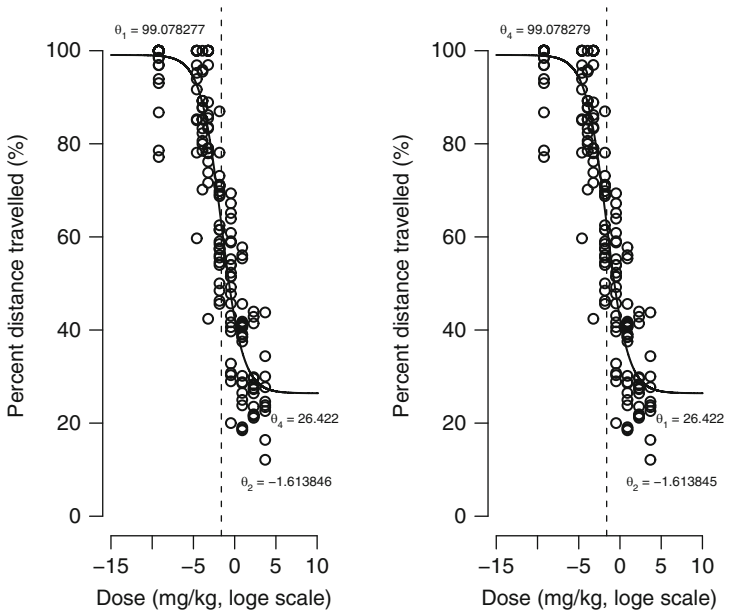

Fig. 4.7 The propulsion data and estimated 4PL models (4.4) and (4.5)

to the shift in position between x_i and θ_2. For graphical consistency, the fitted model for (4.5) in Fig. 4.7 is shown with dose on the $\log_e$ scale, while the model predicted values and θ_2 parameter are on the absolute scale. Models (4.4) and (4.5) are just two examples of alternative parametrizations, and others can be found.

```
#R output for 4PL model (4)
> summary(gnls.dp1)
Generalized nonlinear least squares fit
  Model: ratio ~ (th1 + (th4 - th1)/(1 + (exp((th2 - lmpk)/th3))))
  Data: data2b
       AIC      BIC    logLik
  1627.413 1644.358 -808.7064

Coefficients:
       Value Std.Error   t-value p-value
th1 99.07828 1.2469491  79.45655       0
th2 -1.61385 0.1534648 -10.51607       0
th3  1.27974 0.1237196  10.34384       0
th4 26.42247 2.4654370  10.71715       0

 Correlation:
    th1    th2    th3
th2 -0.237
th3  0.443  0.408
th4 -0.197 -0.728 -0.702

Standardized residuals:
       Min         Q1        Med         Q3        Max
-4.1327652 -0.5091636  0.1135235  0.5000151  2.2931540

Residual standard error: 9.806069
```

```
Degrees of freedom: 219 total; 215 residual

#R output for 4PL model(5)

> summary(gnls.dp2)
Generalized nonlinear least squares fit
  Model: ratio ~ (th1 + (th4 - th1)/(1 + ((mpk/th2)^th3)))
  Data: data2b
       AIC      BIC    logLik
  1627.413 1644.358 -808.7064

Coefficients:
       Value Std.Error  t-value p-value
th1 26.42245 2.4654394 10.71714       0
th2  0.19912 0.0305580  6.51615       0
th3  0.78141 0.0755435 10.34384       0
th4 99.07828 1.2469494 79.45654       0

 Correlation:
    th1    th2    th3
th2 -0.728
th3  0.702 -0.408
th4 -0.197 -0.237 -0.443

Standardized residuals:
       Min         Q1        Med         Q3        Max
-4.1327649 -0.5091627  0.1135234  0.5000143  2.2931538

Residual standard error: 9.806069
Degrees of freedom: 219 total; 215 residual
```

4.3.2 The 1, 2, 3PL Models

When certain parameters of the 4PL model are known to be fixed, the model can be simplified. Assume that, for an increasing curve, the asymptote at infinite dose is known and fixed to be a known constant C. In this case, the 4PL model (4.1) reduces to the 3PL (4.6) model where parameter θ_1 (if $\theta_3 > 0$) is replaced with the known constant C

$$Y_{ij} = C + \frac{\theta_4 - C}{1 + 10^{(x_i - \theta_2)/\theta_3}} + \varepsilon_{ij}\,. \tag{4.6}$$

Similarly, two or three parameters can be fixed, i.e., both the plateau values and/or the slope parameter, and the 4PL model can respectively be simplified to a 2PL or 1PL model. In the R code below, a common 3PL model (4.6) is fitted to the different treatments in the propulsion dataset with $C = 0$.

```
gnls.model3PL<-gnls(ratio~((th4)/(1+(exp((lmpk-th2)*th3)))),
                data = data2b,
                params = list(th4 + th2 + th3 ~1),
                start=c(100,-0.2,1),
                control=gnlsControl(nlsTol=0.1))
```

The parameter estimates are shown in the panel below, and the estimated model is shown in Fig. 4.8a. We notice that since we constrain the model to have a lower

asymptote equal to zero, the parameter estimate for ED_{50} changed to $\hat{\theta}_2 = -0.303$ compared with $\hat{\theta}_2 = -1.61385$ obtained in Sect. 4.2.2 for the 4PL model. The Akaike information criterion (AIC, Akaike 1973) obtained for the 3PL model (1,658.058) is higher than the AIC of the 4PL model (1,627.413), indicating that the 4PL model is to be preferred.

```
> summary(gnls.model3PL)
Generalized nonlinear least squares fit
  Model: ratio ~ ((th4)/(1 + (exp((lmpk - th2) * th3))))
  Data: data2b
       AIC      BIC    logLik
  1658.058 1671.614 -825.029

Coefficients:
        Value Std.Error  t-value p-value
th4 102.36701 1.6565090 61.79683  0.0000
th2  -0.30292 0.1344734 -2.25267  0.0253
th3   0.43089 0.0260811 16.52115  0.0000

 Correlation:
    th4    th2
th2 -0.678
th3 -0.635  0.302

Standardized residuals:
         Min           Q1          Med           Q3          Max
-3.53495283 -0.32513218 -0.01983675  0.43433966  2.67578036

Residual standard error: 10.54038
Degrees of freedom: 219 total; 216 residual
```

4.3.3 The Emax (Hill) Model

Another well-known dose-response model is the so-called Emax model, also called Hill model, given by

$$Y_{ij} = E_0 + \frac{x_i^n \text{Emax}}{x_i^n + \text{ED}_{50}^n} + \varepsilon_{ij}. \tag{4.7}$$

Here, Y_{ij} is the jth observation in dose x_i, E_0 is the base effect, corresponding to the response when the dose equals zero, Emax is the maximum drug effect, ED_{50} is the dose producing half of the Emax effect, and n is the slope parameter, also called Hill factor. A detailed description of the Emax model can be found in Chap. 9 in Ting (2006). In the R code below, a common E_{max} model 4.7 is fitted to the different treatments in the Propulsion dataset

```
gnls.modelEmax2<-gnls(ratio~E0+((Emax*mpk^n)/(ED50^n+mpk^n)),
                  data = data2b,
                  params = list(E0+Emax+ED50+n ~1),
                  start=c(1,70,0.1,-1),
                  control=gnlsControl(nlsTol=0.1))
```

The parameter estimates are shown below, and Fig. 4.8b shows the estimated model. Note that the parameter estimate for the $ED_{50} = 0.19912$ which is similar

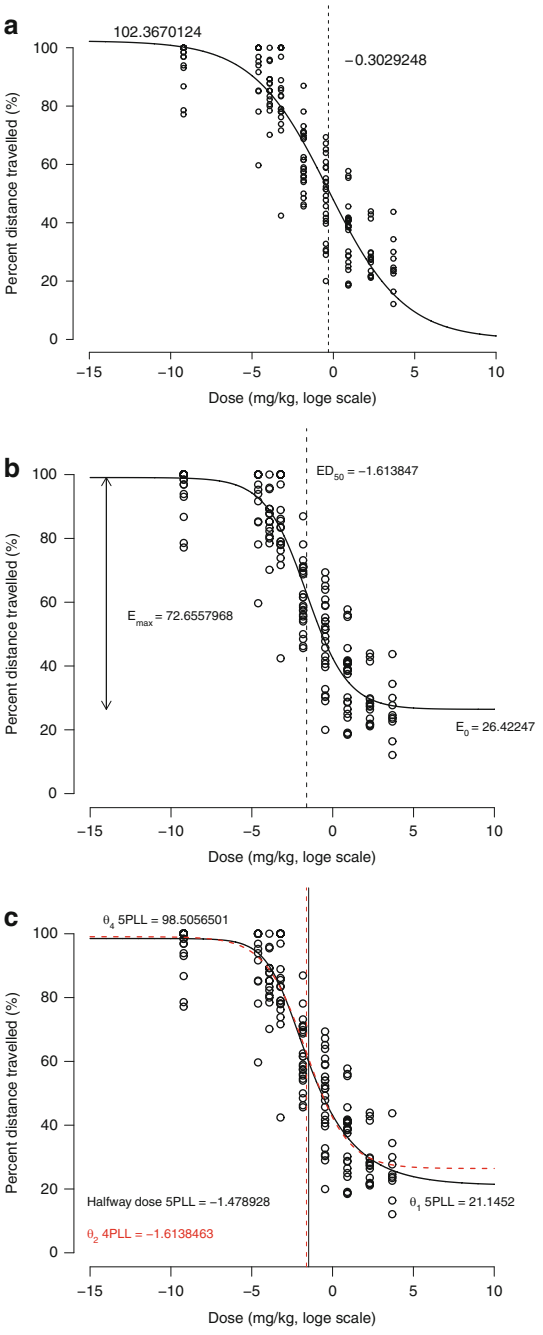

Fig. 4.8 Data and estimated models for the 3PL, the Emax, and the 5PL models. The *vertical line* represents the ED_{50} dose on log scale. (**a**) The 3PL model. (**b**) The Emax model. The *vertical line* is $\ln(ED_{50})$. (**c**) The 5PL (*solid line*) and 4PL (*dotted line*) models

to the exponent of $\hat{\theta}_2$ obtained from the 4PL model in Sect. 4.2.2, $\exp(\hat{\theta}_2) = \exp(-1.61385) = 0.19911$.

```
> summary(gnls.modelEmax2)
Generalized nonlinear least squares fit
  Model: ratio ~ E0 + ((Emax * mpk^n)/(ED50^n + mpk^n))
  Data: data2b
       AIC      BIC    logLik
  1627.413 1644.358 -808.7064

Coefficients:
        Value Std.Error    t-value p-value
E0   26.42248 2.4654354  10.717166       0
Emax 72.65580 2.9739869  24.430436       0
ED50  0.19912 0.0305579   6.516154       0
n    -0.78141 0.0755436 -10.343844       0

 Correlation:
     E0     Emax   ED50
Emax -0.912
ED50 -0.728  0.504
n    -0.702  0.767  0.408

Standardized residuals:
       Min         Q1        Med         Q3        Max
-4.1327655 -0.5091642  0.1135235  0.5000156  2.2931541

Residual standard error: 9.806069
Degrees of freedom: 219 total; 215 residual
```

4.3.4 *Nonsymmetrical Dose-Response Models*

When a dose-response curve is asymmetrical, the 4PL model is not suitable. In this section, we discuss several asymmetric alternatives to the 4PL model. In Sect. 4.3.4.1 we discuss the asymmetric logistic 5PL model and in Sect. 4.3.4.2 two asymmetric growth models, the Gompertz function and the Richards function.

4.3.4.1 Asymmetric Logistic Model (5PL)

A first asymmetric model is obtained by extending the 4PL with a fifth parameter to describe asymmetry. The 5PL model is given by:

$$Y_{ij} = \theta_1 + \frac{\theta_4 - \theta_1}{\left[1 + 10^{(x_i - \theta_2)\theta_3}\right]^{\theta_5}} + \varepsilon_{ij}. \tag{4.8}$$

As before, Y_{ij} is the jth observation on dose x_i, expressed as decadic logarithm (base 10 value), and the parameters θ_1 and θ_4 are the response values in the asymptotic region of the curve. In contrast with the 4PL model, θ_2 is not the dose for which half the effect (difference between the two asymptotes) is obtained,

and θ_3 is no longer a simple slope parameter as explained in Sect. 4.2.2. For an elaborate discussion about the 5PL model and its comparison with the 4PL model including a detailed description on how each parameter effects the sigmoidal pattern, we refer to Gottschalk and Dunn (2005). In this section, we briefly discuss the explicit expression for the dose (on the log10 scale) corresponding to the response halfway between the two asymptotes. Note that, as we mentioned above, it is no longer a parameter in the model but a function of the model parameters given by

$$\mathrm{ED}_{50} = \theta_2 + \frac{1}{\theta_3} \log_{10} \left(2^{\frac{1}{\theta_5}} - 1 \right). \tag{4.9}$$

In the R code below, a common 5PL model is fitted to the different treatments in the propulsion dataset. To be consistent with the applied 4PL model in Sect. 4.2.2, the 5PL model is specified as the natural logarithm (base e value) parametrization.

```
gnls.model5PL002<-gnls(ratio~(th1+(th4-th1)/((1+(exp((lmpk-th2)*th3)))^th5)),
                  data = data2b,
                  params = list(th1 + th2 + th3 + th4 + th5 ~1),
                  start=c(90,-1.6,1,28,1),
                  control=gnlsControl(nlsTol=0.1))
```

The parameter estimates are shown in the panel below, and the estimated model is shown in Fig. 4.8c. The dose corresponding to the response halfway between the two asymptotes was calculated according to (4.9), $\widehat{\mathrm{ED}}_{50} = -1.478$, slightly higher than $\hat{\theta}_2$ obtained for the 4PL model (-1.613).

```
> summary(gnls.model5PL002)
Generalized nonlinear least squares fit
  Model: ratio ~ (th1 + (th4 - th1)/((1 + (exp((lmpk - th2) * th3)))^th5))
  Data: data2b
       AIC      BIC    logLik
  1627.617 1647.951 -807.8085

Coefficients:
       Value Std.Error  t-value p-value
th1 21.14527  6.954185  3.04065  0.0027
th2 -2.95604  0.847753 -3.48691  0.0006
th3  1.07869  0.334077  3.22887  0.0014
th4 98.50565  1.237730 79.58576  0.0000
th5  0.38977  0.283665  1.37404  0.1709

 Correlation:
    th1    th2    th3    th4
th2  0.831
th3 -0.650 -0.909
th4  0.112  0.155 -0.355
th5  0.863  0.986 -0.931  0.248

Standardized residuals:
       Min         Q1        Med         Q3        Max
-4.1760032 -0.5382668  0.1562766  0.4685797  2.2711929

Residual standard error: 9.788737
Degrees of freedom: 219 total; 214 residual
```

4.3.4.2 Growth Models

Other examples of asymmetric alternatives to the 4PL model are members of the so-called growth models, often used in biology (Lindsey 2001; Narinc et al. 2010), such as the Gompertz function and Richards function. Both models were developed for growth data where the phases of growth are asymmetrical. As in the 5PL model (4.8), these models do not have an ED_{50} parameter directly included. These functions can however be applied to asymmetrical dose-response curves, and the ED_{50} parameter can be explicitly calculated (if wanted) through inverse prediction.

For illustrational purposes, we fit the models to the propulsion dataset. Since growth models are typically used to model increasing responses with increasing time, the original response variable, percent distance traveled (%), in the propulsion dataset will be transformed to inhibition as 100 ratio. Notice that the Gompertz and Richards models are not good choices for this dataset since they both assume a lower asymptote equal to zero, which is clearly not the case.

Gompertz Function

A Gompertz curve or Gompertz function is an asymmetrical sigmoidal function, and it originates from Benjamin Gompertz work on human mortality (Gompertz 1825). In biology, it is commonly used to model growth processes where the period of increasing growth is shorter than the period in which growth decreases. The model is given by

$$Y_{ij} = a \times \exp^{-b \times \exp(-c \times t_i)} + \varepsilon_{ij}, \tag{4.10}$$

where Y_{ij} is the jth observation at time t_i and the parameters a, b, and c are constrained to be positive. Note that Y_{ij} equals zero as time approaches $-\infty$ and a is the upper asymptote when time approaches $+\infty$. An example for a Gompertz curve is graphically depicted in Fig. 4.9a with $a = 100$, $b = 2$, and $c = 1$. The inflection point, or the point where the curve changes from concave upward to concave downward, equals to $\ln(b)/c$ and the corresponding response value Y_{ij} is a/e (with e =Eulerian number). Notice that due to the asymmetric nature of the Gompertz curve, the inflection point lies in the lower part of the curve is in contrast with the 4PL model where the inflection point (θ_2) lies in the middle of the curve. The following R code was used to fit the Gompertz model.

```
## Gompertz
gnls.modelG001b<-gnls(inhibition~(a*exp((-b)*exp((-c)*lmpk))),
                      data = data2b,
                      params = list(a + b + c ~1),
                      start=c(100,1,1),
                      control=gnlsControl(nlsTol=0.1))

## parameter estimates
Coefficients:
     Value Std.Error   t-value p-value
a 80.10299  3.863470 20.733432       0
b  0.35444  0.059149  5.992368       0
c  0.45215  0.047345  9.550109       0
```

Figure 4.9b shows the Gompertz model estimated for the propulsion dataset.

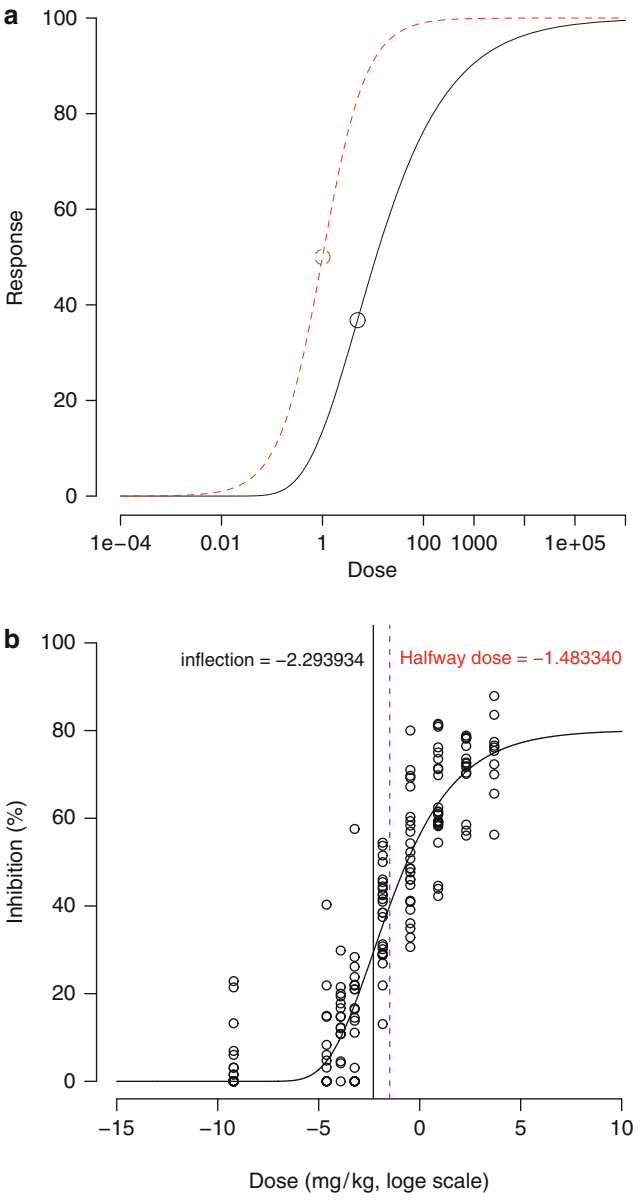

Fig. 4.9 Gompertz model. (**a**) Gompertz model (*solid line*) and 4PL model (*dotted line*). Parameter setting for the 4PL model: $\theta_1 = 100, \theta_2 = 0, \theta_3 = 1$, and $\theta_4 = 0$. The *open circles* are inflection points of the model. (**b**) Gompertz model fitted to the propulsion dataset. The *vertical solid line* represents the doses at the inflection point, and the *dotted vertical line* represents the dose corresponding to half the response between the two asymptotes

Richards Function

The Richards function (Richards 1959) is an asymmetrical sigmoidal growth model given by

$$Y_{ij} = a\left(1 - b\,\mathrm{e}^{-k x_i}\right)^{\frac{1}{1-m}} + \varepsilon_{ij}. \tag{4.11}$$

The Richards function is known for its flexibility as a result of the shape parameter m. In contrast to the 4PL, 5PL, and Gompertz curve (and other nonlinear curves not discussed here), the Richards function has an optional point of inflection which is not in a fixed way related to the asymptote, but its existence and location depend on the value of parameter m. The dose at which the point of inflection is located equals to

$$\frac{1}{k}\ln\left(\frac{b}{1-m}\right)$$

and the response value at the point of inflection equals to

$$a m^{\frac{1}{1-m}}.$$

Figure 4.10a is an illustration of different Richards curves, all with common parameter values with the exception of the parameter m. The circles on the graph indicate the inflection points, and it clearly shows how with increasing m value the inflection point moves from the lower part of the curve to the upper part of the curve.

A common Richards model (4.11) fitted to the different treatments in the propulsion dataset can be implemented in R in the following way:

```
# Richards model
gnls.modelR001<-gnls(inhibition~a*(1-b*exp(-1*k*lmpk))^(1/(1-m)),
                     data = data2b,
                     params = list(a + b + k + m ~1),
                     start=c(78,-0.156,0.55,1.44),
                     control=gnlsControl(nlsTol=0.1))

# Parameter estimates
Coefficients:
     Value Std.Error   t-value p-value
a 75.51039  4.195788 17.996714  0.0000
b -0.19126  0.178131 -1.073730  0.2841
k  0.64348  0.205562  3.130361  0.0020
m  1.62286  0.640525  2.533637  0.0120
```

Figure 4.10b shows the estimated model. The inflection point is a function of the model parameters and can be calculated by

```
> inflec<-(1/(coef(gnls.modelR001)[3]))*
          log(coef(gnls.modelR001)[2]/(1-coef(gnls.modelR001)[4]))
> inflec
        k
-1.834793
```

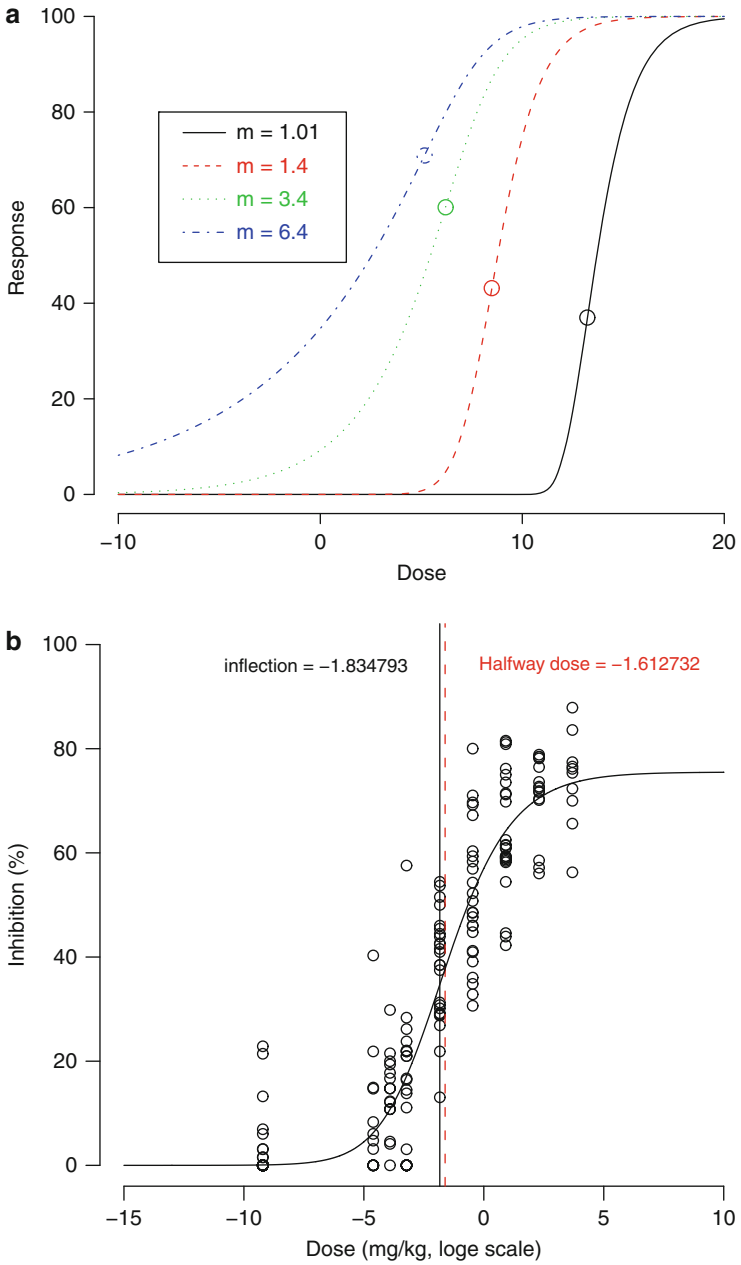

Fig. 4.10 Richards function. (**a**) Richards functions with several values of m. (**b**) Richards function fitted to the propulsion data

4.4 Discussion

Understanding the exposure response relationship of a drug is essential for success in pharmaceutical development. In the clinical development phase, an adequate dose-response model helps in identifying the dose most likely to be successful in the confirmatory phase III clinical trial. In preclinical development, dose-response modeling is a useful tool in ranking multiple compounds with regards to their possible potential and thus is helpful in choosing the most promising compounds to take further into development.

Although dose-response modeling is a very general concept, it often comes down to finding a mathematical model that adequately describes the relationship between dose and response of interest over a continuous range of doses, based on the empirical results of some specific doses used in experiments. A common dose-response pattern involves a monotone symmetrical increase or decrease of response in function of increasing doses where the profile goes from a lower or upper response plateau at zero dose to an upper or lower response plateau at infinite doses in a sigmoidal way. The 4PL model discussed in this chapter is perfectly capable of describing this pattern and allows the estimation of doses or response levels of interest outside the discrete cases studies. Moreover, the model succeeds in summarizing the dose-response pattern using four (or less if appropriate) easily interpretable parameters, two parameters representing the lower and upper plateaus, one slope parameter describing the steepness of the curve, and one parameter which identifies the dose corresponding to the response halfway between the asymptotes. This last parameter is generally known as the ED_{50} (in case of efficacy response) or potency parameter.

While the 4PL model is a commonly used exposure response model to capture monotone symmetrical sigmoidal patterns, it certainly is not the only one available. Some alternatives have been discussed such as different parametrizations, the 3,2, and 1PL model and the Emax model. In case the exposure response pattern would exhibit nonsymmetrical characteristics, the 5PL model, an extension of the 4PL model, as well as some growth models is available.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov, & B. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Academiai Kiado.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London, 115*, 513–585.

Gottschalk, P. G., & Dunn, J. R. (2005). The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Analytical Biochemistry, 343*, 54–65.

Lindsey, J. (2001). *Nonlinear Models in Medical Statistics*. Oxford, Oxford University Press.

Narinc, D., Karaman, E., Firat, M. Z. F., & Aksoy, T. (2010). Comparison of non-linear growth models to describe the growth in Japanese Quail. *Journal of Animal and Veterinary Advances, 9*(14), 1961–1966.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York: Springer.

Plikaytis, B. D., Turner, S. H., Gheesling, L. L., & Carlone, G. M. (1991). Comparisons of standard curve-fitting methods to quantitate Neisseria meningitides group a polysaccharide antibody levels by enzyme-linked immunosorbent assay. *Journal of Clinical Microbiology, 29*(7), 1439–1446.

Richards, F. J. A. (1959). Flexible growth function for empirical use. *Journal of Experimental Botany, 10*(29), 290–300.

Straetemans, R., & Bijnens, L. (2010). Application of the separate ray model to investigate interaction effects. *Frontiers in Bioscience, E2*, 266–278.

Tanila, H., Kauppila, T., & Tana, T. (1993). Inhibition of intestinal motility and reversal of postlaparotomy ileus by selective 2-adrenergic drugs in the rat. *Gastroenterology, 104*, 819–824.

Ting, N. (Ed.). (2006). *Dose finding in drug development*. New York: Springer.

Part II
Dose–Response Microarray Experiments

Chapter 5
Functional Genomic Dose-Response Experiments

Luc Bijnens, Hinrich W.H. Göhlmann, Dan Lin, Willem Talloen, Tim Perrera, Ilse Van Den Wyngaert, Filip De Ridder, An De Bondt, and Pieter Peeters

5.1 Introduction

In the first part of the book, we discussed different aspects of the analysis of dose-response data such as estimation, inference, and modeling. In this part of the book, we focus on dose-response microarray experiments. Within the microarray setting, a dose-response experiment has the same structure as described in part I of the book. The response is the gene expression at a certain dose level. The role of functional genomics, particularly in this setting, is to find indications of both safety and efficacy before the drug is administrated to patients. Studies in human cell lines or in rodents are often used for that purpose. They are usually the next step after the high-throughput experimentation that identifies and validates the biological targets. Preclinical experiments include both discovery and toxicology assays. They are usually carried out before or during the clinical programs.

For a single gene, the data structure is similar to the one we discussed in the first part of the book: a response vector (gene expression) which is measured in a sequence of (increasing) dose levels. Let $\boldsymbol{Y}$ be an expression matrix containing information about the gene expression of m genes in n samples (arrays) measured at $K + 1$ dose levels. Figure 5.1 shows an illustration of the data structure. Here, Y_{ijk} is the expression level of the jth sample (array) of the kth gene measured at dose level i and Y_{0jk} is the gene expression measured at dose zero. Note that at each

L. Bijnens (✉) · H.W.H. Göhlmann · W. Talloen · T. Perrera · I. Van Den Wyngaert · F. De Ridder · An De Bondt · P. Peeters
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: LBIJNENS@its.jnj.com; HGOEHLMA@its.jnj.com; WTALLOEN@its.jnj.com; TPERERA@its.jnj.com; IVDWYNGA@its.jnj.com; FDRIDDER@its.jnj.com; ADBONDT@its.jnj.com; PPEETER3@its.jnj.com

D. Lin
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_5,

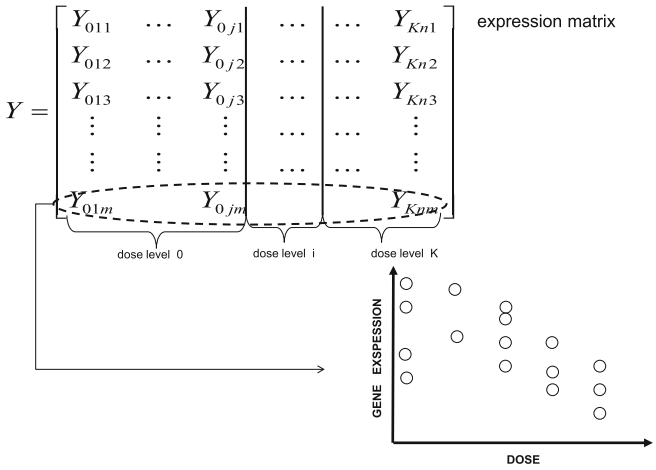

Fig. 5.1 Data structure in dose-response microarray experiments

dose level, there are r_j replicates available for each gene. Hence, each row in the expression matrix consists of a dose-response experiment for a specific gene.

The main goal of the analysis is to detect trends in gene expression caused by increasing doses of compounds. The aim of these microarray experiments is to get insight into the mechanism of action and the safety profile using functional genomics data to identify pathways that are affected by the compound at hand. In this context, gene expression experiments have become important either before or parallel to the clinical testing programs. Dose-response relationships either upregulated or downregulated may give insight in the biological target and the mechanism that the new medicine uses to treat the disease. At the same time, it may generate data on the pathways involved in potential unwanted side effects. When side effects are detected early on in the development, the molecules could potentially be modified chemically so that it no longer has that effect. Dose finding using microarray experiments can also be of importance in the search for biomarkers that could be used as signatures for response (Göehlmann and Talloen 2009). These signatures could then be used to find the target population in which the drug has the best therapeutic effect. Another use is the identification of biomarkers that eventually could be used in phase II or phase III clinical trials as biomarkers endpoints to replace the classical endpoints.

5.2 Order-Restricted Analysis of Microarray Data

Ruberg (1995a,b) and Chuang-Stein and Agresti (1997) formulated four main questions usually asked in dose-response studies: (1) Is there any evidence of the drug effect? (2) For which doses is the response different from the response in the control group? (3) What is the nature of the dose-response relationship? and (4) What is the optimal dose?

5.2.1 Monotonicity and the Simple Order Alternative

Throughout the second part of the book, we address the above questions in several ways. In Chaps. 2 and 3, we discussed the setting of a *simple order*, i.e., $\mu(d_0) \leq \mu(d_1) \leq, \ldots, \leq \mu(d_K)$. In this setting, for a single gene, the first question is related to hypothesis testing of the null hypothesis of no dose effect versus an order alternative. Figure 5.2a shows an example of two genes with and without significant dose effect. The aim of the first step of the analysis is to identify those genes with a significant (and monotone) relationship with doses. Inference on the dose-response relationship under order constraints by using various approaches, i.e., likelihood ratio test and other t-type tests, is discussed in Chap. 7, the significance analysis of microarrays (SAM) in Chap. 9, and the Bayesian approach in Chap. 13. Figure 5.2b illustrates the problem related to the second question. Both genes presented in Fig. 5.2b have a significant monotone relationship with dose. However, the dose in which the gene expression is different from the control dose is not the same. This topic is discussed in details in Chaps. 15–17 where we discuss the topics of the ratio test between means of highest dose and dose zero using the multiple contrast test (MCT), the multiple contrast ratio test, and the FDR-adjusted CIs for the ratio parameters. Figure 5.2c illustrates the problem related to the third question. Two parametric models (or isotonic regression can be used as well), representing different dose-response curve shapes, are fitted to the expression data of two genes, and the problem is to select the best gene-specific model. We employ an exploratory tool, namely, a clustering algorithm to find genes with similar dose-response patterns discussed in Chap. 9, several information criteria to explicitly select the best from a set of fitted possible models discussed in Chap. 10, as well as the parametric modeling of dose-response relationship to obtain ED_{50} the parameter of interest in Chap. 14. Finally, Chap. 12 is devoted to gene set analysis which can be useful for the biological interpretation of the results obtained using the approaches discussed in the book.

5.2.2 Beyond Monotonicity: the Simple Tree and Partial Ordered Alternatives

5.2.2.1 The Simple Tree Alternative

The *simple tree* alternative can be used to test several doses with the control. Within this setting, we consider one-sided comparisons of all dose levels with the control but do not specify any order relationship among the mean gene expression at higher dose levels, i.e., $\mu(d_0) \leq [\mu(d_1), \mu(d_2), \ldots, \mu(d_K)]$ (Robertson et al. 1988). In the case that this is the primary aim of the analysis, Dunnett's test can be used (Dunnett 1955; Robertson et al. 1988) for the one-way ANOVA model specified in (3.5). The simple tree ordering is discussed in Chap. 15 in the context of MCTs.

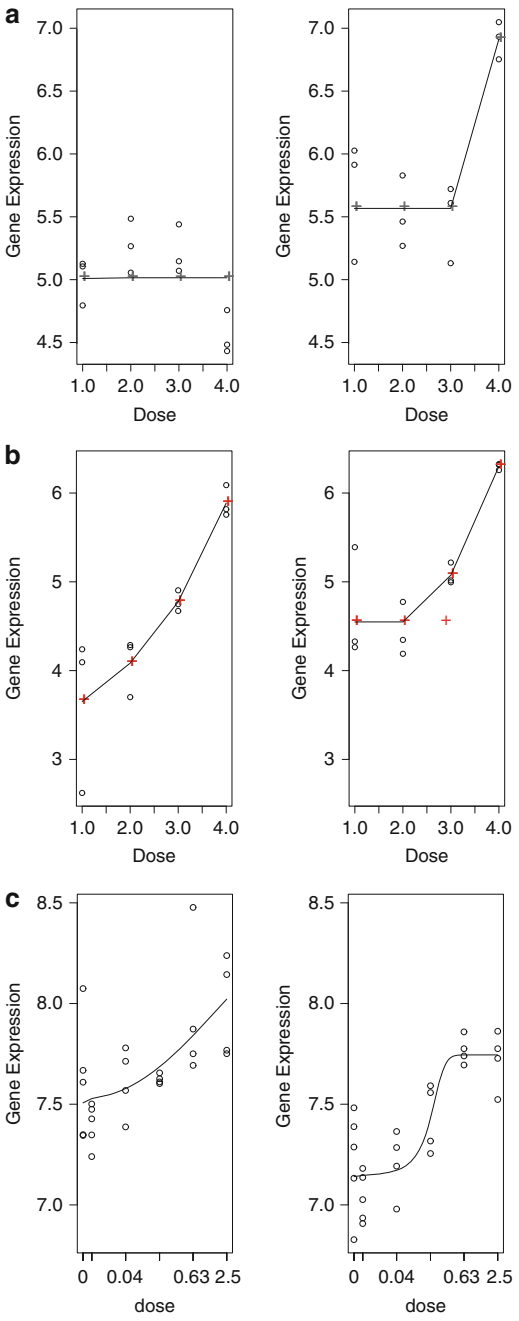

Fig. 5.2 Illustration of dose-response relationship. (**a**) Examples of two genes with fitted isotonic regression means. *Left panel*: a gene for which the dose-response relationship is not significant. *Right panel*: a gene for which the dose-response relationship is significant. (**b**) Examples of two genes with fitted isotonic regression means. *Left panel*: a gene whose expression of control dose is different than the second dose onward. *Right panel*: a gene whose expression of control dose is different than the third dose onward. (**c**) Examples of two genes with fitted parametric models

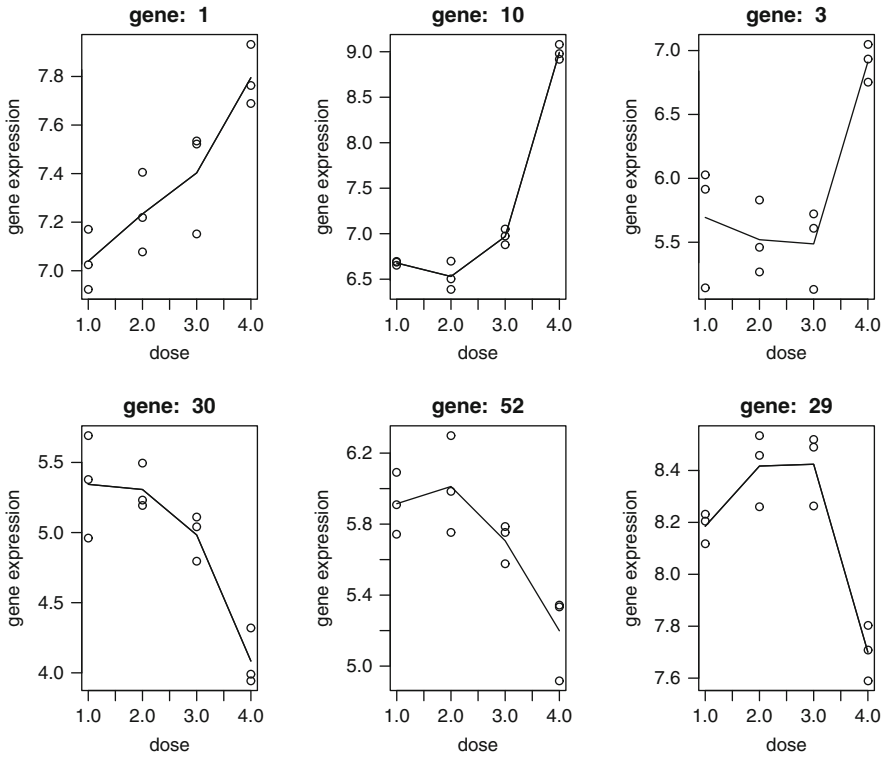

Fig. 5.3 Examples of genes with unimodal partial order (umbrella) mean profile. *First row*: down–up profiles. *Second row*: up–down profiles

5.2.2.2 The Unimodal Partial Ordering

The *unimodal partial ordering (umbrella profile)* is an ordering which imposes the following restrictions $\mu(d_0) \leq \mu(d_1) \leq, \ldots, \leq \mu(d_h) \geq \mu(d_{h+1}) \geq \cdots \geq \mu(d_K)$. Genes for which the mean profile satisfies a unimodal partial ordering restrictions (up–down profile) have a mean profile which increases up to dose level h and thereafter decrease (Robertson et al. 1988; Peddada et al. 2003). An example for genes for which the mean profile follows a unimodal partial ordering (down–up and up–down) is shown in Fig. 5.3.

Gene selection method based on the SAM and clustering of the dose-response curve shapes within this setting are discussed in Chap. 11. We discuss two different approaches. The first, ORIOGEN (Peddada et al. 2003), is a resampling-based method which can be used for both inference and clustering. The second, ORICC (Liu et al. 2009a), is a method based on model selection which can be used for clustering. The umbrella alternative is discussed, in the context of MCTs, in Chap. 15.

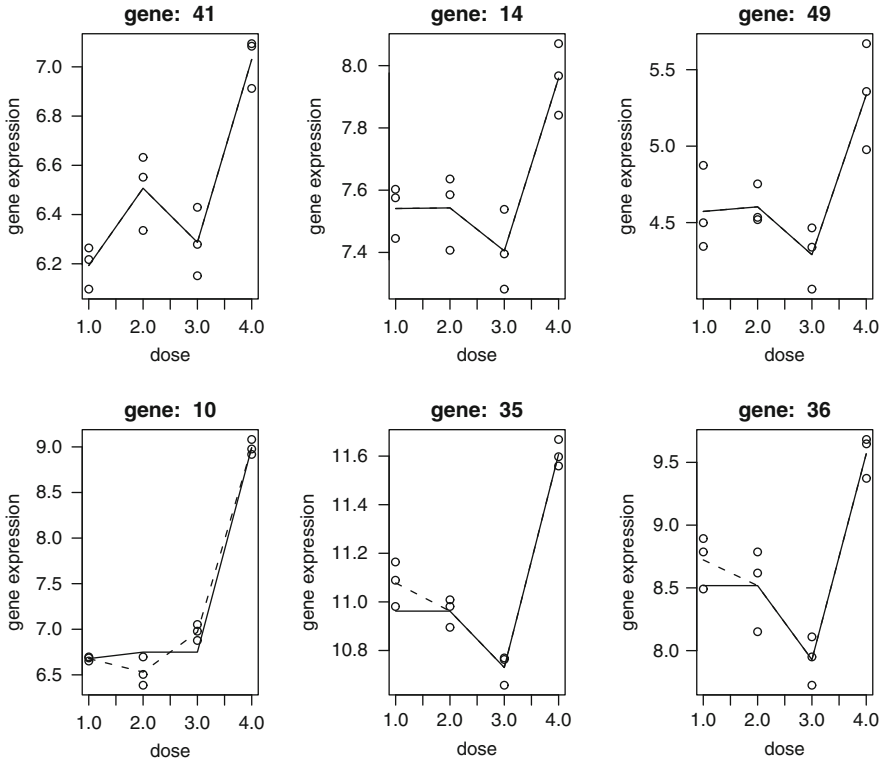

Fig. 5.4 Examples of genes cyclical profiles for $K + 1 = 4$. *First row*: the observed mean gene expression is in cyclical order. *Second row*: observed mean (*dashed line*) and estimated mean (*solid line*) under cyclical ordered alternative

5.2.2.3 The Cyclical Profile

A *cyclical* mean profile (Simmons and Peddada 2007) implies that the mean gene expression has minima and maxima within the dose range. For a dose-response experiment with four dose levels, there is only one possible cyclical model $\mu(d_0) \leq \mu(d_1) \geq \mu(d_2) \leq \mu(d_3)$. For this cyclical mean profile, the mean gene expression turns down at dose level 2 and turns up at dose level 3. Example of genes with cyclical mean profiles for $K + 1 = 4$ are shown in Fig. 5.4. Hypothetical examples of cyclical mean profiles for dose-response experiments with four, five, and six dose levels are shown in Fig. 5.5. Note that for early drug development dose-response experiments, the cyclical mean profile is often not of interest, and we will not discuss this setting further in the book.

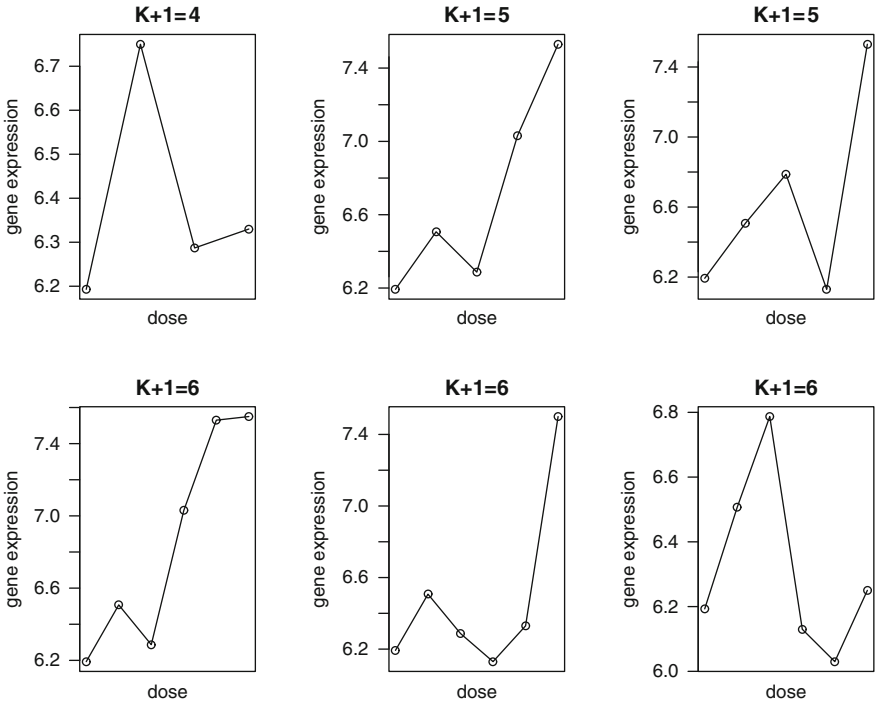

Fig. 5.5 Hypothetical examples of cyclical mean profiles for four, five, and six dose levels

5.3 Case Studies

5.3.1 Example 1: Human Epidermal Squamous Carcinoma Cell Line A431 Experiment

The data come from an oncology experiment designed to better understand the biological effects of growth factors in human tumor. Human epidermal squamous carcinoma cell line A431 (HESCA431) was grown in Dulbecco's modified Eagle's medium, supplemented with L-glutamine (20 mM), gentamycin (5 mg/ml), and 10% fetal bovine serum. The cells were stimulated with growth factor EGF (R&D Systems, 236-EG) at different concentrations (0, 1, 10, and 100 ng/ml) for 24 h. RNA was harvested using RLT buffer (Qiagen). All microarray-related steps including the amplification of total RNAs, labeling, hybridization, and scanning were carried out, as described in the GeneChip Expression Analysis Technical Manual, Rev.4 (Affymetrix 2004). Biotin-labeled target samples were hybridized to human genome arrays U133 A 2.0 containing probe sets interrogation approximately 22,000 transcripts from the UniGene database (Build 133). Hybridization was performed using

15 μg of cRNA for 16 h at 45°C under continuous rotation at 60 rpm. Arrays were stained in Affymetrix Fluidics stations using streptavidin/phycoerythrin staining. Thereafter, arrays were scanned with the Affymetrix scanner 3000, and images were analyzed using the GeneChip Operating System v1.1 (GCOS, Affymetrix). The collected data were quantile normalized in two steps: first within each sample group, and then across all sample groups obtained (Amaratunga and Cabrera 2000, 2004; Bolstad et al. 2002). The resulting dataset consists of 12 samples, 4 dose levels, and 3 microarrays at each dose level, with 16,998 probe sets. For simplicity, we refer probe sets as genes through our book (Hubbell et al. 2002).

5.3.1.1 Reading the Data

A data frame with the $\log_2$ transformed gene intensities is loaded into the R environment. The first ten genes and first six samples are displayed in the table below. The row names of the genes show the probe ID, `X1, X1.1, and X1.2` are the three arrays for dose zero, while `X2, X2.1, and X2.2` are the arrays for the first dose. The data frame is loaded using the function `load()`:

```
> load("data.Rdata")
```

A printout of the first ten lines is given below.

```
> data[1:10,1:6]
                X1     X1.1     X1.2       X2     X2.1     X2.2
31637_s_at 6.923109 7.024719 7.170328 7.219297 7.076908 7.404949
32402_s_at 5.107275 5.092935 5.255918 5.312913 4.893855 4.596591
33646_g_at 5.913526 6.026197 5.141728 5.828770 5.269202 5.461664
34063_at   4.919469 4.908159 3.500307 4.814068 4.139949 4.278321
33494_at   6.002091 5.878718 5.777668 6.214799 5.895586 6.163291
34031_i_at 7.162715 7.294693 6.903935 7.223069 6.972928 7.412160
34449_at   4.049696 4.748409 3.845498 4.780287 4.076589 4.300242
34478_at   3.191931 4.326571 3.771206 3.570291 2.179324 3.988911
35436_at   6.487708 6.285804 6.229814 6.109103 6.340837 5.931840
36711_at   6.695870 6.687039 6.652153 6.503670 6.387794 6.698711
```

5.3.1.2 Exploring the Data

`IsoPlot()` is the plotting function that can be used to explore the data. Figure 5.6 shows a scatter plot for the second gene in the dataset (`data[2,]`), and it can be produced by

```
> x <- c(rep(1,3), rep(2,3), rep(3,3), rep(4,3))
> par(mfrow=c(1,2))
> IsoPlot(x, y=data[2,])
> IsoPlot(x, y=data[2,],add.curve = TRUE)
```

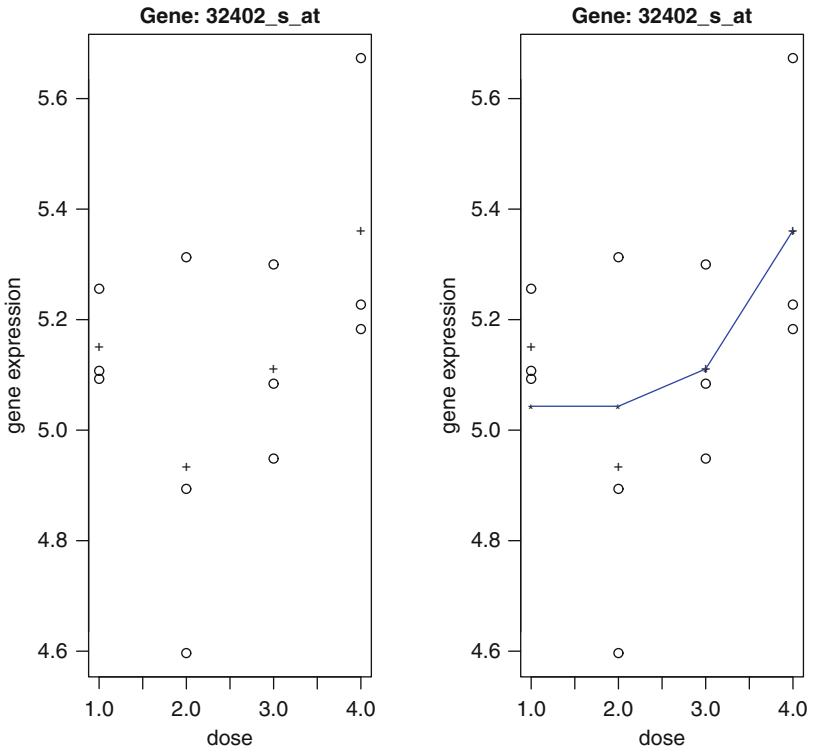

Fig. 5.6 The data points are plotted as *circles*, while sample means as pluses. The *right panel* additionally plots the fitted increasing isotonic regression model (*solid line*)

5.3.2 Example 2: The Pharmacological Activity of Antipsychotics (Extrapyramidal Symptoms)

The data consists of a preclinical evaluation study of six antipsychotic compounds (JnJa, JnJb, JnJc, CompA, CompB, and CompC). In this study, the alterations downstream of the dopamine D2 receptor were investigated. The pharmacological activity of antipsychotics is often based on blocking this receptor. However, it is also this blocking that can cause side effects such as extrapyramidal symptoms (EPS) and hyperprolactinemia (elevated prolactin level in serum) in humans. Two measures that can be used to estimate compound activity are antagonism of apomorphine-induced behavior and receptor occupancy in the brain. An animal model that can be used to study human EPS is the sensitivity to catalepsy in rats. Five compounds (CompB, CompC, JnJa, JnJb, and JnJc) had six dose levels of 0, 0.01, 0.04, 0.16, 0.63, and 2.5 mg/kg, while CompA had six dose levels of 0, 0.16, 0.63, 2.5, and 40 mg/kg. Each dose was given to four to six independent rats. These dose levels

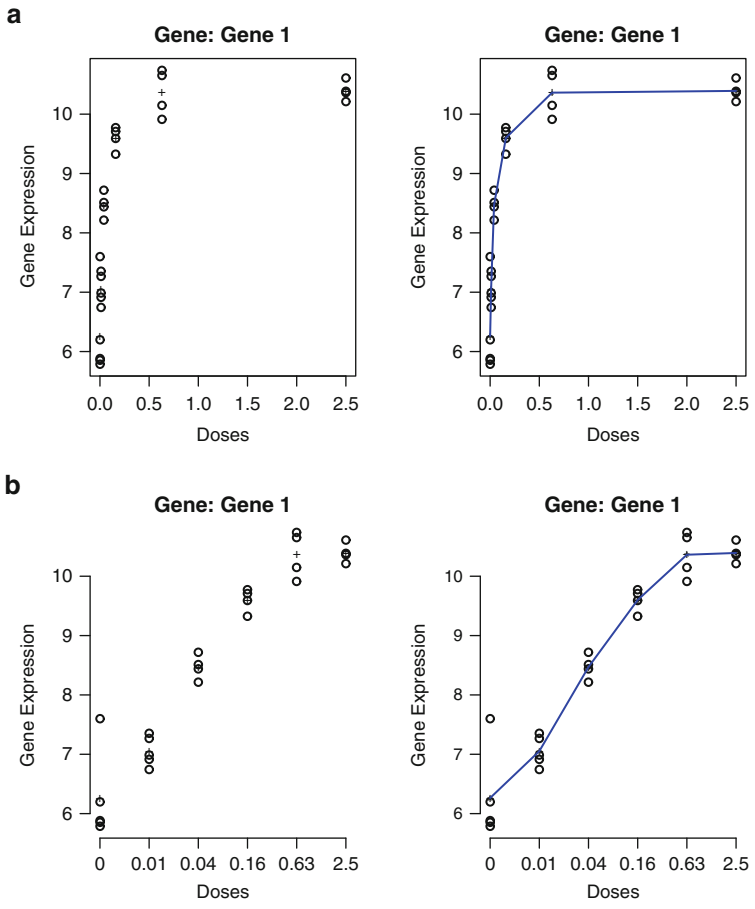

Fig. 5.7 Exploratory plots with `IsoPlot()`. (**a**) The plots produced by `IsoPlot` with original data (*left panel*) and with the isotonic regression model fitted to the data (*right panel*, `add.curve=TRUE`) where dose is treated as a continuous variable. The data points are plotted as *circles*, with sample means as pluses and the fitted isotonic regression model (*solid line*) for the *right panel plot*. (**b**) The plots produced by `IsoPlot` with original data (*left panel*) and with the isotonic regression model fitted to the data (*right panel*) where dose is treated as an ordinal variable (`type = "ordinal"`). The real dose level is presented on the *x*-axis

were based on the ED_{50} value as determined in an independent apomorphine test. Apomorphine is a direct dopamine receptor stimulant which mimics the agonistic action of dopamine on the D2 receptor. The resulting behavioral effects can be antagonized by blockade of central D2 receptors. The efficiency of this antagonism can be a measure for the potency of a compound.

After an acute treatment of 1 hour a specific, relevant brain region (the striatum) was isolated. Then gene expression profiling on all samples was performed. After a quality control check, the dataset consists of four to five arrays at each dose level.

The experiment was performed using the Affymetrix whole human genome array. Each chip consists of 11,562 probe sets. For simplicity, we refer the probe sets as genes. The biological question of this experiment is to know which genes are affected by downstream of the dopamine D2 receptor.

The example data explained above are stored in a workspace "`data.RData`" which contains two objects: `dose` and `express` (containing gene intensities). To load these objects, we can use the function `load`:

```
> load("data.RData")
> dose
 [1] 0.00 0.00 0.01 0.01 0.04 0.04 0.16 0.16 0.63 0.63 2.50 2.50 0.00 0.00
[15] 0.00  0.01 0.01 0.01 0.04 0.04 0.16 0.16 0.63 0.63 2.50 2.50
```

The data contain information about 11,562 genes and 26 arrays:

```
> dim(express)
[1] 11562    26
```

As illustrated in the previous example, we use the `IsoPlot` function to produce the dose-response scatterplots. The function has the options to specify the dose levels by using `type` = "`ordinal`" or "`continuous`" (default). The scatterplots for the first gene in the dataset (`express[1,]`), shown in Fig. 5.7, can be produced by:

```
> IsoPlot(dose,express[1,],type="continuous")
> IsoPlot(dose,express[1,],type="continuous",add.curve = TRUE)

> IsoPlot(dose,express[1,],type="ordinal")
> IsoPlot(dose,express[1,],type="ordinal", add.curve=TRUE)
```

References

Affymetrix GeneChip. (2004). *Expression analysis technical manual, Rev.4*. Santa Clara: Affymetrix.

Amaratunga, D. & Cabrera, J. (2004). Exploration and Analysis of DNA Microarray and Protein Array Data, Hoboken, NJ, Wiley-Interscience-John Wiley and Sons, Inc.

Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2002). A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics, 19*, 185–193.

Chuang-Stein, C., & Agresti, A. (1997). Tutorial in biostatistics: A review of tests for detecting a monotone dose-response relationship with ordinal response data. *Statistics in Medicine, 16*, 2599–2618.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *JASA, 50*, 1096–1121.

Goehlmann, H., & Talloen, W. (2009). *Gene expression studies using Affymetrix microarrays*. Boca Raton: Chapman & Hall/CRC.

Hubbell, E., Liu, W. M., & Mei. R. (2002) Robust estimators for expression analysis. *Bioinformatics, 18*(12), 1585–1592.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009a). *Order-restricted information criterion-based clustering algorithm. Reference manual*. `http://cran.r-project.org/web/packages/ORIClust/`.

Peddada, S., Lobenhofer, E. K., Li, L., Afshari, C. A., Weinberg, C. R., & Umbach, D. M. (2003). Gene selection and clustering for time-course and dose-response microarray experimants using order-restricted inference. *Bioinformatics, 19*(7), 834–841.

Simmons, S. J., & Peddada, S. (2007). Order-restricted inference for ordered gene expresion (ORIOGEN) data under heteroscedastic variances. *Bioinformation, 1*(10), 414–419.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Ruberg, S. J. (1995a). Dose response studies. I. Some design considerations. *Journal of Biopharmaceutical Statistics, 5*(1), 1–14.

Ruberg, S. J. (1995b) Dose response studies. II. Analysis and interpretation. *Journal of Biopharmaceutical Statistics, 5*(1), 15–42.

Chapter 6
Adjustment for Multiplicity

Daniel Yekutieli, Dan Lin, Ziv Shkedy, and Dhammika Amaratunga

6.1 Introduction

The number of comparisons of gene expression level studied in a microarray experiment has been growing literarily at an exponential rate since the beginning of the 1990s. Considering a microarray dataset analyzed by testing each gene, multiple testing is an immediate concern. When many hypotheses are tested, the probability that a type I error is committed increases sharply with the number of hypotheses. This problem of multiplicity is not unique to microarray technology, yet its magnitude here, where a single experiment may involve many thousands of genes, dramatically intensifies the problem.

Dudoit et al. (2002) is one of the first studies to recognize the importance of the multiplicity problem as one of the key statistical issues arising in microarray data analysis. The Westfall and Young step-down algorithm, herein maxT (Westfall and Young 1993), a permutation-based procedure, is used to adjust for multiplicity by

D. Yekutieli (✉)
Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel
e-mail: yekutiel@post.tau.ac.il

D. Lin
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

Z. Shkedy
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be

D. Amaratunga
Biostatistics and Programming, Janssen Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ, USA
e-mail: DAMARATU@its.jnj.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_6,

controlling the Familywise Error Rate (FWER), the probability of making at least one type I error.

While there are many cases in which FWER control is needed, the purpose of gene expression data analysis is to find genes that are potential candidates for further investigation, and several erroneous rejections will not distort the conclusions at this stage of the investigation, as long as their proportion is small. Thus, controlling the probability of making even one erroneous rejection is overconservative and will result in reduced experimental efficiency due to unnecessary loss of power. A more suitable option may be controlling the false discovery rate (FDR), defined in Benjamini and Hochberg (1995) as the proportion of errors in the set of identified differentially expressed genes.

In this chapter, we discuss a few procedures controlling for the FWER, such as the Bonferroni, Holm, and the maxT procedures. However, the focus of this chapter is controlling the FDR, since it admits a more powerful outcome. We discuss several variations of the Benjamini and Hochberg step-up procedure (BH-FDR 1995), the permutation-based FDR controlling procedures, and the significance analysis of microarrays (SAM) approach of Tusher et al. (2001) and Efron et al. (2001), and Storey (2002, 2003) Bayesian interpretation of the FDR within the context of microarray data.

6.2 Controlling Type I Error

Benjamini and Hochberg (1995) considered the case where there are m hypotheses (Table 6.1) that need to be tested, among which m_0 are true null hypotheses and m_1 are false null hypotheses. Let V be the number of true null hypotheses that we reject and let R be the total number of rejected hypotheses. Note that the values of m_0 and m_1 are unknown in practice.

The FWER (Hochberg and Tamhane 1987) is defined as the probability to reject erroneously at least one true null hypothesis, i.e., FWER $= P(V > 0)$. Here the term "family" refers to the collection of null hypotheses $H_{01}, \ldots, H_{0m}$ that is being considered for joint testing. Once the family has been defined, strong control of the FWER (at a joint level α) requires that FWER $\leq \alpha$ for all possible constellations of true and false null hypotheses (Lehmann and Romano 2005).

The FDR introduced by Benjamini and Hochberg (1995) is defined as the expected proportion of false rejections among the rejected hypotheses, FDR $= E(Q)$, where $Q = V/R$ when $R > 0$, and $Q = 0$ otherwise. Approaches based on the control of the FDR have gained their popularity in the microarray setting, because they lead to higher power as compared to the methods that control the FWER.

Recently, two new error rates were introduced to control the number of false positives. The k-FWER, proposed by Hommel and Hoffman (1998), is defined as the probability of rejecting at least k true null hypotheses, i.e., k-FWER $= P(V > k)$. Controlling the k-FWER overcomes the stringency of controlling the FWER, and has the advantage of controlling the number of mistakes as compared to controlling the FWER or FDR (Xu and Hsu 2007). A generalization of the FDR, similar in spirit

Table 6.1 Decisions in multiple testing (Benjamini and Hochberg 1995)

	# Not rejected	# Rejected	Total
# True null hypotheses	U	V	m_0
# Not true null hypotheses	T	S	m_1
Total	W	R	m

to the way the k-FWER generalizes the FWER, was proposed by Sarkar (2007). The k-FDR is defined as the expected proportion of k or more false rejections among all rejections, i.e., $k\text{-FDR} = E(Q_k)$ with $Q_k = V/R$ if $V > k$ and $Q_k = 0$ if $V \leq k$. The procedures controlling the k-FWER and k-FDR are not implemented in this book.

6.3 Procedures Adjusting for Multiple Testing

In this section, we outline several procedures used to control the FWER. These include Bonferroni's and Holm's (1979) approaches. We also describe procedures for controlling the FDR: Benjamini and Hochberg step-up procedure (BH-FDR) and Benjamini and Yekutieli procedure (BY-FDR). Some other resampling-based procedures, such as the maxT (Westfall and Young 1993) for controlling the FWER, the adaptive resampling-based procedure for controlling the FDR, the SAM, and the linear mixed model for microarray (Limma), will be discussed.

6.3.1 Procedures Controlling the FWER

6.3.1.1 Bonferroni's Procedure

Let P_i be the raw p value for the test statistic t_i for gene i $(i = 1 \ldots m)$ and let H_{0i} be the corresponding null hypothesis.

The single-step Bonferroni procedure rejects H_i if $P_i \leq \alpha/m$, where α is the desired level of the FWER. This ensures that the probability that at least one true hypothesis is rejected is not greater than α. In order to keep the significance level irrespective of the number of tests performed, the adjusted p values can be computed to simplify the multiple testing procedure. The adjusted p value for Bonferroni can be written as $\tilde{P}_i = \min(m\, P_i, 1)$ and H_{0i} is rejected if $\tilde{P}_i \leq \alpha$.

6.3.1.2 Holm's Procedure

Let $P_{(1)} \leq P_{(2)} \ldots \leq P_{(i)} \ldots \leq P_{(m)}$ be the ordered p values and let $H_{0(1)}, H_{0(2)}, \ldots, H_{0(m)}$ be the corresponding null hypotheses. Holm's procedure (Holm 1979) is a step-down procedure that compares $P_{(1)}$ to α/m and sequentially compares the

ordered raw p values $P_{(i)}$ with $\alpha/(m-i+1)$ and rejects all the hypotheses stopping at the first i for which $P_{(i)}$ is smaller than the corresponding critical value. Holm's adjusted p values are given by

$$\begin{aligned}
\tilde{P}_{(1)} &= \min\left(mP_{(1)}, 1\right) \\
\tilde{P}_{(2)} &= \min\left(\max(\tilde{P}_{(1)}, (m-1)P_{(2)}), 1\right) \\
&\vdots \\
\tilde{P}_{(i)} &= \min\left(\max(\tilde{P}_{(i-1)}, (m-i+1)P_{(i-1)}), 1\right) \\
&\vdots \\
\tilde{P}_{(m)} &= \min\left(\max(\tilde{P}_{(m-1)}, P_{(m)}), 1\right).
\end{aligned}$$

$H_{(i)}$ is rejected if $\tilde{P}_{(i)} \leq \alpha$.

6.3.2 Procedures Controlling the FDR

6.3.2.1 The Linear Step-Up Procedure (BH-FDR)

For a desired FDR level q, the ordered p value $P_{(i)}$ is compared to the critical value $q \cdot i/m$. Let $k = \max\{i : P_{(i)} \leq q \cdot i/m\}$. Then reject $H_{(1)}, \ldots, H_{(k)}$, if such a k exists. Benjamini and Hochberg (1995) showed that when the test statistics are independent, this procedure controls the FDR at the level $q \cdot m_0/m \leq q$. Benjamini and Yekutieli (2001) further showed that the FDR $\leq q \cdot m_0/m$ for positively dependent test statistics as well. The technical condition under which the control holds is that of positive regression dependency on each test statistic corresponding to the true null hypotheses.

6.3.2.2 The BY-FDR Procedure

Benjamini and Yekutieli (2001) further derived a universal bound for the FDR of the level q BH-FDR procedure $q \cdot (\sum_{j=1}^{m} 1/j) \cdot m_0/m$. This yields the BY-FDR procedure, a modification of the BH-FDR procedure that controls the FDR for any joint test statistic distribution: reject $H_{(1)}, \ldots, H_{(k)}$, if $k = \max\{i : P_{(i)} \leq q \cdot i/[m \cdot (\sum_{j=1}^{m} 1/j)]\}$ exists.

The adjusted p values for the BH-FDR and BY-FDR procedures are given by

$$\tilde{P}_{(i)} = \min_{k=i,\ldots,m} \left[\min\left(\frac{m \cdot C}{i} P_{(i)}, 1 \right) \right], \tag{6.1}$$

where $C = 1$ for the BH-FDR procedure and $C = \sum_{j=1}^{m} 1/j$ for the BY-FDR procedure.

6.4 Procedures Controlling for the FWER and FDR in R

6.4.1 A Small Example

The procedures discussed above are implemented in the R function `p.adjust()`:

```
p.adjust(p, method = p.adjust.methods, n = length(p))
p.adjust.methods
# c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none")
```

Suppose we have the ordered p values as follows:

```
0.0001, 0.0004, 0.0019,0.0095,0.0201, 0.0278,0.0298,
0.0344, 0.0459, 0.3240, 0.4262, 0.5719, 0.6528,0.7590,1
```

The adjusted p values by using these four procedures discussed above can be obtained using the following code:

```
> praw <-c(0.0001, 0.0004, 0.0019,0.0095,0.0201, 0.0278,0.0298,
0.0344, 0.0459, 0.3240, 0.4262, 0.5719, 0.6528,0.7590,1)
> p.Bonf <- p.adjust(praw, method = "bonferroni", n = length(praw))
> p.Holm <-p.adjust(praw, method = "holm", n = length(praw))
> p.BH <-p.adjust(praw, method = "BH", n = length(praw))
> p.BY <-p.adjust(praw, method = "BY", n = length(praw))
> cbind(praw, p.Bonf, p.Holm, p.BH, p.BY)
       praw p.Bonf p.Holm       p.BH        p.BY
 [1,] 0.0001 0.0015 0.0015 0.00150000 0.004977343
 [2,] 0.0004 0.0060 0.0056 0.00300000 0.009954687
 [3,] 0.0019 0.0285 0.0247 0.00950000 0.031523175
 [4,] 0.0095 0.1425 0.1140 0.03562500 0.118211908
 [5,] 0.0201 0.3015 0.2211 0.06030000 0.200089208
 [6,] 0.0278 0.4170 0.2780 0.06385714 0.211892623
 [7,] 0.0298 0.4470 0.2780 0.06385714 0.211892623
 [8,] 0.0344 0.5160 0.2780 0.06450000 0.214025770
 [9,] 0.0459 0.6885 0.3213 0.07650000 0.253844518
[10,] 0.3240 1.0000 1.0000 0.48600000 1.000000000
[11,] 0.4262 1.0000 1.0000 0.58118182 1.000000000
[12,] 0.5719 1.0000 1.0000 0.71487500 1.000000000
[13,] 0.6528 1.0000 1.0000 0.75323077 1.000000000
[14,] 0.7590 1.0000 1.0000 0.81321429 1.000000000
[15,] 1.0000 1.0000 1.0000 1.00000000 1.000000000
```

The following R code can be used to produce Fig. 6.1:

```
> plot(praw,ylab="p-values")
> lines(praw)
> points(p.Bonf,pch=2)
> lines(p.Bonf,lty=2)
> points(p.Holm,pch=3)
> lines(p.Holm,lty=3)
> points(p.BH,pch=4)
> lines(p.BH,lty=4)
> points(p.BY,pch=5)
> lines(p.BY,lty=5)
> title("Multiplicity Adjustment of P-values")
```

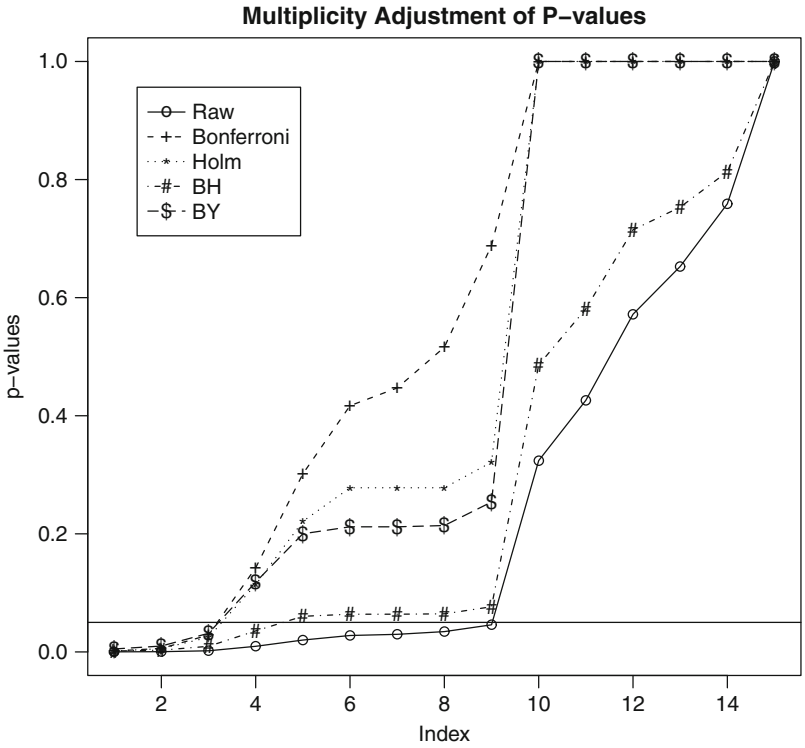

Fig. 6.1 The multiplicity adjusted p values

There are nine raw p values (solid line) below 0.05. From the adjusted p values, the BH procedure (with #) discovers four genes as differentially expressed, while all the other procedures yield three significant genes at the significance level of 0.05 (horizontal solid line).

6.4.2 A Microarray Experiment on the Animal Behavior Study

This case study was obtained from a behavioral experiment, in which 24 males, experimentally naive Long-Evans rats obtained from Janvier (France), weighing 300–370 g at the start of the experiment, were randomized into two treatment groups (12 rats in each group). Quinpirole hydrochloride (Sigma-Aldrich) (treatment 1) was dissolved in physiological saline and administered at a dose of 0.5 mg/kg s.c. (the method used by Szechtman et al. 1998). Equivalent volumes of saline (treatment 0) were used in solvent injections. Animals were tested in a large open field. Two datasets encompassed behavioral response and microarray data. Rat behavior data

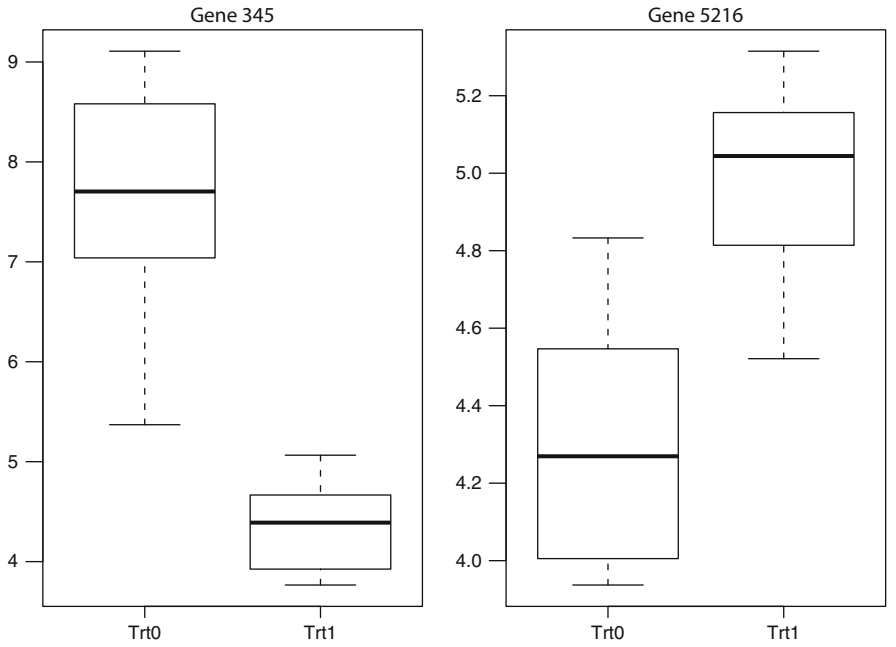

Fig. 6.2 Boxplot of downregulated expression for gene 345 (panel **a**) and upregulated expression for gene 5216 (panel **b**) in the two treatment groups

parameters, suggested by Szechtman et al. (1998) for the definition of compulsive checking, were recorded systematically. In particular, the parameter of primary interest is defined as the distance traveled by the rats. The active response to the treatment doses is expected to increase this distance. After dose administration, rat microarrays were taken by cutting the rat's brain into the frontal, striatum, and thalamus parts. Thus, there are three microarrays for each animal, and thousands of gene expressions are measured within each array. Each array measured the expression levels of 5,644 genes for each rat. All microarray-related steps, including the amplification of total RNAs, labeling, hybridization, and scanning, were carried out as described in the GeneChip Expression Analysis Technical Manual, Rev.4 (Affymetrix 2004). For analysis, we consider the gene expression levels from the thalamus part of the brain. Figure 6.2a, b show boxplots of gene expression (for genes 345 and 5216, respectively) under two treatment conditions. The two genes seem to be differentially expressed between the two treatments.

We use a two-sample t-test to obtain asymptotic p values for testing the treatment effects and apply the multiplicity adjustment methods discussed above to these p values. The results of the differentially expressed genes detected after adjustment by Bonferroni, Holm, the BH-FDR, and BY-FDR procedures are given in Table 6.2, and R code used to produce the results is given by

Table 6.2 Raw and adjusted p values for differentially expressed genes found by using t-tests with multiplicity adjustment (only adjusted p values <0.05 are reported)

Gene	Raw p value	Bonferroni p value	Holm p value	BH p value	BY p value
59	0.0000	0.0215	0.0215	0.0027	0.0248
60	0.0000	0.0318	0.0318	0.0032	0.0293
158	0.0001			0.0477	
214	0.0000	0.0134	0.0134	0.0021	0.0194
345	0.0000	0.0006	0.0006	0.0002	0.0019
486	0.0000	0.0012	0.0012	0.0003	0.0028
662	0.0000	0.0002	0.0002	0.0001	0.0011
1962	0.0000	0.0019	0.0019	0.0004	0.0035
2247	0.0000	0.0289	0.0289	0.0032	0.0293
4297	0.0001			0.0477	
4447	0.0000	0.0002	0.0002	0.0001	0.0011
5216	0.0000	0.0147	0.0147	0.0021	0.0194
5614	0.0001			0.0452	

```
> pval <- apply(gene.exp,1,function(x)t.test(x~trt)$p.value)
> p.Bonferroni <- p.adjust(as.numeric(pval), method="bonferroni")
> p.Holm <- p.adjust(as.numeric(pval), method="holm")
> p.BH <- p.adjust(as.numeric(pval), method="BH")
> p.BY <- p.adjust(as.numeric(pval), method="BY")
```

The number of differentially expressed genes found after Bonferroni, Holm, and the BY-FDR adjustment is the same (i.e., 10), while the BH-FDR adjustment yields 13 differentially expressed genes with smaller adjusted p values than those of the other methods.

6.5 Resampling-based Multiple Testing Procedures

In a microarray setting, resampling methods to adjust for multiplicity are often used (Reiner et al. 2003; Tusher et al. 2001; Ge et al. 2003). The main motivation is to avoid inference based on asymptotic distribution of the test statistics, which, within the microarray setting, can be problematic because of either typically small sample sizes or departure from the assumption about the distribution of the response. Also, in some cases, the asymptotic distribution of the test statistics is unknown (Tusher et al. 2001).

6.5.1 Obtaining p Values from Permutations

Consider a microarray with m genes and let $t_1, \ldots, t_i, \ldots, t_m$ be the test statistics of primary interest. By permuting the labels of arrays randomly, the (permutation) test statistics of m genes are recalculated.

Let T be the permutation matrix i.e.,

$$T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1B} \\ t_{21} & t_{22} & \dots & t_{2B} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ t_{m1} & t_{m2} & \dots & t_{mB} \end{pmatrix}, \tag{6.2}$$

where B is the total number of permutations and each element t_{ib} of matrix T is the test statistic for the ith gene in the bth permutation. The raw p values are calculated as

$$P_i = \frac{\#(b : |t_{ib}| \geq |t_i|)}{B}, \tag{6.3}$$

where t_i is the observed test statistic for gene i. Note that if the number of observations is relatively small, then the distribution of the permutation-based p values is granular. Alternatively, less discrete p values can be obtained from the marginal distribution of permutation-based test statistics for all the genes in the array:

$$P_i = \frac{\sum_{b=1}^{B} \sum_{j=1}^{m} (|t_{jb}| \geq |t_i|)}{B \times m}. \tag{6.4}$$

The raw p values obtained by using (6.3) or (6.4) can be used as an input in the procedures for multiple testing adjustment discussed in the next section. The following code can be applied to the animal behavior case study to obtain the raw p values and adjusted Bonferroni and the BH-FDR p values.

```
> perm.mat <- matrix(0, nrow(gene.exp),100)
> niter=100
> xiter.index <- t(sapply(1:niter, function(i) sample(trt)))
> library(multtest)
> for (i in 1: 100){
 set.seed(i)
perm.mat[,i] <- mt.teststat(gene.exp, xiter.index[i,])
print(i)
}
> # obs0 is the observed t-test statistics
> p.raw <- sapply(obs0, function(x)
sum(abs(x) <= abs(perm.mat))/niter/length(obs0))
> p.Bonf <- p.adjust(p.raw, method = "bonferroni", n = length(p.raw))
> p.BH <-p.adjust(p.raw, method = "BH", n = length(p.raw))
```

The histogram of the p values obtained based on the permutations by using the method in (6.4) seems to be uniformly distributed (see Fig. 6.3a). This indicates that there are few genes which seem to be differentially expressed. The number of differentially expressed genes found by using both Bonferroni and the BH-FDR

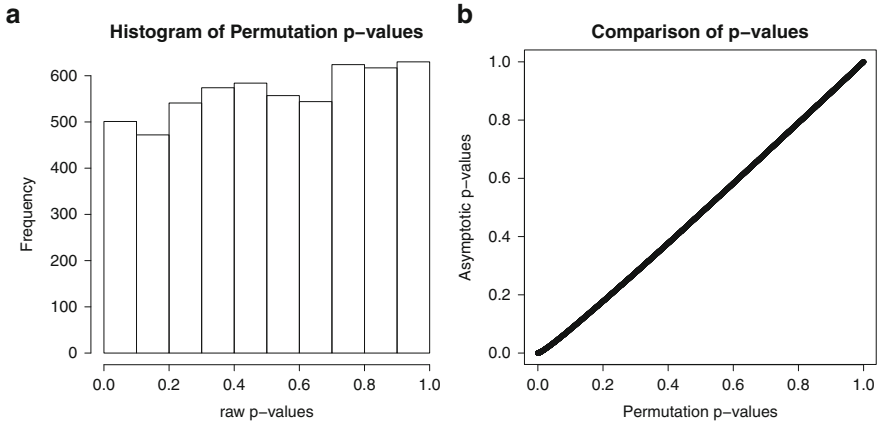

Fig. 6.3 The raw p values. (**a**) Histogram of raw p values based on permutations using the method in (6.4). (**b**) Comparison of raw p values from permutations and asymptotic t-tests

procedures is nine, indicating which are the most differentially expressed among all the 5,644 genes. All the nine genes found have permutation p values equal to zero and are contained in the list of differentially expressed genes (except genes 158, 2247, 4297, and 5614) detected using the asymptotic p values by the t-tests and adjusted by Bonferroni and the BH-FDR procedures. The comparison between the asymptotic p values and permutation p values shows a high agreement, as shown in Fig. 6.3b.

6.5.2 *The maxT Procedure*

The maxT procedure, proposed by Westfall and Young (1993) in order to control the FWER, is discussed in the context of microarray analysis by Ge et al. (2003). The starting points are the observed statistics and the permutation matrix T. Let $t_{(1)} \geq t_{(2)} \geq, \ldots, \geq t_{(m)}$ be the ordered values of the test statistics. The permutation matrix T is sorted based on the original order of the observed statistics (Westfall and Young 1993). For each column of the permutation matrix, the adjusted test statistics are calculated in the following way:

$$\begin{aligned} u_{(m),b} &= |t_{(m),b}| && \text{for } i = m, \\ u_{(i),b} &= \max(u_{(i+1),b}, |t_{(i),b}|) && \text{for } i = m-1, \ldots, 1. \end{aligned}$$

Once the adjusted matrix of the test statistics is obtained, the adjusted p values are calculated (over the rows of the matrix):

$$\tilde{P}^{\star}_{(i)} = \frac{\#(b : u_{(i),b} \geq |t_{(i)}|)}{B}.$$

6.5.2.1 Implementation of the maxT Procedure

For a two sample comparison, the maxT procedure calculates the raw p values of t-test statistics or Wilcoxon test statistics by using the sampling-based approach given by (6.3). This is implemented in the R function `mt.maxT` in the `multtest()` package :

```
> library(multtest)
> maxt <- mt.maxT(gene.exp,trt,test="t",side="abs",fixed.seed.sampling="y",
B=10000)
> sum(maxt[,4]<=0.05)
> 10
> maxt[1:10,]
ID   teststat  rawp   adjp
1962  -10.407638 1e-04 0.0001
 345   -9.702417 1e-04 0.0001
4447   -9.200213 1e-04 0.0001
 662   -8.272687 1e-04 0.0006
  60   -7.602661 1e-04 0.0010
 486   -7.477946 1e-04 0.0011
  59   -6.580940 1e-04 0.0051
5216    6.367395 2e-04 0.0080
 214   -6.311827 1e-04 0.0085
2247   -6.086100 1e-04 0.0141
```

In the animal behavior study, ten genes are found to be differentially expressed by using the maxT approach. This procedure gives one more significant gene (gene 2247) as compared to Bonferroni and the BH-FDR approach based on the raw p values obtained from the permutations using the method in (6.4).

6.5.3 *FDR Control in R Using the `FDR-AME` Package*

The `FDR-AME` (FDR adjustments of microarray experiments) package computes FDR adjustments for p values generated in multiple hypothesis testing of gene expression data obtained by the animal behavior experiment. It applies both theoretical distribution-based and resampling-based multiple testing procedures and presents as output adjusted p values and p value plots, as described in Reiner et al. (2003). It also computes p value adjustments that correspond to the adaptive two-stage FDR controlling procedures in Benjamini et al. (2006), and can be applied to discover differences in expression between many classes using one-way ANOVA models.

6.5.3.1 Two-Stage Approach

Since the BH procedure controls the FDR at a level too low by a factor of m_0/m, it is natural to try to estimate m_0 and use $q^* = qm/m_0$ instead of q to gain more power. Benjamini et al. (2006) suggest a simple two-stage procedure: use the BH-FDR procedure once to reject r_1 hypotheses; then use the BH-FDR at the second stage at level $q^* = qm/(m - r_1)(1 + q)$. This two-stage procedure has proven

FDR controlling properties under independence and simulation studies support for controlling properties under positive dependence.

6.5.3.2 Resampling-based FDR Adjustments

The underlying assumption in p value resampling is that the joint distribution of p values corresponding to the true null hypotheses, which is generated through the p value resampling scheme, is identical to its joint distribution for any configuration of false null hypotheses (this property is referred to as subset pivotality in Westfall and Young's monograph). Thus, for each value of p, the distribution of the number of resampling-based p values less than p, denoted by $V^*(p)$, stochastically dominates the distribution of $V(p)$, the number of p values corresponding to true null hypotheses less than p.

Yekutieli and Benjamini (1999) introduced resampling-based FDR adjustments in which, for each value of p, they first conservatively estimate the number of false null hypotheses less than p, denoted by $\hat{s}(p)$, and then estimate the FDR adjustment by

$$\mathrm{FDR}^{\mathrm{est}}(p) = E_{V^*(p)} \frac{V^*(p)}{V^*(p) + \hat{s}(p)}.$$

Two estimation methods for $\hat{s}(p)$ are suggested differing by their strictness level. The FDR local estimator is conservative on the mean, and the FDR upper limit bounds the FDR with probability 95%.

6.5.3.3 BH-FDR Adjustment for Resampling-based p Values

A third alternative uses the BH procedure to control the FDR, but rather than using the raw p values, it applies the BH-FDR adjustment to resampling-based p values similar to those computed in (6.4).

6.5.3.4 Implementation of the `fdrame` Package

The `fdr.ma()` is the main function in the `fdrame` package. A general call for the function has the form of

```
fdr.ma(exp.arr,design,
       p.method=c("theoretic",""resampling"),
       fdr.adj=c("BH-LSU","adaptive","point.est","upper.est"),
       equal.var=TRUE,
       plot=c("pvlVSrank","adjVSstat"),perms.num=100)
```

The `fdr.ma()` function provides four methods for adjusting the FDR: the BH-FDR approach with theoretic and resampling-based p values, the two-stage

adaptive procedure can be used in combination with both the theoretical and resampling-based p values, while the upper and point estimates of the FDR are obtained only using the resampling-based procedure. The following code can be used to obtain the number of differentially expressed genes for each of the four procedures for the animal behavior case study:

```
> library(fdrame)
> fdr.BH.theoretic <- fdr.ma(data.matrix(gene.exp),trt,
  p.method="theoretic",fdr.adj="BH-LSU",equal.var=TRUE)

> fdr.BH.resampling <- fdr.ma(data.matrix(gene.exp),trt,
  p.method="resampling",fdr.adj="BH-LSU",equal.var=TRUE,perms.num=100)

> fdr.adaptive.theoretic <- fdr.ma(data.matrix(gene.exp),trt,
  p.method="theoretic",fdr.adj="adaptive",equal.var=TRUE)

> fdr.adaptive.resampling <- fdr.ma(data.matrix(gene.exp),trt,
  p.method="resampling",fdr.adj="adaptive",equal.var=TRUE)

> fdr.upper.resampling <- fdr.ma(data.matrix(gene.exp),trt,
  p.method="resampling",fdr.adj="upper.est",equal.var=TRUE,perms.num=100)

> fdr.point.resampling <- fdr.ma(data.matrix(gene.exp),trt,
  p.method="resampling",fdr.adj="point.est",equal.var=TRUE,perms.num=100)
```

The resampling-based inference (`p.method=resampling`) with two-stage adaptive procedure (`fdr.adj=adaptive`) leads to a higher number of significant genes:

```
> sum(fdr.BH.theoretic$adj<=0.05)
[1] 11
> sum(fdr.BH.resampling$adj<=0.05)
[1] 11
> sum(fdr.adaptive.theoretic$adj<=0.05)
[1] 13
> sum(fdr.adaptive.resampling$adj<=0.05)
[1] 20
> sum(fdr.upper.resampling$adj<=0.05)
[1] 11
> sum(fdr.point.resampling$adj<=0.05)
[1] 18
```

The table produced by the code bellow lists the top 20 differentially expressed genes with the adjusted p values for all the six methods:

```
> Toplist <- which(fdr.adaptive.resampling$adj<=0.05)
> TopTable <- cbind(fdr.BH.theoretic$adj[Toplist],
fdr.BH.resampling$adj[Toplist], fdr.adaptive.theoretic$adj[Toplist],
fdr.adaptive.resampling$adj[Toplist], fdr.upper.resampling$adj[Toplist],
fdr.point.resampling$adj[Toplist])
> colnames(TopTable) <- c("BH.theo","BH.rsmpl","adpt.theo",
  "adpt.rsmpl","up.rsmpl","pnt.rsmpl")
> rownames(TopTable) <- Toplist
> TopTable
```

In the edited top table (see Table 6.3), genes are ordered according to their row numbers in the dataset, and we omit some of the p values which are greater

Table 6.3 Edited top table: top 20 differentially expressed genes declared by the various resampling-based FDR procedures with the adjusted p values

	BH.theo	BH.rsmpl	adpt.theo	adpt.rsmpl	up.rsmpl	pnt.rsmpl
59	0.0012	0.0012	0.001	0	0	0.0011
60	0.0002	0	0.0002	0	0	0
158			0.0477	0.0315		0.0335
214	0.0017	0.0013	0.0015	0.0011	0	0.0011
345	0	0	0	0	0	0
486	0.0002	0	0.0002	0	0	0
522				0.0485		
662	0.0001	0	0	0	0	0
1022				0.0439		0.0464
1316				0.0424		0.0456
1962	0	0	0	0	0	0
2247	0.0025	0.0033	0.0023	0.003	0	0.002
2489			0.0533	0.0329		0.0368
3170				0.0424		0.0456
4297			0.0477	0.0283		0.0335
4447	0	0	0	0	0	0
5216	0.0017	0.0012	0.0015	0.0011	0	0.0011
5352				0.0485		
5356				0.0424		0.0456
5614	0.0347	0.037	0.0315	0.0191	0.0355	0.0248

Genes are ordered according to their row number

than 0.05. The FDR adaptive procedure yields the largest number of differentially expressed genes (20), which contains all the significant genes found by the other methods. Depending on the random permutation seed, the output of these results can vary; therefore, a fixed random seed (123) was used to produce the number of differentially expressed genes above.

6.5.4 SAM Procedure

The SAM (Tusher et al. 2001) is a widely used testing procedure that estimates the FDR by using permutations under the assumption that all null hypotheses are true. The procedure consists of three components: (1) the adjusted test statistics, (2) an approximation of the distribution of the test statistics based on permutations, and (3) the control of the FDR.

For a two-group setting, the modified test statistic for gene i in SAM is given by,

$$t_i^{\text{SAM}} = \frac{\bar{X}_1 - \bar{X}_2}{s_i + s_0}, \tag{6.5}$$

where

$$\bar{X}_1 = \frac{\sum_{j=1}^{n_1} x_{1j}}{n_1}, \qquad \bar{X}_2 = \frac{\sum_{j=1}^{n_2} x_{2j}}{n_2},$$

and

$$s_i = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{\sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2 + \sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2}{n_1 + n_2 - 2}}.$$

Here, s_0 is the fudge factor which is estimated from the data and is discussed later (in Sect. 8.3).

The SAM requires that the test statistic for each permutation is sorted for all the genes, such that the first row of the sorted matrix is the minimum test statistic across permutations and the last row is the maximum. The expected values of the observed ordered statistics are approximated by the means of the rows of the sorted permutation matrix $\boldsymbol{T}^{\text{SAM}}$:

$$\boldsymbol{T}^{\text{SAM}} = \begin{pmatrix} t_{(1)1} & t_{(1)2} & \dots & t_{(1)B} \\ t_{(2)1} & t_{(2)2} & \dots & t_{(2)B} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ t_{(m)1} & t_{(m)2} & \dots & t_{(m)B} \end{pmatrix} \Rightarrow \boldsymbol{T}^{\text{SAM}^{\text{e}}} = \begin{pmatrix} \bar{t}_{(1)1} \\ \bar{t}_{(2)1} \\ \vdots \\ \bar{t}_{(m)1} \end{pmatrix}. \tag{6.6}$$

Let $t_i^{\text{SAM}^{\text{o}}}$ and $t_i^{\text{SAM}^{\text{e}}}$ be the observed and the expected values of the test statistic for the ith gene, respectively. A gene is declared differentially expressed whenever

$$|t_i^{\text{SAM}^{\text{o}}} - t_i^{\text{SAM}^{\text{e}}}| \geq \Delta.$$

The value of Δ is chosen in order to control the FDR at a desired level. For the details of the SAM procedure, approximating the FDR, and choice of the fudge factor s_0 used in the test statistic, we refer to Tusher et al. (2001) and Chap. 8.

6.5.4.1 Implementation of the SAM Using the `samr` Package

The R package `samr` can be used to perform the SAM in R. In this section, we illustrate how to use the `samr()` function for the animal behavior experiment described in Sect. 6.4.2. For a two-sided t-test we use

```
> dim(data) ## gene expression
> d=list(x=data,y=trt+1,geneid=as.character(1:5644),
genenames=paste("gene", as.character(1:5644)),logged2=False)
> samr.obj <- samr(d,  resp.type="Two class unpaired")
```

Table 6.4 Delta table produced by the SAM procedure

	Delta	# med FALSE pos	90th perc FALSE pos	# called	Median FDR	90th perc FDR	cutlo	cuthi
[1,]	0.0000	3035.4621	3183.3891	2934	1.0346	1.0850	–1.4691E+00	3.3951E-02
[2,]	0.0026	3032.7725	3178.8706	2932	1.0344	1.0820	–1.4796E+00	3.3951E-02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
[16,]	0.5831	1.0758	4.3033	16	0.0672	0.2690	–2.6728E+00	3.8381E+00
[17,]	0.6633	1.0758	3.2275	14	0.0768	0.2305	–2.7930E+00	3.8381E+00
[18,]	0.7488	0.0000	2.1517	13	0.0000	0.1655	–3.0525E+00	3.8381E+00
[19,]	0.8395	0.0000	2.1517	13	0.0000	0.1655	–3.0525E+00	3.8381E+00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
[49,]	5.9701	0.0000	0.0000	1	0.0000	0.0000	–9.1204	1.0000E+10
[50,]	6.2214	0.0000	0.0000	0	NaN	NaN	–1.0000e+10	1.0000E+10

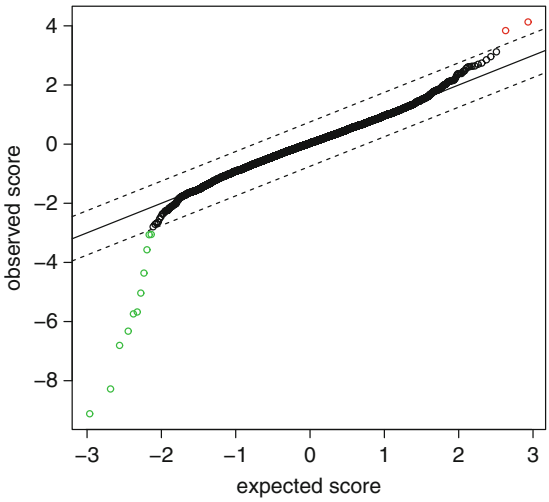

Fig. 6.4 Differentially expressed genes identified by the SAM with expected and observed test statistics

As we mentioned above, we need to select the value of Δ. In practice, the `samr()` calculates the FDR for several values of Δ:

```
> delta.table <- samr.compute.delta.table(samr.obj)
> delta.table
```

We obtain the delta table as follows (Table 6.4).

By choosing the FDR under 0.05, the corresponding Δ value is 0.75. The SAM plot (in Fig. 6.4) and the list of significant genes can be obtained accordingly. This yields 13 genes to be differentially expressed between comparison of gene expression in two treatments:

```
> delta=.75
> samr.plot(samr.obj,delta)
> siggenes.table<-samr.compute.siggenes.table(samr.obj,delta,
d, delta.table)
> siggenes.table

$genes.up
        Row Gene ID Gene Name Score(d) Numerator(r) Denominator(s+s0)
        Fold Change q-value(%)
[1,] "5217" "gene 5216" "5216"    "4.1331" "0.6936" "0.1678" "1.1613" "0"
[2,] "5615" "gene 5614" "5614"    "3.8381" "1.0157" "0.2646" "1.3296" "0"

$genes.lo
        Row Gene ID Gene Name Score(d) Numerator(r) Denominator(s+s0)
        Fold Change q-value(%)
 [1,] "1963" "gene 1962" "1962" "-9.1204" "-4.3426"  "0.4761" "0.2665" "0"
 [2,] "346"  "gene 345"  "345"  "-8.2776" "-3.3194"  "0.4010" "0.5671" "0"
 [3,] "61"   "gene 60"   "60"   "-6.8054" "-3.8216"  "0.5616" "0.2773" "0"
 [4,] "487"  "gene 486"  "486"  "-6.3247" "-2.4150"  "0.3818" "0.5814" "0"
 [5,] "4448" "gene 4447" "4447" "-5.7429" "-0.8990"  "0.1567" "0.8394" "0"
 [6,] "60"   "gene 59"   "59"   "-5.6754" "-2.4290"  "0.4280" "0.5621" "0"
 [7,] "663"  "gene 662"  "662"  "-5.0379" "-0.7587"  "0.1506" "0.8880" "0"
 [8,] "215"  "gene 214"  "214"  "-4.3599" "-0.8303"  "0.1904" "0.8927" "0"
 [9,] "159"  "gene 158"  "158"  "-3.5751" "-0.8809"  "0.2464" "0.8682" "0"
[10,] "1570" "gene 1569" "1569" "-3.0626" "-1.0422"  "0.3403" "0.4991" "0"
[11,] "2248" "gene 2247" "2247" "-3.0525" "-0.3606"  "0.1181" "0.9486" "0"

$color.ind.for.multi
NULL

$ngenes.up
[1] 2

$ngenes.lo
[1] 11
```

6.5.5 *The Limma: Empirical Bayes Approach*

The `Limma` package fits a hybrid frequentist empirical Bayes (eBayes) linear model (Smyth 2004; Efron et al. 2001) for the expression levels of the genes in the array. The expression level variances of the genes in the array, $\sigma_1^2 \cdots \sigma_m^2$, are elicited a scaled inverse chi-squared marginal prior density. The hyperparameters of this prior distribution, s_0^2—the prior variance and ν_0—the prior degrees of freedom, are derived by applying the Limma eBayes function to the sample variances $s_1^2 \cdots s_m^2$. Limma produces moderated t and F statistics, computed by dividing the standard frequentist numerator with a denominator in which the ν degree of freedom sample variance s_i^2 is replaced with $\tilde{s}_i^2 = (\nu_0 s_0^2 + \nu s_i^2)/(\nu_0 + \nu)$ the posterior mean of $\sigma_i^2 | s_i^2$. The significance levels provided in the Limma are the corresponding tabulated t or F values with augmented $\nu_0 + \nu$ degrees of freedom.

6.5.5.1 Analysis Using the `Limma` Package

The empirical Bayes inference procedure is implemented in the R `Limma` package in `Bioconductor`. The main function `lmFit` calls the following input:

```
lmFit(object,design=NULL,ndups=1,spacing=1,block=NULL,correlation,
weights=NULL,method="ls", . . .)
```

where the `object` should be of class numeric, matrix, MAList, EList, marrayNorm, ExpressionSet, PLMset containing log-ratios, or log-values of expression for a series of microarrays, and `design` defines the dummy coded design matrix for exploratory variables in the model. The function also supports two different correlation structures. By defining the `block` structure, different arrays are assumed to be correlated; while assigning `block` as NULL and `ndups` greater than one, replicate spots on the same array are assumed to be correlated. However, it is not possible to fit models with both a block structure and a duplicate-spot correlation structure simultaneously.

By loading the library and specifying the data matrix and the design matrix from the animal behavior study, the following code can be used to illustrate the use of empirical Bayes inference:

```
> library(limma)
> design <- cbind(rep(1,24),trt)
> design
          trt
 [1,] 1   1
 [2,] 1   1
 [3,] 1   1
 [4,] 1   1
 [5,] 1   1
 [6,] 1   1
 [7,] 1   1
 [8,] 1   1
 [9,] 1   1
[10,] 1   1
[11,] 1   1
[12,] 1   1
[13,] 1   0
[14,] 1   0
[15,] 1   0
[16,] 1   0
[17,] 1   0
[18,] 1   0
[19,] 1   0
[20,] 1   0
[21,] 1   0
[22,] 1   0
[23,] 1   0
[24,] 1   0
```

```
> fit <- lmFit(gene.exp,design)
> fit1 <- eBayes(fit)
```

After obtaining the model fit from the `lmFit()`, the output can be directly passed to the function `eBayes()`, which computes moderated *t*-statistics, moderated *F*-statistic, and log-odds of differential expression by empirical Bayes shrinkage of the standard errors toward a common value. And some additional information can be obtained for the prior variance (common for all genes), sample variance, the prior degrees of freedom (common for all genes), and posterior variance:

```
> fit1$t
$sigma
        3         4         5         6         7
0.3175772 0.2390783 0.3410579 0.5360771 0.6993906
5639 more elements . . .

$df.prior
[1] 2.885634

$s2.prior
[1] 0.07502008

$s2.post
          3           4           5           6           7
0.09785952 0.05922962 0.11153145 0.26275448 0.44112675
5639 more elements . . .
```

The results (the moderated t-test statistics and adjusted p values) of the analysis are shown as

```
> fit1$t
$t
                    trt
3  65.35861 -0.88594877
4 124.81012 -1.76788239
5  91.84175 -0.56723023
6  27.67912 -0.02735882
7  13.60096 -1.26273256
5639 more rows . . .
> fit1$p.value
$p.value
                      trt
3 2.311827e-29 0.38412651
4 2.479289e-36 0.08934106
5 5.038620e-33 0.57563939
6 3.265455e-20 0.97839176
7 5.043676e-13 0.21838979
5639 more rows . . .
```

After the multiplicity adjustment (the BH-FDR procedure), 14 genes are declared to be differentially expressed with the significance level of 0.05:

```
> topTable <- topTable(fit1,number=nrow(gene.exp),coef=2)
> topTable[topTable[,6]<=0.05,]
       ID      logFC  AveExpr          t      P.Value    adj.P.Val       B
1962 1964 -4.3425482 3.749374 -11.017402 4.632269e-11 2.614453e-07 13.4208
345   347 -3.3194123 6.007432 -10.247585 2.052915e-10 5.793325e-07 12.3089
4447 4449 -0.8999684 5.154405  -9.040867 2.454054e-09 4.616894e-06 10.3814
60     62 -3.8216171 3.377434  -8.059791 2.121030e-08 2.454592e-05  8.6380
662   664 -0.7587221 6.394576  -8.048792 2.174514e-08 2.454592e-05  8.6176
486   488 -2.4150473 4.560990  -7.891463 3.110503e-08 2.925947e-05  8.3226
59     61 -2.4290157 4.331910  -6.957489 2.784001e-07 2.244700e-04  6.4861
214   216 -0.8302660 7.326145  -6.415855 1.042949e-06 7.358006e-04  5.3567
5216 5218  0.6936288 4.646642   6.347682 1.234510e-06 7.741747e-04  5.2114
2247 2249 -0.3606362 6.830275  -5.344121 1.555993e-05 8.782022e-03  3.0036
5614 5616  1.0157272 3.589754   5.151527 2.552640e-05 1.309737e-02  2.5678
158   160 -0.8808504 6.243623  -4.883900 5.092234e-05 2.395047e-02  1.9580
4297 4299  0.4650609 7.425946   4.649801 9.330014e-05 4.050661e-02  1.4220
1316 1318  0.6048107 4.465285   4.564982 1.161982e-04 4.684448e-02  1.2274
```

By comparing the differentially expressed genes identified by the methods discussed above (the maxT, the BH-FDR, and BY-FDR procedures based on the permutation p values, the resampling-based FDR procedures, Limma, and the SAM) given by the top tables, we can conclude that the majority of genes (12) identified are in common, with a few additional genes declared significant by some procedures. All the methods yield a consistent list of differentially expressed genes.

6.6 Discussion

The procedures that control the FDR are preferred over those that control of the FWER as they result in tests that have higher power. The BH-FDR, BY-FDR, maxT, the adaptive procedures, and the SAM procedures can be combined with resampling-based inference (RBI) to adjust for multiple testing without strong distributional assumptions.

In this chapter, we have discussed the resampling-based procedures for controlling the FWER and FDR. The advantage of such resampling-based procedures is to avoid the distributional assumption of gene expression data. Approaches, such as p values obtained from permutations adjusted by the BH-FDR, BY-FDR procedure, the adaptive procedures of the FDR, and the SAM procedure, can be used to find differentially expressed genes. However, an issue related to RBI is that, because of the small samples that are typically used in microarray experiments, the RBI p value distributions can be coarse or granular. As a result, it will often be difficult to obtain p values (as in (6.3)) that are below some specified level. To overcome this problem, three approaches are discussed in this chapter, i.e., (1) raw p values are obtained using (6.4) and adjusted by the BH-FDR procedure, (2) the SAM procedure combines resampled test statistics across all genes as the null distribution to obtain very small p values, and (3) the Limma empirical Bayes approach uses moderated test statistics and estimates the common variance for all the genes. The procedures preserve the correlation among test statistics of all the genes. Moreover, using Limma one can borrow strength for variance estimation across the genes and derive more powerful rejection regions for testing by assuming a statistic from a mixture of the null and alternative distributions, as well as from the pure null distribution (Efron et al. 2001). However, this is based on two assumptions: that the null distribution of the test statistic is the same for all genes; and that genes are independent of each other. Unfortunately, neither of the assumptions is necessarily valid.

References

Affymetrix GeneChip. (2004). *Expression analysis technical manual, Rev.4*. Santa Clara: Affymetrix.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Soceity B, 57*, 289–300.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up false discovery rate controlling procedures. *Biometrika, 93*(3), 491–507.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annal of Statistics, 29*(4), 1165–1188.

Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association, 98*, 77–87.

Efron, B., Tibshirani, R., Storey, J. D., & Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association, 96*, 1151–1160.

Ge, Y., Dudoit, S., & Speed, P. T. (2003). *Resampling based multiple testing for microarray data analysis* (Technical report, 633). Berkeley: University of Berkeley.

Hochberg, Y., & Tamhane, Y. C. (1987). *Multiple comparison procedures*. New York: Wiley.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.

Hommel, G., & Hoffman, T. (1998). Controlled uncertainty. In P. Bauer, G. Hommel, & E. Sonnemann (Eds.), *Multiple hypotheses testing* (pp. 154–161). Heidelberg: Springer.

Lehmann, E.L., & Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist., 33*(3), 1138–1154.

Reiner, A., Yekutieli, D., & Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics, 19*(3), 368–375.

Sarkar, S. K. (2007). Procedures controlling generalized FWER and generalized FDR. *The Annals of Statistics, 35*(6), 2405–2420.

Smyth, G. K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology, 3*(1), Article 3.

Storey, J. D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society B, 64*(Pt 3), 479–498.

Storey, J. D. (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics, 31*(6), 2013–2035.

Szechtman, H., Sulis, W., & Eilam, D. (1998). Quinpirole induces compulsive checking behavior in rats: A potential animal model of obsessive-compulsive disorder (OCD). *Behavioral Neuroscience, 112*, 1475–1485.

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrys applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences, 98*, 5116–5121.

Westfall, P. H., & Young, S. S. (1993). *Resampling based multiple testing*. New York: John Wiley & Sons.

Xu, H., & Hsu, J. C. (2007). Using the partitioning principle to control the generalized family error rate. *Biometrical Journal, 49*, 52–67.

Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference, 82*, 171–196.

Chapter 7
Single Contrast Tests

Dan Lin, Ziv Shkedy, Daniel Yekutieli, Tomasz Burzykowski, Hinrich W. H. Göhlmann, An De Bondt, Tobias Verbeke, and Luc Bijnens

7.1 Introduction

In Chap. 3, we discussed several testing procedures for testing the null hypothesis of no dose effect against ordered alternatives. Williams (1971, 1972) proposed a step-down procedure to test for the dose effect. The tests are performed sequentially from the comparison between the isotonic mean of the highest dose and the sample mean of the control, to the comparison between the isotonic mean of the lowest dose and the sample mean of the control. The procedure stops at the dose level where the null hypothesis of no dose effect is not rejected. Marcus (1976) proposed a modification of Williams' procedure, in which the sample mean of the control was replaced by the isotonic mean of the control. The likelihood ratio test, discussed by Bartholomew (1961), Barlow et al. (1972), and Robertson et al. (1988), uses the ratio between the

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

Z. Shkedy · T. Burzykowski
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be; Tomasz.burzykowski@uhasselt.be

D. Yekutieli
Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel
e-mail: yekutiel@post.tau.ac.il

H.W.H. Göhlmann · A. De Bondt · L. Bijnens
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: HGOEHLMA@its.jnj.com; ADBONDT@its.jnj.com; LBIJNENS@its.jnj.com

T. Verbeke
OpenAnalytics BVBA, Heist-op-den-Berg, Belgium
e-mail: tobias.verbeke@openanalytics.be

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_7,

variance calculated under the null hypothesis and the variance calculated under an ordered alternative. In the context of dose-response microarray data, Hu et al. (2005) proposed a test statistic that was similar to Marcus' statistic, but with the variance estimator calculated under the ordered alternative. Lin et al. (2007) proposed a modification for Hu's test statistic in which the degrees of freedom for the variance estimator are not fixed for all genes.

The chapter is organized as follows. In Sects. 7.2 and 7.3, we review the M test statistics of Hu et al. (2005) and Lin et al. (2007). Directional inference and the multiplicity issue are discussed in Sect. 7.4. In Sect. 7.5, we show how to use the `IsoGene` package to compute the five test statistics discussed above and how to obtain a list of significant genes using the FDR correction. Finally, in Sect. 7.6, we compare the results of the analysis of the case study using the five tests discussed. In particular, we present the comparison of the likelihood ratio test and the modified M test statistic.

7.2 The M Test Statistic of Hu et al. (2005)

Hu et al. (2005) discussed a setting in which the comparison of primary interest is the difference between the highest dose level (K) and the control dose and proposed a t-type test statistic given by

$$M = \frac{\hat{\mu}_K^\star - \hat{\mu}_0^\star}{\sqrt{\sum_{i=0}^{K}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i^\star)^2/(N-K)}}. \tag{7.1}$$

The numerator of the M test statistic is the same as that of Marcus' statistic, while the denominator is an estimate of the standard error under an ordered alternative. This is in contrast to Williams' and Marcus' approaches that use the unrestricted means to derive the estimate for the standard error.

Hu et al. (2005) evaluated the performance of the $\bar{E}_{01}^2$ and M test statistics by comparing the ranks of genes obtained by using both statistics, and reported similar findings for simulated and real-life data sets.

7.3 A Modification to the M Test Statistic

For the variance estimate, Hu et al. (2005) used $N-K$ degrees of freedom. However, the unique number of isotonic means is not fixed but changes across the genes. For that reason, Lin et al. (2007) proposed a modification to the standard error estimator used in the M statistic by replacing it with $\sqrt{\sum_{i=0}^{K}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i^\star)^2/(N-I)}$, where I is the unique number of isotonic means for a given gene. Such a modification is expected to improve the standard error estimates across all the genes.

7.4 Directional Inference in Isotonic Regression

The five test statistics discussed in Sects. 3.3, 7.2, and 7.3 should be calculated assuming a particular direction of the ordered alternative. However, the direction of the test is unknown in advance. In this section, we address the issue of how to obtain the two-sided p value for the five testing procedures and how to determine the direction of the trend from the two-sided p value.

We focus on the two possible directions of the alternatives: H_1^{Up}, defined in Eq. (3.3) and H_1^{Down}, defined in Eq. (3.4). Let p^{Up} and T^{Up} denote the p value and the corresponding test statistic computed to test H_0 vs. H_1^{Up}, and let p^{Down} and T^{Down} denote the p value and the corresponding test statistic computed to test H_0 vs. H_1^{Down}. Barlow et al. (1972) showed that, for $K > 2$, a $\bar{\chi}^2$ statistic for testing H_0 may actually yield $p^{\text{Up}} < \alpha$ and $p^{\text{Down}} < \alpha$. However, $p = 2\min(p^{\text{Up}}, p^{\text{Down}})$ is always a conservative p value for the two-sided test of H_0 vs. either H_1^{Up} or H_1^{Down}.

Hu et al. (2005) adapted the approach by taking the larger of the likelihoods of H_1^{Up} or H_1^{Down}, i.e., the larger of T^{Up} and T^{Down} as the test statistic for two-sided inference. In contrast to Hu et al. (2005), we obtain two-sided p values by taking $p = \min(2\min(p^{\text{Up}}, p^{\text{Down}}), 1)$, where p^{Up} and p^{Down} are calculated for T^{Up} and T^{Down} using permutations to approximate the null distribution of these test statistics.

After rejecting the null hypothesis against the two-sided test, there is still a need to determine the direction of the trend. The direction can be inferred by the following procedure. If $p^{\text{Up}} \leq \alpha/2$, then reject H_0 and declare H_1^{Up}; if $p^{\text{Down}} \leq \alpha/2$, then reject H_0 and declare H_1^{Down}. The validity of this directional inference is based on the following property: under H_1^{Up}, p^{Down} is stochastically larger than $U[0, 1]$, and under H_1^{Down}, p^{Up} is stochastically larger than $U[0, 1]$. Thus, the probability of falsely rejecting H_0 is $\leq \alpha$, and the probability of declaring a wrong direction for the trend is $\leq \alpha/2$. It is also important to note that the event $p^{\text{Up}} < \alpha/2$ and $p^{\text{Down}} < \alpha/2$ may be observed. Under H_0, H_1^{Up}, or H_1^{Down}, this event is unlikely. However, it is likely if the treatment has a large and non-monotone effect. An example of this unique situation, in which the null hypothesis can be rejected for both directions, is given in Sect. 7.6.1.

In order to verify whether the property needed for directional inference applies to the five test statistics, we conduct a simulation study to investigate the distribution of the p^{Up} and p^{Down} values. For each simulation, data are generated under H_1^{Up}: the means are assumed to be equal to $(1,\ 2,\ 3,\ 4)/\sqrt{5}$ for the four doses, respectively, and the variance is equal to $\sigma^2 = 1$. The test statistics T^{Up} and T^{Down} are calculated for the two possible alternatives H_1^{Up} and H_1^{Down}. Their corresponding p^{Up} and p^{Down} values are obtained using 10,000 permutations.

Figure 7.1 shows the cumulative distribution of p^{Up} and p^{Down}. Clearly, the simulations show that the cumulative distribution of p^{Down} (the p value of the test statistics calculated assuming the wrong direction, dotted line in Fig. 7.1) is stochastically higher than $U[0, 1]$ (solid line in Fig. 7.1), which is the distribution of the p values under the null hypothesis. Moreover, the distribution of p^{Up} (the p value for the test statistics calculated assuming the right direction, dashed

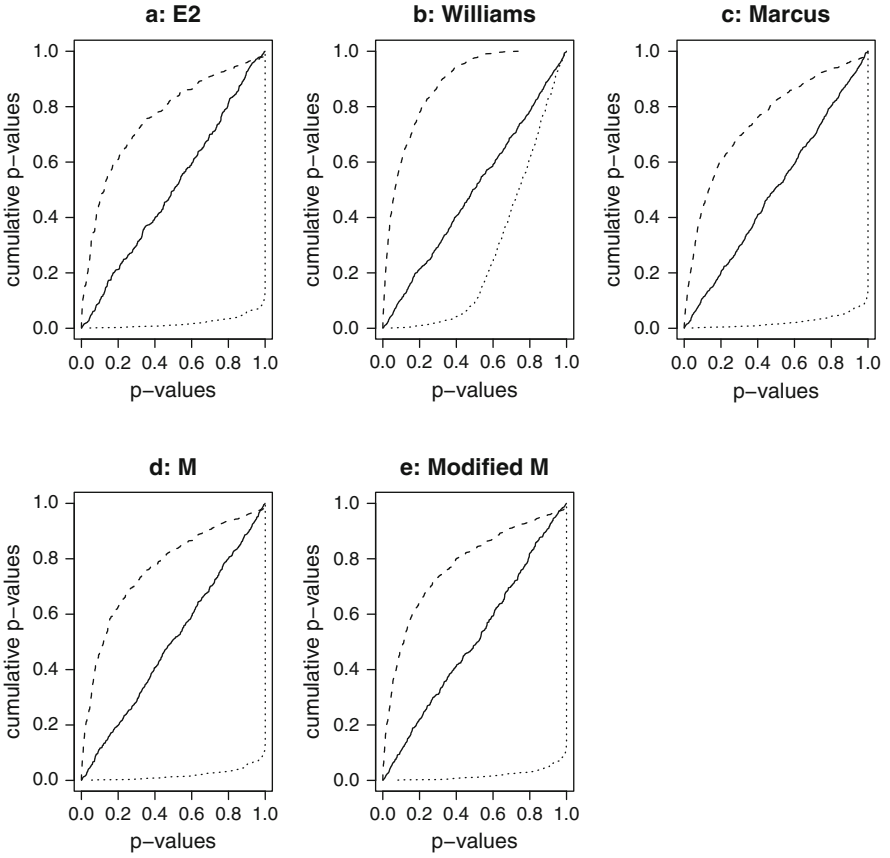

Fig. 7.1 The cumulative distribution of p^{Up} values (*dashed line*) and p^{Down} values (*dotted line*) for the five test statistics. Data are generated under H_1^{Up} with isotonic means $(1, 2, 3, 4)/\sqrt{5}$ for the four doses. *Solid line*: cumulative distribution of $H_0 \sim U[0, 1]$

line in Fig. 7.1) is, as expected, stochastically smaller than $U[0, 1]$. Similar results (not shown) are obtained when the data are generated under H_1^{Down}. The results imply that all the five test statistics possess the property required for the directional inference: under H_1^{Up} the distribution of p^{Down} is stochastically greater than $U[0, 1]$.

Figure 7.2 shows the values of test statistics, which were calculated under H_1^{Up} and H_1^{Down}, for data generated under H_1^{Up}. The five test statistics are calculated for testing H_0 vs. H_1^{Down} (the x-axis of each test statistic in Fig. 7.2). The behavior of Marcus', M, and the modified M statistics is similar, as they all use the difference between the highest and the lowest isotonic means. The maximum value of the test statistics (when calculated assuming the wrong direction) is equal to zero. In contrast, Williams' test statistic for testing H_0 vs. H_1^{Down} (shown on the x-axis of

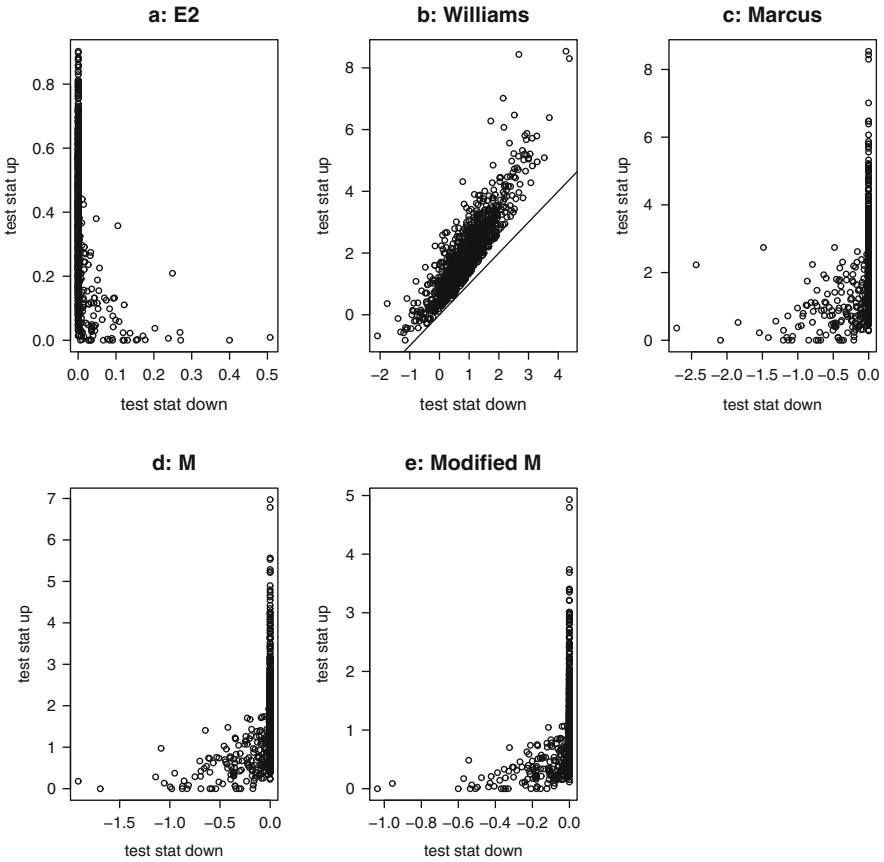

Fig. 7.2 The five test statistics calculated for H_0 vs. H_1^{Up} (y-axis) and H_0 vs. H_1^{Down} (x-axis)

the panel b) can be positive or negative, because the sample mean of control group is used instead of the isotonic mean. A similar pattern was observed in Chap. 3 for the bootstrap distribution of Williams' and Marcus' statistics. Note that we reject the null hypothesis in favor of H_1^{Down} for negative values of the test statistic. Further, the value of the test statistics for testing H_0 vs. H_1^{Up} (the y-axis of Fig. 7.2) is higher than the value of the test statistics calculated for testing H_0 vs. H_1^{Down} (the x-axis of Fig. 7.2).

7.4.1 Control of the Directional FDR

When FDR controlling procedures are used to adjust for multiplicity in the microarray setting, the set of two-sided p values computed for each gene is adjusted

by using the BH-FDR or BY-FDR procedures described in Sect. 6.3. A discovery in this case is a rejection of H_0 for some gene; a false discovery is to reject H_0 when H_0 is true. As mentioned before, in a microarray dose-response experiment, we are also interested in the direction of the dose-response trend.

Benjamini and Yekutieli (2005) provided a framework for addressing the multiplicity problem when attempting to determine the direction of multiple parameters: a discovery is to declare the sign of a parameter as either being positive or negative. Three types of false discoveries are possible: declaring a zero parameter either as negative or as positive, declaring a negative parameter as positive, and declaring a positive parameter as negative. The FDR corresponding to these discoveries is termed the mixed directional FDR (MD-FDR). In the current setting, the MD-FDR is the expected value of the number of genes, for which H_0 is true, that are erroneously declared to have either a positive or negative trend plus the gene that have a monotone trend but the direction of the declared trend is wrong, divided by the total number of genes declared to have a trend. Benjamini and Yekutieli (2005) proved that if p values poses the directional property described in Sect. 7.4, then applying the BH procedure at level q to the set of two-sided p values computed for each gene, and declaring the direction of the trend corresponding to the smaller one-sided p value, controls the MD-FDR at level $q/2 \cdot (1 + m_0/m)$, where m is the total number of genes and m_0 is the number of genes, for which H_0 holds.

In general, directional inference is a more general setting than hypothesis testing (Benjamini and Yekutieli 2005). Nevertheless, as a false discovery is made based on the p value that is stochastically larger than $U[0, 1]$, then the resampling-based methods that control the FDR (Yekutieli and Benjamini 1999) also control the MD-FDR. This is achieved by simply applying the resampling-based procedure to test H_0, and if H_0 is rejected, declaring the direction of the trend according to the minimum one-sided p value. For each rejected null hypothesis, it is also advisable to examine if the larger p value is $\leq\alpha$. If this is the case, this may serve as an indication of a non-monotone dose-response relationship.

7.5 Application Using the `IsoGene` Package

7.5.1 A Quick Start

The first stage of the analysis, which is also the time-consuming stage, consists of permutations under the null hypothesis in order to obtain the distribution of the test statistic under the null hypothesis. Note that, by default, all five test statistics discussed above are calculated. The function `IsoRawp()` is used to perform the permutation. A general call of the function `IsoRawp()` has the form of

```
IsoRawp(x, data.frame, niter=1000)
```

Here, x is a vector which contains the dose levels, and data is the R data frame which contains the information about gene expression and gene names. Once the permutation stage is completed, the FDR-adjusted p values can be obtained using the function IsoTestBH(). The function calculates the adjusted p values for each statistic using either the BH-FDR or BY-FDR for multiplicity adjustment. The user can specify one of the five test statistics discussed above or use the default call; in the latter case, adjusted p values for all test statistics will be calculated. A typical call of the IsoTestBH() function has the form:

```
IsoTestBH(rawp, FDR, type,stat)
```

The first argument rawp is an R object which contains all the output produced by the function IsoRawp(). The argument FDR and type are vectors used to specify the error rate and the adjustment type (BH or BY or both). The last argument stat is a vector in which we specify the test statistics to be used. For example, a simple call of IsoTestBH()

```
IsoTestBH(rp, FDR=0.05, type=c("BH","BY"),
stat=c("E2","Williams","Marcus","M","ModifM"))
```

will produce the adjusted p values for BH-FDR and BY-FDR at levels of 5% for all test statistics. In what follows, we illustrate in more details the use of the functions in the IsoGene package.

7.5.2 Calculating the Test Statistics

The five test statistics described above can be obtained by using the function IsoGene1() (for a single gene) and the function IsoGenem() (for all the genes simultaneously). We can calculate the test statistics with the call

```
> IsoGene1(dose levels, gene expression vector)
```

For example, the test statistics for gene 1 from the human epidermal carcinoma case study can be calculated using the following code

```
> stat1 <- IsoGene1(x, gene1)
```

The object stat1 contains the information about the five test statistics and the direction for which the likelihood is maximized:

```
> stat1
$ E2.up [1] 0.2010957
$ Williams.up [1] 0.6712589
$ Marcus.up [1] 1.3646790
$ M.up [1] 1.0004640
$ ModM.up [1] 1.0611520
$ E2.dn [1] 0.0098105
$ Williams.dn [1] -0.2850243
$ Marcus.dn [1] -0.2850243
$ M.dn [1] -0.1876899
$ ModM.dn [1] -0.2098437
$ direction [1] "u"
```

The first ten objects are the values calculated for the five test statistics under increasing and decreasing trends. The last object indicates the higher likelihood of isotonic regression with "u" meaning an increasing trend or "d" meaning a decreasing trend.

In a similar way, the test statistics can be calculated for the first ten genes using the function `IsoGenem()`:

```
> statm <- IsoGenem(x, data[1:10,])
```

```
> statm
$E2.up
          2           3           4           5           6
0.00000000 0.20109571 0.50774178 0.24835414 0.00000000
          7           8           9          10          11
0.16263545 0.43080221 0.00000000 0.06367646 0.00000000
$Williams.up
 [1] -1.1298186  0.6712589  2.4888850  0.8911883 -0.6520746
 [6]  0.5582301  2.1412458 -0.5774471  0.8895008 -1.6655641
$Marcus.up
 [1] 0.0000000 1.3646791 2.4888850 1.3929232 0.0000000
 [6] 1.4262255 2.1412458 0.0000000 0.8895008 0.0000000
$M.up
 [1] 0.0000000 1.0004635 2.0194520 0.9386711 0.0000000
 [6] 0.7196721 1.6954778 0.0000000 0.4960717 0.0000000
$ModM.up
 [1] 0.0000000 1.0611518 2.1419523 1.0494662 0.0000000
 [6] 0.8046179 1.7983258 0.0000000 0.5261635 0.0000000
$E2.dn
           2           3           4           5           6
0.275992755 0.009810531 0.000000000 0.002919082 0.139779913
           7           8           9          10          11
0.079338232 0.000000000 0.099457466 0.124447675 0.456682221
$Williams.dn
 [1] -1.66138158 -0.28502430  1.72265475  0.06394548
 [5] -1.10255870 -1.01756814  1.31114235 -0.76992952
 [9]  0.12749125 -3.83085270
$Marcus.dn
 [1] -1.6613816 -0.2850243  0.0000000 -0.1743749 -1.1025587
 [6] -1.1502473  0.0000000 -0.7699295 -1.0674972 -3.8308527
$M.dn
 [1] -1.2705909 -0.1876899  0.0000000 -0.1020262 -0.9002354
 [6] -0.5535348  0.0000000 -0.6266432 -0.6156540 -2.0821512
$ModM.dn
 [1] -1.3476651 -0.2098437  0.0000000 -0.1140688 -0.9002354
 [6] -0.6188707  0.0000000 -0.7006084 -0.6883221 -2.2084548
$direction
  2   3   4   5   6   7   8   9  10  11
"d" "u" "u" "u" "d" "u" "u" "d" "d" "d"
```

The output from `IsoGenem` has the same structure as the one for the `IsoGene1`, but each object contains the values of the test statistics and the likely direction of the isotonic regression model for all the genes.

7.5.3 Obtaining Raw p Values

As discussed above, we use permutations to obtain the raw p values for the five test statistics:

```
> set.seed(1234)
> rawp <- IsoRawp(x=x, y=data, niter=1000)
```

The R object `rawp` contains four objects with p values for the five test statistics: the first one contains the two-sided p values, the second contains the one-sided p values, the third contains p^{Up} values, and the last one contains p^{Down} values. Below we print a part of the object with two-sided p values for illustration:

```
> rawp.twosided=rawp[[2]]
```

The first ten rows in of the object `rawp.twosided` are

```
> rawp.twosided[1:10,]
   Probe.ID      E2  Williams  Marcus  M   ModM
1  31637_s_at 0.000 0.000  0.000 0.000 0.000
2  32402_s_at 0.129 0.225  0.124 0.123 0.125
3  33646_g_at 0.003 0.004  0.003 0.003 0.002
4  34063_at    0.487 0.379  0.500 0.467 0.474
5  33494_at    0.071 0.185  0.035 0.063 0.064
6  34031_i_at 0.082 0.220  0.086 0.103 0.086
7  34449_at    0.357 0.445  0.432 0.400 0.391
8  34478_at    0.472 0.516  0.535 0.518 0.511
9  35436_at    0.151 0.116  0.148 0.150 0.140
10 36711_at    0.000 0.000  0.000 0.000 0.000
```

The second output object from the `rawp` is a matrix with six columns, where the first column indicates the probe ID. Columns from the second to the sixth are p values for each of the five test statistics, respectively. The remaining three output objects (`rawp[[1]]`, `rawp[[3]]`, `rawp[[4]]`) are structured in the same way.

7.5.4 Plot of p Values for a Single Gene

For a single gene, the function `IsopvaluePlot()` can be used to show the p^{Up}- and p^{Down} values for a given test statistic:

```
IsopvaluePlot(x, y, niter,stat = c("E2", "Williams",
"Marcus", "M", "ModifM"))
```

We use one gene as an example to illustrate how p^{Up} and p^{Down} values (in the upper and lower panels of Fig. 7.3) are obtained. In Fig. 7.3, the observed test statistics are drawn as the dashed line, and the values of the test statistics obtained from permutations are spread over the x-axis. For this gene, the p^{Up} is much smaller as compared to the p^{Down} since $T^{\text{Up}} \gg T^{\text{Down}}$, which implies a possible increasing trend in the data:

```
> gene2 <- data[2,]
> set.seed(123)
> IsopvaluePlot(x, gene2, niter=1000, stat="E2")
```

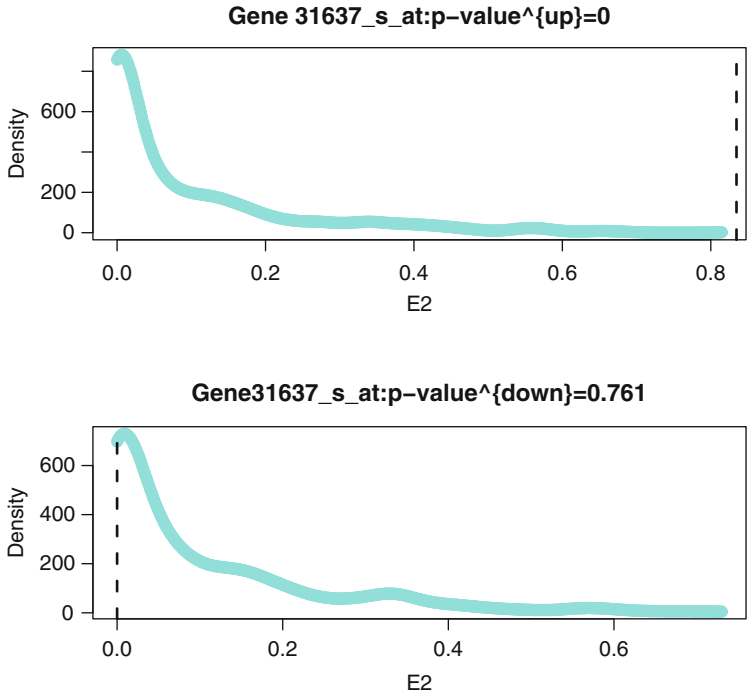

Fig. 7.3 The p^{Up} and p^{Down} values using $\bar{E}^2_{01}$ for an example gene. The *dashed line* is the observed test statistic value. In the *upper panel*, the *dashed line* (at the right) is larger than most of the test statistics from permutations, which results in a small p^{Up} value. In the *lower panel*, the *dashed line* (close to zero) is smaller than most of the test statistics from permutations, which results in a large p^{Down} value

7.5.5 BH-/BY-FDR Procedures for Adjusting for Multiple Testing

With the two-sided p values, we now need to select one of the five test statistics, the FDR level, and the type of multiplicity adjustment (BH-FDR or BY-FDR) to obtain the list of significant genes. The following example shows the use of the likelihood ratio test $\bar{E}^2_{01}$, the FDR level of 0.05, and the BH-FDR procedure controlling the FDR:

```
> E2.BH <- IsoTestBH(rawp.twosided, FDR=0.05, type="BH", stat="E2")
```

The first ten rows in the R object `E2.BH` list the sorted raw and adjusted p values for the $\bar{E}^2_{01}$ test statistic:

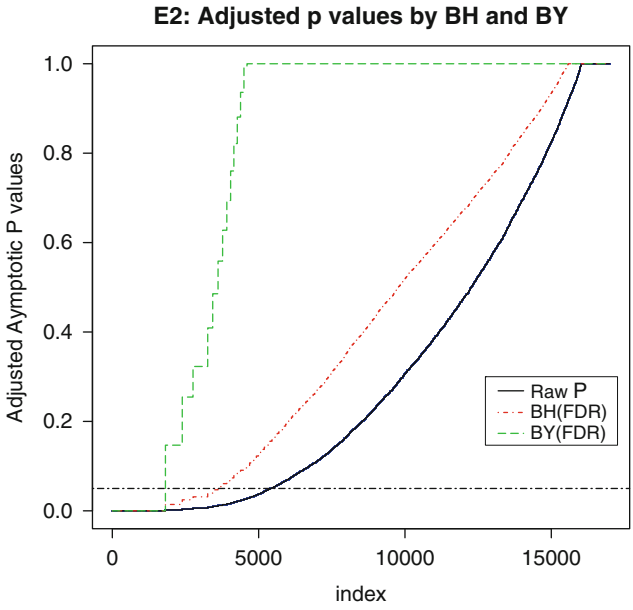

Fig. 7.4 The unadjusted (*solid line*) and the BH-FDR (*dotted* and *dashed red line*) and BY-FDR (*dashed line*) adjusted p values for the $\bar{E}_{01}^2$

```
> E2.BH[1:10,]
    Probe.ID    row.name raw p-values BH adjusted p values
1  31637_s_at    1        0.000          0.000000000
2  33646_g_at    3        0.003          0.015647131
3  36711_at      10       0.000          0.000000000
4  37079_at      12       0.001          0.007115111
5  37117_at      13       0.011          0.042679297
6  37152_at      14       0.003          0.015647131
7  38158_at      29       0.003          0.015647131
8  38241_at      30       0.000          0.000000000
9  39248_at      35       0.003          0.015647131
10 39249_at      36       0.008          0.033642751
```

Note that order of the list of genes found significant is based on the row number. Moreover, the function `IsoBHPlot()` can be used to plot the raw and adjusted p values in order to visualize the number of significant findings by using the BH-FDR and BY-FDR procedures for the specified test statistic. A general call of `IsoBHPlot()` has the same structure as the function `IsoTestBH()`:

```
IsoBHPlot(rp, FDR=c(0.05,0.1),
stat=c("E2","Williams","Marcus","M","ModifM"))
```

Figure 7.4 shows the unadjusted (solid line) and the BH-FDR (dotted and dashed line) and BY-FDR (dashed line) adjusted p values for $\bar{E}_{01}^2$. It is obtained using the following call for the function `IsoBHPlot()`:

```
> IsoBHPlot(rawp.twosided, FDR=0.05, stat="E2")
```

7.6 Results

In this section, we present results of an application of the five testing procedures to the human epidermal carcinoma case study. We compare the performance of each of the five test statistics in combination with the Bonferroni, Holm, the maxT, and the BH-FDR multiple testing adjustment procedures. In Sect. 7.6.1, we examine the number of significant genes for all the testing procedures. In Sect. 7.6.2 we make a comparison between the global likelihood ratio test $\bar{E}_{01}^2$ and the two t-type test statistics: M and the modified M. The complete R code which was used to conduct the analysis, is given below. Raw p values were obtained with the `IsoGene` package, while the adjusted p values were obtained using the function `p.adjust()`:

```
> seed <- 1234
> pval.maxT <- IsomaxT(x=x, y=data,niter=1000)
> adjp.Bonf <- apply(rawp.twosided,2,
+ function(x)p.adjust(x,method="bonferroni"))
> adjp.Holm <- apply(rawp.twosided,2, function(x)p.adjust(x,method="holm"))
> adjp.BH <- apply(rawp.twosided,2, function(x)p.adjust(x,method="BH"))
> adjp.BY <- apply(rawp.twosided,2, function(x)p.adjust(x,method="BY"))
```

7.6.1 *Number of Significant Findings for each Statistic Using Different Multiple Testing Adjustments*

The testing procedures discussed in the previous sections are applied to the case study data. For each test statistic, p^{Up} and p^{Down} are obtained based on the permutation matrix, in which the null distribution of the test statistics T^{Up} and T^{Down}, respectively, is approximated using 1,000 permutations. The inference is made based on the two-sided p values. Table 7.1 shows the number of rejected hypotheses for several multiplicity adjustment methods and for the five test statistics that are tested at the significance level of 0.05. Figure 7.5 shows the adjusted p values for the five test statistics. Clearly, the adjusted p values for the maxT, Bonferroni, and BY-FDR are larger than the adjusted p values obtained for the BH-FDR. For instance, for the $\bar{E}_{01}^2$, without adjusting for multiple testing, we reject

Table 7.1 Number of rejected null hypotheses for various testing procedures at the significance level of 0.05

Method	$\bar{E}_{01}^2$	Willams	Marcus	M	Modified M
Unadjusted	5,457	5,238	5,465	5,449	5,451
maxT	224	215	223	265	251
Bonferroni	1,814	1,592	1,669	1,755	1,745
Holm	1,814	1,592	1,669	1,755	1,745
BH-FDR	3,613	3,209	3,533	3,562	3,567
BY-FDR	1,814	1,592	1,669	1,755	1,745

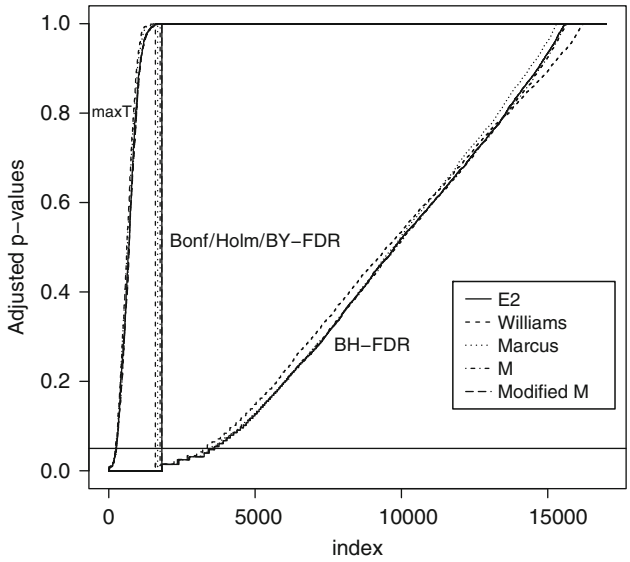

Fig. 7.5 Adjusted p values using the Bonferroni, BH(FDR), and maxT procedures for the five test statistics

the null hypothesis for 5,457 genes. With Bonferroni, Holm, and the BY-FDR adjustment procedures, we obtain the same number of significant genes, i.e., 1,814. Using the maxT for controlling the FWER seems to be the most conservative approach with only 224 genes declared significant.

Note that due to the incidence that a different random permutation seed is used for the analysis of $\bar{E}_{01}^2$ for the human epidermal carcinoma case study, 3,499 genes are obtained as compared to 3,613 obtained as above, which will be used throughout the remaining chapters of the book:

```
> rbind(colSums(rawp.twosided[,-1]<=0.05),colSums(pval.maxT[,-1]<=0.05),
                                    colSums(adjp.Bonf[,-1]<=0.05),
                                    colSums(adjp.Holm[,-1]<=0.05)
                                    colSums(adjp.BH[,-1]<=0.05)
                                    colSums(adjp.BY[,-1]<=0.05))
```

Note that the number of significant genes obtained for each test statistic for a given multiple testing adjustment is similar. For example, for the BH-FDR adjustment, we find 3,613, 3,562, and 3,567 significant genes for $\bar{E}_{01}^2$, M, and the modified M statistic, respectively. This method of adjustment for multiple testing yields more liberal results as compared to the other methods. For that reason, the FDR adjustment for multiplicity is commonly used within the microarray framework (Ge et al. 2003; Tusher et al. 2001; Storey and Tibshirani 2003). Moreover, the BH-FDR procedure controls the MD-FDR, as discussed in Sect. 7.4.1. Therefore, in what follows, we use the BH-FDR procedure to investigate the performance of the considered test statistics.

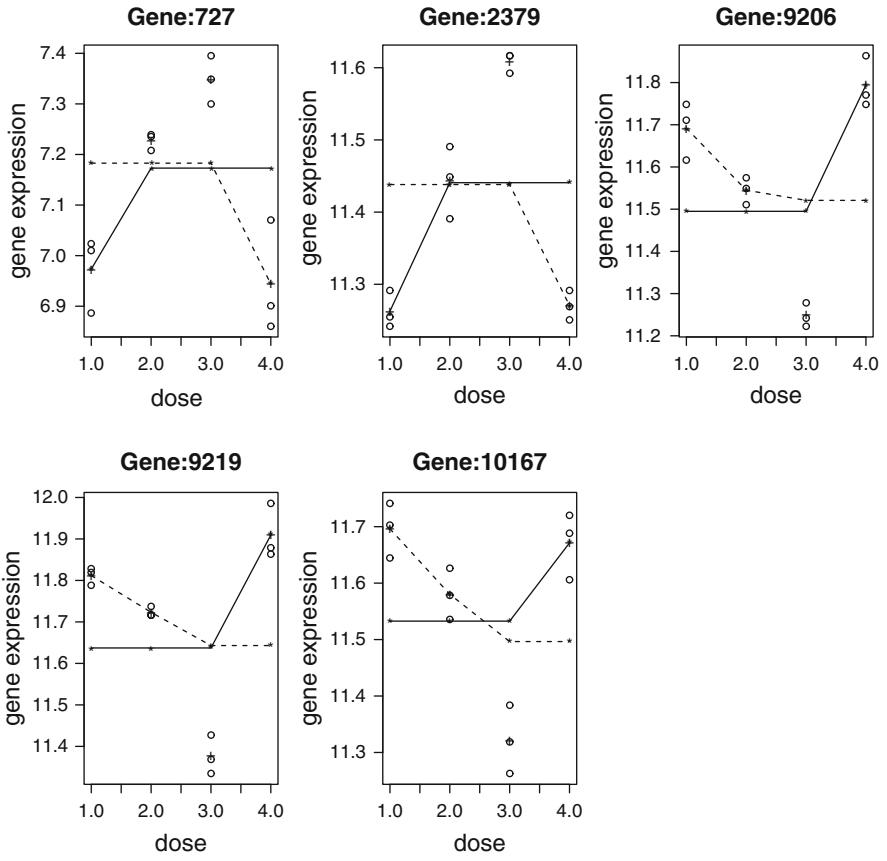

Fig. 7.6 Five genes rejected by Marcus' statistics with both p^{Up} and p^{Down} values smaller than the rejection threshold. *Solid line*: the isotonic means obtained for testing H_0 against H_1^{Up}. *Dashed line*: the isotonic means obtained for testing H_0 against H_1^{Down}

As we argue in Sect. 7.4, there is a possibility, although unlikely, that the null hypothesis is rejected for both directions, i.e., $p^{\text{Up}} \leq \alpha/2$ and $p^{\text{Down}} \leq \alpha/2$. For the analysis discussed above, the null hypothesis is rejected for only five genes when using Marcus' statistic with p^{Up} and p^{Down} smaller than the rejection threshold (with multiple testing adjustment), suggesting a non-monotonic trend. The five genes are shown in Fig. 7.6.

For Marcus' statistic, the large values of T^{Up} and T^{Down} are obtained from the large difference between the isotonic mean of the highest and control doses, relative to the variance calculated under the unrestricted alternative. Instead, $\bar{E}_{01}^2$, M, and the modified M use the variance estimator calculated under the order alternative, that results in smaller test statistic values. Hence, using these test statistics, the null hypothesis is not rejected. If the difference between the highest isotonic mean and control sample mean exists, Williams' test statistic will tend to reject the null hypothesis as well.

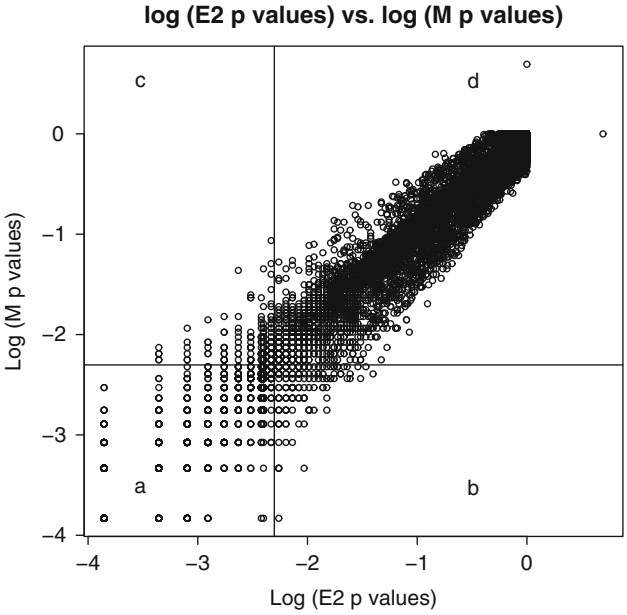

Fig. 7.7 Logarithm of p values (two-sided) for $\bar{E}^2_{01}$ and M. Panel (**a**): 3,420 genes rejected by both $\bar{E}^2_{01}$ and M statistics; panel (**b**): 142 genes are rejected by M statistic only; panel (**c**): 193 genes in are rejected from $\bar{E}^2_{01}$ only; panel (**d**): 13,244 genes are not rejected by either statistic

In particular, for the five genes, the estimates of σ^2 for Williams' and Marcus' test statistic calculated under the unrestricted alternatives are equal, respectively, to 0.0414, 0.0075, 0.0204, 0.0145, and 0.0232. They are smaller than the estimates for σ^2 for $\bar{E}^2_{01}$, M, and the modified M procedures calculated under the ordered alternative H_1^{Up}, which are equal, respectively, to 0.2995, 0.1788, 0.3277, 0.3317, and 0.2437, and under H_1^{Down}, which are equal, respectively, to 0.2608, 0.1868, 0.4679, 0.4401, and 0.2065.

7.6.2 Comparison Between $\bar{E}^2_{01}$, M, and the Modified M Test Statistics

Although in our case study, the number of significant genes obtained for the five testing procedures is very similar, there are some discrepancies. In this section, we investigate the subset of genes not commonly found by $\bar{E}^2_{01}$, M, and the modified M statistics.

We first compare genes identified as significant or nonsignificant by M and $\bar{E}^2_{01}$. The logarithm of two-sided p values for these genes is shown in Fig. 7.7. Among the total of 16,998 genes, 3,420 genes are found significant for monotonic trends

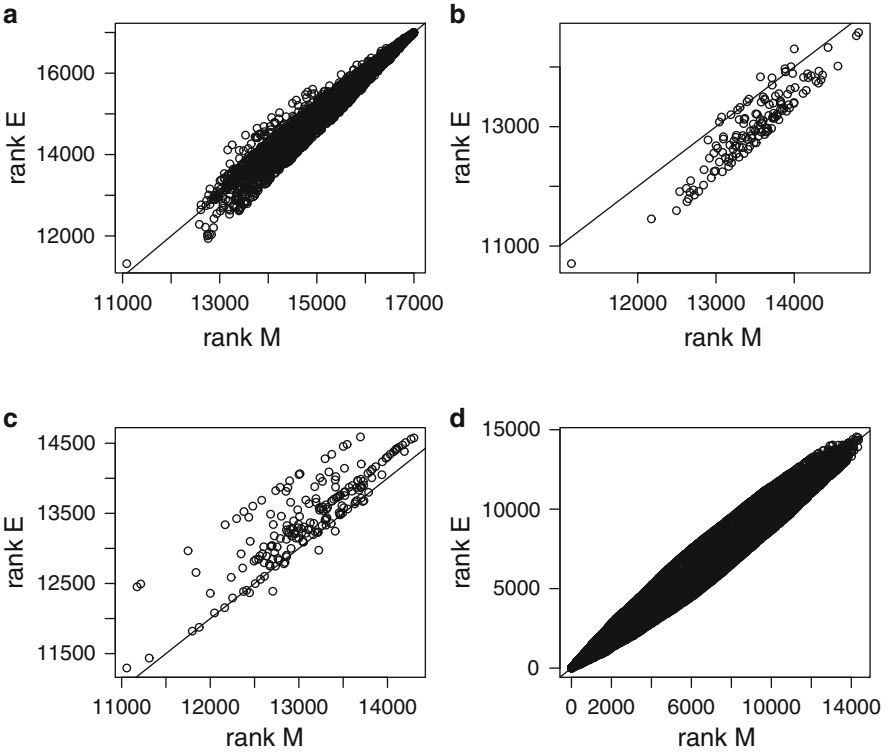

Fig. 7.8 Correlation between $\bar{E}^2_{01}$ and M. Panel (**a**): correlation (0.98) between rankings of 3,420 genes rejected both from $\bar{E}^2_{01}$ and M. Panel (**b**): correlation (0.92) between rankings of 142 genes rejected only from M. Panel (**c**): correlation (0.85) between rankings of 193 genes rejected only from $\bar{E}^2_{01}$ and panel (**d**): correlation (0.99) between rankings of 13,244 genes not rejected from $\bar{E}^2_{01}$ and M

for both statistics. However, 193 genes are found to be significant for $\bar{E}^2_{01}$ and nonsignificant for M-test statistic, while for 142 genes, the reversed order is observed. These genes account for 8.9% $((193 + 142)/(3{,}420 + 193 + 142))$ of the total significant findings for both test statistics, which is not negligible.

Similar to Hu et al. (2005), we compare the ranking of M and $\bar{E}^2_{01}$ of all the genes. In both Hu et al. (2005) and our example, the correlation of the ranks is equal to 0.99. Based on this observation, Hu et al. (2005) concluded that the two statistics perform similarly. However, in our data, the correlation of ranks of 142 genes found significant only for the M statistic (panel c of Fig. 7.8) is 0.92, while the correlation of ranks of 193 genes significant only for $\bar{E}^2_{01}$ (panel b) is 0.85. Both are somewhat lower than the correlation for genes in panel a (3,420 genes significant for both statistics, correlation of 0.98) and in panel d (genes nonsignificant by either statistic, correlation of 0.99). The discrepant conclusions (rejecting the null only for

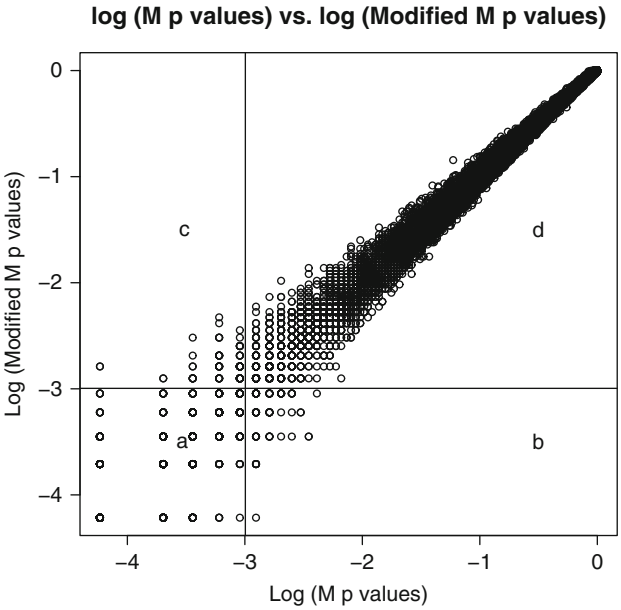

Fig. 7.9 Logarithm of p values (two-sided) for the M and the modified M. Panel (**a**): 3,478 genes are rejected by both M and the modified M statistics; panel (**b**): 86 genes are rejected by M statistic only; panel (**c**): 89 genes are rejected from the modified M only; panel (**d**): 13,345 genes are not rejected by either statistic

one of statistics) can be explained by the fact that the M statistic looks for the mean difference between the highest dose and the control. On the other hand, $\bar{E}_{01}^2$ is a global test for the monotonic trend.

The logarithm of the two-sided p values for the genes identified as significant or nonsignificant by M and the modified M statistics is shown in Fig. 7.9. Among the total of 16,998 genes, 3,478 genes are found significant for monotonic trends for both tests. However, 86 genes are found to be significant for M statistic and non-significant for the modified M test, while for 89 genes, the reverse is true. These genes account for about 4.8% $((86+89)/(86+89+3,478))$ of the total significant findings for both test statistics.

The overall correlation between the ranks of genes obtained for M and the modified M test statistics is 0.99. The correlation between genes in each panel of Fig. 7.10 is also very high, with 0.97 (in panel b) for genes declared significant only by the modified M, 0.98 (in panel c) for genes declared significant only by M, 0.99 (in panel a) for genes declared significant by both of the test statistics, and 0.998 (in panel c) for genes declared significant by neither of the test statistics. The difference between the two statistics lies in the adjustment of the degrees of freedom in the standard error estimator of the modified M test statistic. Nevertheless, the discrepancy is not substantial.

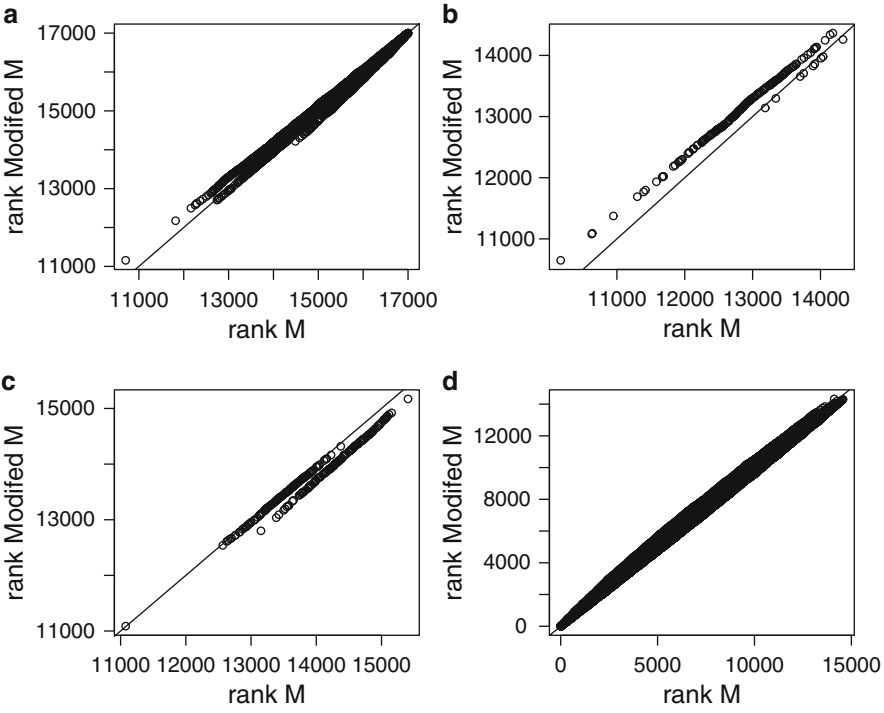

Fig. 7.10 Correlation between M and the modified M. Panel (**a**): correlation (0.99) between rankings of 3,478 genes rejected both from M and the modified M. Panel (**b**): correlation (0.97) between rankings of 89 genes rejected only from the modified M. Panel (**c**): correlation (0.98) between rankings of 86 genes rejected only from M and panel (**d**): correlation (0.998) between rankings of 13,345 genes not rejected from M and the modified M

7.7 Discussion

In this chapter, we introduced another two t-type test statistics to test for the monotone trends of gene expression with respect to doses, namely, M and the modified M tests, which complements Williams' and Marcus' tests discussed in Chap. 3. Further we discussed the issue of directional FDR in order-restricted inference in the dose-response microarray setting. Different approaches for directional FDR are discussed in Guo et al. (2010) in the context of ordered categorical predictor and in Sun and Wei (2011) in the context of multiple testing for pattern identification for time course data. This chapter also introduced the use of the `IsoGene` package to implement the various test statistics aforementioned, the resampling-based inference, and multiple testing adjustment procedures, such as the BH-FDR and BY-FDR procedures.

References

Barlow, R. E., Bartholomew, D. J., Bremner, M. J., & Brunk, H. D. (1972). *Statistical inference under order restriction*. New York: Wiley.

Bartholomew, D. J. (1961). Ordered tests in the analysis of variance. *Biometrika, 48*, 325–332.

Benjamini, Y., & Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association, 100*, 71–81

Ge, Y., Dudoit, S., & Speed, P. T. (2003). *Resampling based multiple testing for microarray data analysis* (Technical report, 633). Berkeley: University of Berkeley.

Guo, W., Sarkar, S., & Peddada, S. (2010). Controlling false discoveries in multidimensional directional decisions, with applications to gene expression data on ordered categories. *Biometrics, 66*(2), 485–492.

Hu, J., Kapoor, M., Zhang, W., Hamilton, S. R., & Coombes, K. R. (2005). Analysis of dose response effects on gene expression data with comparison of two microarray platforms. *Bioinformatics, 21*(17), 3524–3529.

Lin, D., Shkedy, Z., Yekutieli, D., Burzykowki, T., Göhlmann, H. W. H., De Bondt, A., et al. (2007). Testing for trend in dose-response microarray experiments: comparison of several testing procedures, multiplicity, and resampling-based inference. *Statistical Application in Genetics and Molecular Biology, 6*(1). Article 26.

Marcus, R. (1976). The powers of some tests of the quality of normal means against an ordered alternative. *Biometrika, 63*, 177–83.

Storey, J. D., & Tibshirani R. (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In The Analysis of Gene Expression Data: Methods and Software, by G Parmigiani, ES Garrett, RA Irizarry and SL Zeger (editors). Springer, New York.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Sun, W., & Wei, Z. (2011). Multiple testing for pattern identification, with applications to microarray time-course experiments. *Journal of the American Statistical Association, 106*(493), 73–88.

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrys applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences, 98*, 5116–5121.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics, 27*, 103–117.

Williams, D. A. (1972). The comparision of several dose levels with a zero dose control. *Biometrics, 28*, 519–531.

Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference, 82*, 171–196.

Chapter 8
Significance Analysis of Dose-Response Microarray Data

Dan Lin, Ziv Shkedy, Hinrich W. H. Göhlmann, An De Bondt, Luc Bijnens, Dhammika Amaratunga, and Willem Talloen

8.1 Introduction

In Sect. 6.5.4, we discussed the use of the significance analysis of microarrays (SAM) procedure to detect differentially expressed genes in a two-sample setting. In this chapter, we adapt the SAM procedure to the test of monotonic trends in the dose-response microarray setting. As in the previous chapter, we focus on two types of test statistics: the t-type test statistics and the likelihood ratio test. The regularized test statistic in the SAM for Williams' test statistic is defined as

$$t_i^{\text{SAM}} = \frac{\mu_i^* - \bar{y}_0}{s + s_0} = \frac{\hat{\Lambda}_i}{s + s_0}, \tag{8.1}$$

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

Z. Shkedy
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be

H.W.H. Göhlmann · An. De Bondt · L. Bijnens · W. Talloen
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: HGOEHLMA@its.jnj.com; ADBONDT@its.jnj.com; LBIJNENS@its.jnj.com; WTALLOEN@its.jnj.com

D. Amaratunga
Biostatistics and Programming, Jansenn Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ, USA
e-mail: DAMARATU@its.jnj.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_8,

where μ_i^* is the isotonic mean at dose i ($i = 1, \ldots, K$), $\bar{x}_0$ is the sample mean at dose zero, and

$$s = \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_0}\right) \sum_{i=0}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \frac{1}{N-K}},$$

The regularized test statistics for Marcus, M, and the modified M test statistics can be obtained in the same way.

The likelihood ratio test statistic, which is the ratio between the variance under the null and the ordered hypotheses, respectively, is regularized in a similar way as the SAM F statistic (Chu et al. 2011). The modified test statistic is given by

$$\bar{E}_{01}^{2\text{SAM}} = \frac{\sqrt{\hat{\sigma}_{H_0}^2 - \hat{\sigma}_{H_1}^2}}{\sqrt{\hat{\sigma}_{H_0}^2} + s_0}. \tag{8.2}$$

Note that in both (8.1) and (8.2) s_0 is the fudge factor which is estimated from the data as discussed in Sect. 6.3. Based on the regularized SAM test statistics, the SAM procedure can be carried out to find significant genes with monotonic increasing/decreasing trends while controlling the FDR empirically. Before discussing the implementation of the SAM procedure within the `IsoGene` package in Sect. 8.4, we discuss in the following section the main ideas behind the SAM procedure, while in Sect. 8.3, we discuss the choice of the fudge factor in (8.1) and (8.2).

8.2 Graphical Interpretation of the SAM Statistic

To show the effect of the SAM test statistic, we compare the values of the usual t-test statistics and the SAM test statistics and investigate how the SAM test statistic values are affected by the fudge factor. We use the human epidermal squamous carcinoma cell line data to illustrate the change in the modified M test statistics by adding the fudge factor. Figure 8.1 shows the effect size (numerator of t-test statistics) vs. the absolute values of the SAM t-test statistics without (Fig. 8.1a) and with the fudge factor (Fig. 8.1b). We observe that a large number of genes have large test statistic values with small effect sizes, which are represented by points lying along the zero vertical line (Fig. 8.1a). With the introduction of the fudge factor, the points are gathering more around the zero crossing point (Fig. 8.1b). However, the values of test statistics for all the genes decrease simultaneously. Figure 8.2 illustrates how the fudge factor affects genes with different variances.

The two axes of Fig. 8.2a represent the numerator (absolute value of effect size) and denominator (standard error) of the t-test statistics. The angle α (between the y-axis and the solid line) for the three genes using small, median, and large standard

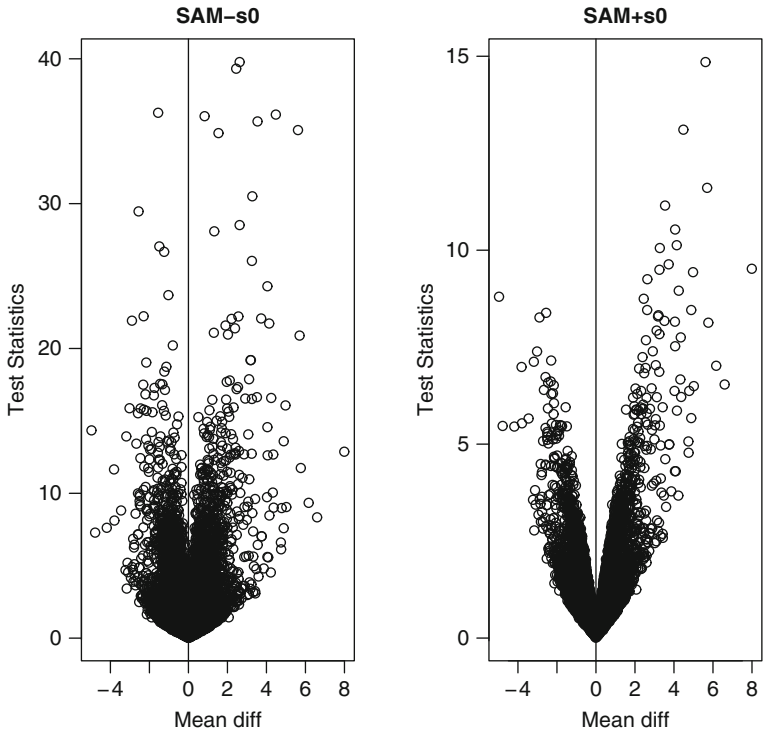

Fig. 8.1 Comparison of the SAM test statistics for the modified M test (absolute values) without the fudge factor (*left panel*: SAM $- s_0$) and with the fudge factor (*right panel*: SAM $+ s_0$) using the case study data. $s_0 = 0.218$ (45th percentile of the standard errors in the data, see Sect. 8.3)

errors ($s_1 < s_2 < s_3$) and corresponding effect sizes ($\Lambda_1 < \Lambda_2 < \Lambda_3$) constitutes the same value of the t-test statistic, i.e., $t_1 = \Lambda_1/s_1 = t_2 = \Lambda_2/s_2 = t_3 = \Lambda_3/s_3$. Note that the test statistic value for these three genes is equal to $\cot(\alpha)$. When the fudge factor s_0 is added in the denominator (extending the standard errors, s_1, s_2, and s_3 by s_0, respectively), the new angles are formed by increasing β_1, β_2, and β_3 on the basis of α, respectively. The three newly formed angles are provided between the y-axis and the dotted line ($\alpha + \beta_1$), short dashed line ($\alpha + \beta_2$), and the long dashed line ($\alpha + \beta_3$), respectively. Thus, the SAM test statistics for the three genes become $\cot(\alpha + \beta_1)$, $\cot(\alpha + \beta_2)$, and $\cot(\alpha + \beta_3)$, respectively. The values of the SAM test statistics are illustrated by the cotangent function in Fig. 8.2b. The left panel of Fig. 8.2b shows the same t-test statistic value of the three genes with angle α. However, the introduction of the SAM fudge factor decreases the values of the SAM test statistics for three genes simultaneously, in particular, $t_1^{\text{SAM}} < t_2^{\text{SAM}} < t_3^{\text{SAM}}$ (see the right panel of Fig. 8.2b) due to $s_1 < s_2 < s_3$.

Let $s_{(1)}, s_{(2)}, \ldots, s_{(m)}$ be the order statistics of the standard error in the microarray experiment with m genes. Let $s^{(0)}, s^{(1)}, s^{(l)}, \ldots, s^{(100)}$ be the lth percentile of

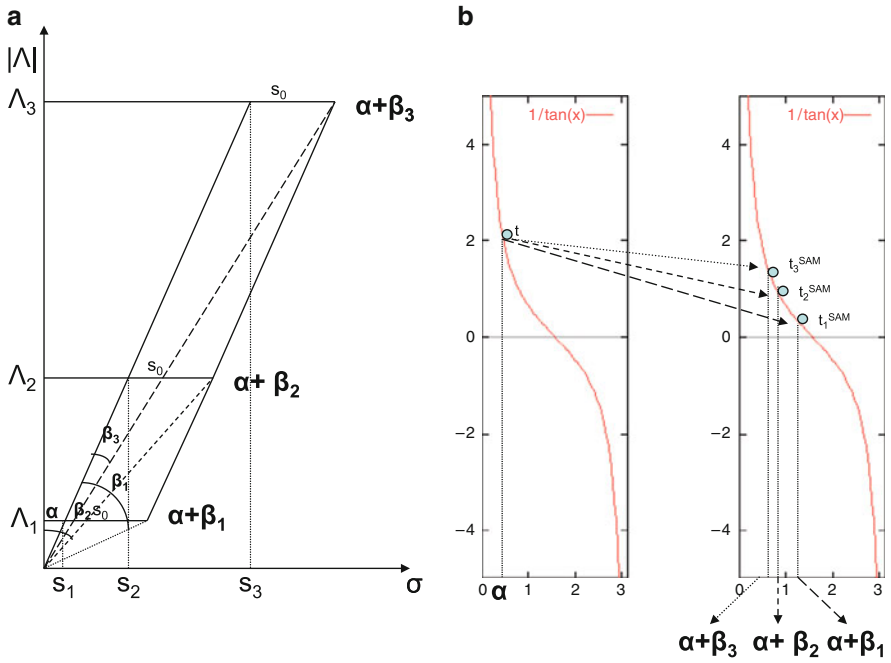

Fig. 8.2 Graphical interpretation of the SAM test statistic: (**a**) the SAM test statistics and (**b**) the cotangent function

$s_{(1)}, s_{(2)}, \ldots, s_{(m)}$. Let the fudge factor $s_0 = s^{(q)}$; it is easy to see that for gene g and for a given dose i, the SAM test statistic with the fudge factor (t_g^{SAM}) and the t-test statistic (t_g) have the following relationship:

$$\begin{cases} t_g^{\text{SAM}} < \frac{1}{2} t_g & s_{(g)} < s^{(q)}, \\ t_g^{\text{SAM}} = \frac{1}{2} t_g & s_{(g)} = s^{(q)}, \\ t_g^{\text{SAM}} > \frac{1}{2} t_g & s_{(g)} > s^{(q)}. \end{cases}$$

Hence, the SAM test statistic with the fudge factor is smaller than 1/2 of the t-test statistics for genes with their standard errors smaller than the fudge factor. Moreover, the ratio between the SAM with and without the fudge factor is $s_g/(s_g + s_0)$ since $|t|_g^{\text{SAM}} = |\Delta|/s_g \times s_g/(s_g + s_0)$. Depending on s_g, the standard error of gene g, the SAM test statistic becomes smaller by ratio of $s_g/(s_g + s_0)$.

8.3 The Choice of the Fudge Factor

As pointed out in Tusher et al. (2001) and Chu et al. (2011), the fudge factor is selected in order to minimize the coefficient of variation (CV) of median absolute deviance (MAD) of the test statistics. Several candidates for s_0 are the percentiles

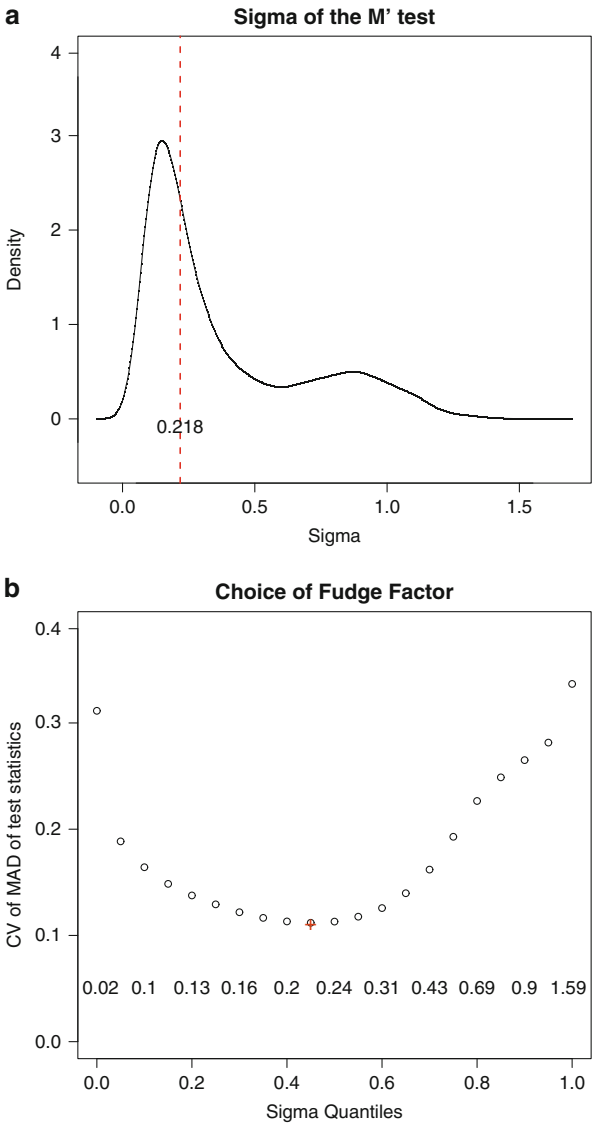

Fig. 8.3 The choice of the fudge factor. (**a**) Distribution of $\hat{s}$. s_0=0.218 (45th percentile of the standard errors in the data). (**b**) CV of MAD of the modified M test statistic vs. standard errors. The fudge factor s_0=0.218 is the value which minimizes the CV of MAD of the test statistics

of the distribution of the standard error in the sample. Figure 8.3a shows the distribution of the standard error of the M test statistic, which seems to be bimodal. Several percentiles of the distribution, $s_{(5\%)}, s_{(10\%)}, \ldots, s_{(95\%)}$, are chosen, and the CV of MAD of the modified M test statistics is calculated for each one of the chosen percentiles. As shown in Fig. 8.3b, the value of s_0 in this case study is equal to 0.218, which is the 45th percentile of the standard errors of the test statistics. Note that, as we discussed in Chap. 6, the fudge factor is chosen before the permutation, and it is kept fixed during the permutations.

8.4 The Significance Analysis of Dose-Response Microarray Data Using the IsoGene Package

In addition to the resampling-based multiple testing procedure discussed in the previous chapter, the IsoGene package has the capacity to perform a SAM analysis for dose-response data. The main function for the SAM procedure is the IsoTestSAM(). A general call of the function has the form:

```
IsoTestSAM(x, data, fudge, niter, FDR=0.05, stat)
```

The first two arguments x and data are R objects that contain information about the dose levels and the expression matrix. The argument fudge can be used in order to specify whether the analysis will be conducted without the fudge factor adjustment (fudge="none") or with automatic selection of the fudge factor (fudge="pooled"). The IsoTestSAM() combines several functions which can be used to carry out different parts of the analysis:

1. Isofudge() calculates the fudge factor in the SAM regularized test statistic.
2. IsoGenemSAM() is used to obtain the SAM test statistics.
3. Isoqqstat() calculates the SAM test statistic for the required number of permutations specified by user.
4. Isoallfdr() obtains the delta table in the SAM procedure.
5. Isoqval() computes q values of the SAM.
6. IsoSAMPlot() produces the graphical display output of the analysis.

8.4.1 Calculation of the Fudge Factor

The R object fudge.factor is a vector of the fudge factors for the five test statistics and it can be produced with the function Isofudge() in the following way:

```
> fudge.factor <- Isofudge(x, data)
> fudge.factor
[1] 0.05433744 0.22199918 0.19684197 0.25662444 0.21811043
```

Note that the fudge factor of the $\bar{E}_{01}^2$ is obtained based on the algorithm for F test statistics given by Tusher et al. (2001) and should be used with caution. The performance of using the fudge factor as compared to the t-type test statistics has not yet investigated in terms of power and control of the FDR. Therefore, it is advisable to use the fudge factor in the t-type test statistics.

8.4.2 Calculation of the Test Statistics

After the initial step in which we calculate the fudge factor, the observed test statistics are calculated with the function IsoGenemSAM().

```
> SAMtest.stat <- IsoGenemSAM(x, data, fudge.factor)
> names(SAMtest.stat)
```

The output of the function gives

```
[1] "E2" "Williams" "Marcus" "M" "ModM" "direction"
```

The object `SAMtest.stat` contains all test statistics and the direction of the trend.

```
> SAMtest.stat[[1]][1:10]
[1] 0.26447986 0.18999517 0.48012060 0.23728408 0.10796339
 0.14052809 0.42501333 0.09426217 0.11186881 0.41814965
> SAMtest.stat[[2]][1:10]
[1] -0.89438193  0.33788341  1.17016612  0.44809154 -0.37792503
0.16928575  1.40636038 -0.42466794  0.04789247 -1.26568858
> SAMtest.stat[[3]][1:10]
[1] -1.0376186  0.8067224  1.3904565  0.8226705 -0.4703604
0.5463295  1.5671898 -0.4904282 -0.4930563 -1.5825487
> SAMtest.stat[[4]][1:10]
[1] -0.8990584  0.6872440  1.2812581  0.6610949 -0.4531919
0.4158778  1.3448651 -0.4422525 -0.3838333 -1.2290796
> SAMtest.stat[[5]][1:10]
[1] -0.9924257  0.7608038  1.4288688  0.7586334 -0.4985868
0.4826826  1.4670386 -0.5074367 -0.4436651 -1.3792490
> SAMtest.stat[[6]][1:10]
2   3   4   5   6   7   8   9  10  11
"d" "u" "u" "u" "d" "u" "u" "d" "d" "d"
```

8.4.3 The Delta Table

As we mentioned in Chap. 6, for a given test statistic, a gene is declared differentially expressed whenever

$$|t_g^{obs} - t_g^{exp}| \geq \Delta,$$

where, t_g^{obs} and t_g^{exp} are gene specific observed and expected test statistics, respectively, and Δ is a threshold which is chosen in order to control the FDR at a given level. To obtain the SAM test statistics for one of five test statistic values, for example, for the modified M test, with the required number of permutations specified by users, to compute the delta table in the SAM procedure, we can use the functions `Isoqqstat()` and `Isoallfdr()` as follows:

```
> set.seed(123)
> qqstat <- Isoqqstat(x, data, fudge="pooled", niter=100)
> dtable <-  Isoallfdr(qqstat, , stat="ModifM")
> dtable
     Ddelta FalsePositive50% FalsePositive90% Called FDR50% FDR90%
  [1,]   0.01        9374.1753        9488.2337  16616 0.5642 0.5710
  [2,]   0.07        8322.9868        8605.0284  15578 0.5343 0.5524
  [3,]   0.13        7266.9118        7698.1376  14469 0.5022 0.5320
  [4,]   0.19        3416.0033        4345.8898  10270 0.3326 0.4232
.
.
 [12,]   0.67         130.5003         205.5811   2476 0.0527 0.0830
 ------------------------------------------------------------------------
```

```
 [13,]   0.73          94.5696         144.4701   2148 0.0440 0.0673
------------------------------------------------------------------------
.
.
 [36,]   2.11          0.5749           1.7247    308 0.0019 0.0056
 [37,]   2.17          0.5749           1.1498    281 0.0020 0.0041
 [38,]   2.23          0.5749           1.1498    269 0.0021 0.0043
 [39,]   2.29          0.0000           1.1498    253 0.0000 0.0045
```

Note that in the `Isoallfdr()`, the value of Δ is left unspecified, with default value taken from the data, i.e., all the percentiles of the standard errors. By fixing the 50% FDR at 0.05, the corresponding delta value is 0.73 (marked in-between the dashed lines) as we obtain from the delta table above, the number of differentially expressed genes is 2,148 with potential 144 genes as false positives.

8.4.4 The q Values

The q value of a gene is the FDR for the list of all gene declared differentially expressed including that gene and all genes that are more significant (Chu et al. 2011). By specifying the desired Δ value, delta table, and the user-defined test statistic in the function `Isoqval()`, we can obtain the q value of each gene from the SAM procedure.

```
> qval <- Isoqval(delta=0.73, allfdr=dtable, qqstat=qqstat,
stat="ModifM")
> dim(qval[[1]])
[1] 16998     3
> qval[[1]]
      Row.names    t.stat q.val
 [1,]      5131 -8.802030     0
 [2,]      3009 -8.385913     0
 [3,]      9453 -8.266634     0
 [4,]      7279 -7.390471     0
.
.
[16995,]      4625 11.152476595 0.0000
[16996,]      4624 11.606319834 0.0000
[16997,]      7760 13.106895443 0.0000
[16998,]        51 14.852821743 0.0000
> dim(qval[[2]])
[1] 2148    3
> qval[[2]]
     Row.names    t.stat q.val
 [1,]      5131 -8.802030     0
 [2,]      3009 -8.385913     0
 [3,]      9453 -8.266634     0
 [4,]      7279 -7.390471     0
.
.
[2145,]      4625 11.152477 0.0000
[2146,]      4624 11.606320 0.0000
[2147,]      7760 13.106895 0.0000
[2148,]        51 14.852822 0.0000
```

The first object of the output is the list of q values for all the genes, ranked from the smallest test statistic value to the largest, while the second object is the

list of q values for the 2,148 differentially expressed genes at the 50% FDR level of 0.05, ranked from the smallest test statistic value to the largest. The first column of the output matrices is the row name of genes in the dataset, the second column is the observed modified M test statistic value, and the last column is the q value of the SAM procedure for both objects.

8.4.5 Analysis in One Step

Instead of using the functions above to produce separate parts of the analysis, we can use the function `IsoTestSAM()` to summarize all the steps above and give results of a list of significant findings, which is the same second output of the function `Isoqval()` and the output of the function `Isoqqstat()` and `Isoallfdr()` as well.

```
IsoTestSAM(x, y=data, fudge=c(0,"pooled"), niter=100,
FDR=0.05, stat=c("E2","Williams","Marcus","M","ModifM"))
```

Specifying the same options as above in this function, we can obtain the list of significant genes. Note that the last two columns of the `IsoTestSAM` gives additional information by calculating the permutation p values based on the SAM permutation matrix and adjusting these p values using the BH-FDR procedure.

```
> set.seed(123)
> IsoSAM.obj <- IsoTestSAM (x, y=data, fudge="pooled",
niter=100, FDR=0.05, stat="ModifM")
> IsoSAM.obj[[1]]
    Probe.ID row.number  stat.val qvalue       pvalue  adj.pvalue
1       g5131       5131 -8.802030 0.0000 0.000000e+00 0.0000000000
2       g3009       3009 -8.385913 0.0000 1.176609e-06 0.0007407407
3       g9453       9453 -8.266634 0.0000 1.176609e-06 0.0007407407
4       g7279       7279 -7.390471 0.0000 1.764914e-06 0.0009677419
.
.
2145    g4625       4625 11.152477 0.0000 0.000000e+00 0.0000000000
2146    g4624       4624 11.606320 0.0000 0.000000e+00 0.0000000000
2147    g7760       7760 13.106895 0.0000 0.000000e+00 0.0000000000
2148      g51         51 14.852822 0.0000 0.000000e+00 0.0000000000
```

8.4.6 Graphical Displays

Finally, the graphic output of the SAM procedure can be produced using the function `IsoSAMPlot()`.

```
> IsoSAMPlot(qqstat, allfdr, FDR=0.05, stat=c(E2","Williams",
"Marcus","M","ModifM"))
```

This function requires the use of output from `Isoqqstat()` and `Isoallfdr()`, given a user-defined test statistic, and the FDR level to control. We still take the modified M test statistic, for example, at the FDR of 0.05. There are four

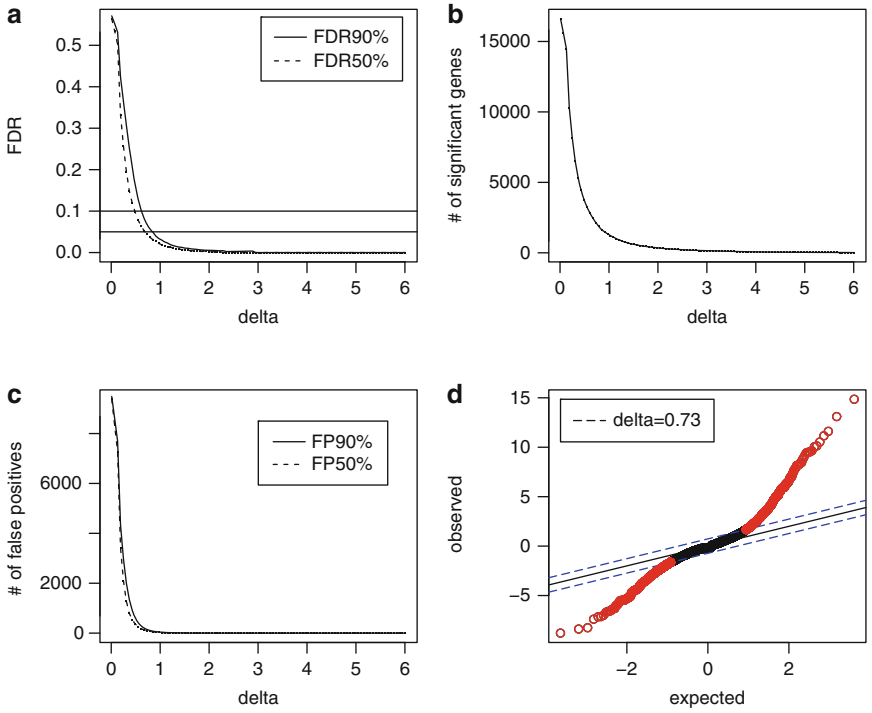

Fig. 8.4 The SAM plots: (**a**) plot of the FDR vs. delta; (**b**) plot of number of significant genes vs. delta; (**c**) plot of number of false positives vs. delta; (**d**) plot of observed vs. expected test statistics

plots generated by the SAM procedure. Panel a shows the FDR [either 50% or 90% (more stringent)] vs. Δ, from which, user can choose the delta value with the corresponding desired FDR. Panel b shows the number of significant genes vs. Δ, and panel c shows the number of false positives (either 50% or 90%) vs. Δ. Finally, panel d shows the observed vs. the expected (obtained from permutations) test statistics, in which dots beyond the two dashed lines are those genes called differentially expressed (Fig. 8.4).

```
> IsoSAMPlot(qqstat=qqstat, allfdr=dtable, FDR=0.05, stat="ModifM")
```

8.5 Discussion

In this chapter, we discussed the issue of genes with small variance and the solution that the SAM procedure provides for this problem. The SAM procedure was adapted in the dose-response microarray setting in case of small variance, by modifying the likelihood ratio test and the four t-type test statistics with a fudge factor. This

adjustment of fudge factor for the five test statistics is new; therefore, we suggest to conduct a simulation study to further investigate the performance of these adjusted test statistics.

References

Chu, G., Li, J., Narasimhan, B., Tibshirani, R., & Tusher, V. (2011). *SAM: "Significance Analysis of Microarrays" users guide and technical document*. http://www-stat.stanford.edu/~tibs/SAM/sam.pdf.

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrys applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences, 98*, 5116–5121.

Chapter 9
δ-Clustering of Monotone Profiles

Adetayo Kasim, Suzy Van Sanden, Martin Otava, Sepp Hochreiter, Djork-Arné Clevert, Willem Talloen, and Dan Lin

9.1 Introduction

In Chaps. 7 and 8, we discussed several testing procedures to detect differentially expressed genes with monotone relationship with respect to dose. The second question of primary interest in dose-response studies is the nature (or the shape of curve) of the dose-response relationship. In the context of dose-response microarray experiments, we wish to group (or classify) genes with similar dose-response relationship. Similar to the previous chapters, the subset of genes with monotone relationship is of interest.

In the following three chapters, we discuss four methods to cluster possible dose-response profiles using order-restricted clustering, inference-based methods, and model selection procedures based on information theory. Assuming a monotone

A. Kasim (✉)
Wolfson Research Institute, Durham University, Durham, UK
e-mail: a.s.kasim@durham.ac.uk

S. Van Sanden · W. Talloen
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: SVSANDE1@its.jnj.com; WTALLOEN@its.jnj.com

M. Otava
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Martin.otava@uhasselt.be

D. Lin
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

S. Hochreiter · D.-A. Clevert
Institute of Bioinformatics, Johannes Kepler University, Linz, Austria
e-mail: hochreit@bioinf.jku.at; okko@clevert.de

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_9,

Table 9.1 The set of seven possible monotonic dose-response models for an experiment with four dose levels

Model	Up: mean structure	Down: mean structure
g_1	$\mu_0 = \mu_1 = \mu_2 < \mu_3$	$\mu_0 = \mu_1 = \mu_2 > \mu_3$
g_2	$\mu_0 = \mu_1 < \mu_2 = \mu_3$	$\mu_0 = \mu_1 > \mu_2 = \mu_3$
g_3	$\mu_0 < \mu_1 = \mu_2 = \mu_3$	$\mu_0 > \mu_1 = \mu_2 = \mu_3$
g_4	$\mu_0 < \mu_1 = \mu_2 < \mu_3$	$\mu_0 > \mu_1 = \mu_2 > \mu_3$
g_5	$\mu_0 = \mu_1 < \mu_2 < \mu_3$	$\mu_0 = \mu_1 > \mu_2 > \mu_3$
g_6	$\mu_0 < \mu_1 < \mu_2 = \mu_3$	$\mu_0 > \mu_1 > \mu_2 = \mu_3$
g_7	$\mu_0 < \mu_1 < \mu_2 < \mu_3$	$\mu_0 > \mu_1 > \mu_2 > \mu_3$

μ_i is the mean response of dose level

relationship, the dose-response curve could be either linear, nonlinear, concave, or convex. Furthermore, for an experiment with $K + 1$ dose levels, there is a fixed number of monotonic models that can be fitted. For instance, in a dose-response experiment with four dose levels, upon the establishment of a monotonic relationship between gene expression and doses, there is a set of seven models, shown in Table 9.1 and Fig. 9.1, that can be fitted to the data. As mentioned above, in the following two chapters, we present two different methods for clustering subsets of genes into homogenous groups in terms of dose-response curve shape. In Chap. 10, we focus on classification based on information criteria, while in this chapter, we focus on clustering of dose-response microarray experiments in order to find clusters of genes with coherent expression profiles. In both Chaps. 9 and 10, we focus on monotone profiles. In Chap. 11, we discuss two methods for testing and clustering for order-restricted, but not necessary monotone, gene profiles.

9.2 Order-Restricted Curve Clustering

Traditional clustering methods such as K-means (Hartigan and Wong 1979), hierarchical methods (Johnson and Wichern 2008), and self-organizing maps (Kohonen 2001) have been used extensively in microarray experiments to identify clusters of co-regulated genes. While the clustering methods find co-regulated genes based on similarity across all the conditions in an experiment, it has been argued that co-expression under subsets of conditions is more biologically intuitive and relevant. As such, several methods for biclustering have been developed in the recent years to find co-regulated genes under a subset of conditions (Madeira and Oliviera 2004). The major difference between clustering and biclustering is that clustering focuses on global patterns in gene expression data, while biclustering focuses on local structures that are inherent in gene expression data. However, both clustering and biclustering methods assume that genes are profiled under multiple nominal experimental conditions. More recently, special attention has been paid to microarray experiments profiled under experimental conditions with trends, such as time course, dose-response, and temperature (Madeira and Oliviera 2004). The aim of the analysis discussed in this chapter is to cluster genes with similar

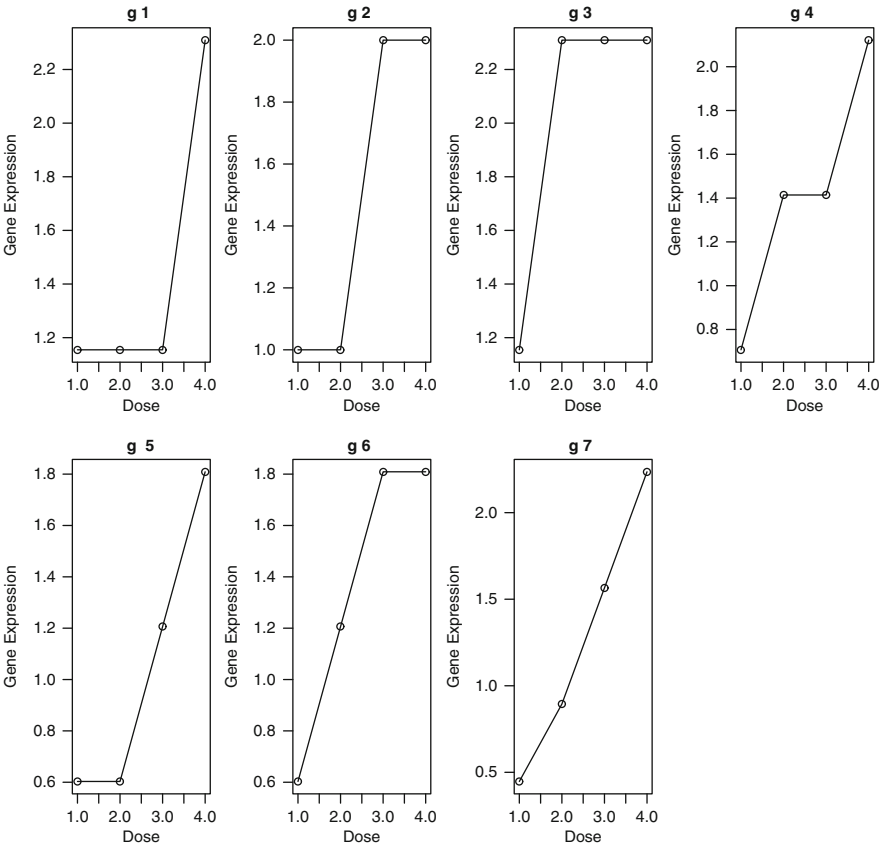

Fig. 9.1 Illustrative example. The set of all possible upward monotone increasing dose-response curves for an experiment with four dose levels

dose-response curve shapes under monotone constraints. Hence, for the analysis presented in this chapter, the dimension of the dose is fixed, and our goal is to find subsets of genes with similarly shaped dose-response curves.

9.2.1 The δ-Biclustering Method

δ-Biclustering is a node deletion-based algorithm introduced by Cheng and Church (2000) to find a subset of genes and conditions with a high similarity score. The similarity between members of a bicluster is defined in terms of the mean squared residue score. The lower the mean squared residue score, the more homogeneous is the cluster. The δ-biclustering method relies on the assumption that every entry in a gene expression matrix can be expressed in terms of its row mean, column mean, the overall mean of the expression matrix, and random error. In other words, for a

gene expression matrix $\mathbf{Y}$ with entries y_{mi}, the residue of expression value of the mth gene under condition (dose) i can be expressed as :

$$r_{mi} = y_{mi} - y_{Mi} - y_{mI} + y_{MI}, \tag{9.1}$$

and the mean squared residue score of matrix $\mathbf{Y}$ is defined as:

$$H_{MI} = \frac{1}{|M||I|} \sum_{m \in M i \in I} r_{mi}^2, \tag{9.2}$$

where y_{MI} is the overall mean of the expression matrix $\mathbf{Y}$, y_{mI} is the mean expression of gene m, and y_{Mi} is the mean expression of condition i. The numbers of genes and conditions are denoted with $|M|$ and $|I|$, respectively. Note that the model for the residual in (9.1) can be expressed in the form of a two-way ANOVA model without an interaction term:

$$y_{mi} = \mu + \alpha_m + \beta_i + r_{mi}, \tag{9.3}$$

with $\mu = y_{MI}$, $\alpha_m = y_{mI} - y_{MI}$ and $\beta_i = y_{Mi} - y_{MI}$.

As an illustration, we present an example of two expression matrices. A is an example of a perfect cluster with coherent values, and B is an example of a cluster for which the genes have coherent values except for the genes in the last two rows of the matrix. Based on (9.1), the mean squared residue score for A is zero since the total variability of the cluster can be explained by the row means, column means, and overall mean of the matrix. However, for B, the mean squared residue score is 8.11. This means that genes in A are more similar than those in B. Suppose that the last two rows of B are excluded; then the mean squared score becomes zero.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 30 & 31 & 32 & 33 & 34 \\ 32 & 33 & 34 & 35 & 36 \\ 81 & 82 & 83 & 84 & 85 \\ 91 & 92 & 93 & 94 & 95 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 30 & 31 & 32 & 33 & 34 \\ 32 & 33 & 34 & 35 & 36 \\ 42 & 43 & 30 & 30 & 31 \\ 37 & 30 & 36 & 35 & 34 \end{pmatrix} \tag{9.4}$$

In microarray experiments, a perfect cluster/bicluster such as A is unlikely given the noise level of the technology. It may therefore be sufficient to find clusters/biclusters of genes whose mean squared residue scores are less than a pre-specified threshold δ. Cheng and Church (2000) proposed the δ-biclustering method for gene expression data based on a suit of node deletion algorithms that evolve in cycles starting from the input gene expression matrix until a bicluster that satisfies the δ-criterion is found. Several cycles of the algorithm are then applied to the data by replacing the initially found biclusters with random data.

9.2.2 *Clustering of Order-Restricted Dose-Response Profiles*

In some settings, the column effects β in (9.3) have inherent ordering, which may be time, temperature, or, in our example, increasing doses of a therapeutic compound. The aim is therefore to find clusters of genes that are similar in intensities and trends. δ-Clustering, a variant of δ-biclustering by Cheng and Church (2000), may in general be used.

9.2.2.1 The δ-Clustering Method

Applying the δ-biclustering algorithm in only one dimension offers a clustering method for which the number of clusters is not required to be specified but implicitly controlled by the degree of homogeneity assumed for a cluster. However, the choice of a δ value to achieve a desired degree of homogeneity is not readily available (Prelic et al. 2006). We propose a relative δ criterion, where a cluster is a subset of genes with mean squared residue score smaller than a certain proportion λ ($0 \leq \lambda \leq 1$) of the heterogeneity in the observed data. Additionally, our aim is to find non-overlapping clusters of genes. To achieve this, we propose that members of initially found clusters be deleted from the observed data before another cycle of the node deletion algorithm is applied. To overcome the problem of local minima (Prelic et al. 2006), we introduce an additional parameter ϕ that indicates the minimum number of genes in a cluster. Note that for $\lambda = 0$, the algorithm searches for clusters of genes with mean squared residue score of 0, which may result in as many clusters as the number of genes in the data set. On the other hand, specifying λ to be 1 means to consider all the genes as one cluster. Any value of λ between 0 and 1 reflects the degree of homogeneity expected of a cluster. We define the algorithm to carry out this task as Algorithm 1, which can also be applied to cluster time-course microarray experiments.

Algorithm 1: δ-clustering

Input: Y, a matrix of real numbers; ϕ, minimum number of genes in a cluster; and λ: $0 \leq \lambda \leq 1$.

Output: Y_{sub}, a sub-matrix with number of rows or columns less than or equal to the number of the rows or columns of the original matrix Y.

Initialization: $\delta = \lambda * H_P$, where H_P is the mean squared residue score of the observed data.

Iteration:

1. Apply node deletion algorithms proposed by Cheng and Church (2000) only to the genes while the dose levels are kept fixed.
2. If the mean squared residue of the reduced matrix satisfies the δ criterion or the number of genes in the reduced matrix is at least ϕ, then output the reduced matrix as a cluster.
3. Delete members of the cluster found in step 2.
4. Repeat steps 1–3 on the non-clustered genes until every gene belongs to a cluster.

The algorithm discussed above can be applied to any setting of an ordered design variable (time, temperature, dose, etc.), but it does not require a monotone gene expression profile. In other words, using Algorithm 1, we will be able to cluster subsets of genes with a similar dose-response curve shape, but not necessarily monotone. In the following section, we discuss an algorithm in which isotonic regression is used in the dose dimension in order to cluster genes with similar monotone dose-response curve shapes. Note that the δ-clustering algorithm is applied to an expression matrix after an initial filtering in which genes with non significant dose-response relationship are excluded from the analysis.

9.2.2.2 δ-Clustering of Dose-Response Monotone Profiles

A typical dose-response microarray data Y has entries y_{mij} corresponding to the expression level of gene m under dose i from subject/sample j. Usually, different subjects/samples are used for different doses. In order to find clusters of genes with a similar monotone dose-response relationship, it is required that gene expression measurements under increasing doses are constrained to be monotone using isotonic regression. Thus, a new matrix Y^* of the isotonic means is obtained. The gene effects (α_m), isotonic dose effects (β_i^*), and overall mean (μ) can be defined as shown below:

$$\begin{aligned}
\mu &= \textstyle\sum_{m\in M, i\in I} \frac{y^*_{mi}}{|M||I|} && \text{overall mean,} \\
\alpha_m &= \textstyle\sum_{i\in I} \frac{y^*_{mi}}{|I|} - \mu && \text{effect of the } m\text{th row (gene),} \\
\beta_i^* &= \textstyle\sum_{m\in M} \frac{y^*_{mi}}{|M|} - \mu && \text{effect of the } i\text{th column (dose).}
\end{aligned} \tag{9.5}$$

The clustering algorithm is applied specifically to each direction in order to find clusters of genes with monotone increasing or decreasing trends. The linear model for the δ-clustering algorithm using a reduced gene expression matrix based only on the isotonic means is given by the model in (9.6) and is described in Algorithm 2:

$$y^*_{mi} = \mu + \alpha_m + \beta_i^* + r^*_{mi}. \tag{9.6}$$

Algorithm 2 : Order-restricted clustering based only on the isotonic means

Input: Y^*, a matrix of isotonic means; ϕ, minimum number of genes in a cluster; and λ: $0 \leq \lambda \leq 1$.

Output: Y_{sub}, a sub-matrix with fewer rows or columns than the rows or columns of the original matrix Y^*.

Initialization: $\delta = \lambda * H_P$, where H_P is the mean squared residue score of Y^*

Iteration:

1. Using the likelihood ratio statistic, assign a direction to each gene.
2. Apply Algorithm 1 using the linear model in Eq. (9.6), specifically to each direction.

9.2.3 Application to the Data

9.2.3.1 Step I: Initial Filtering

The initial filtering for dose-response microarray data is performed by applying a likelihood ratio test, discussed in Chap. 7, to establish a dose-response relationship under order-restricted constraints. Raw p values for the genes are calculated based on permutations under the null hypothesis. In order to adjust for multiplicity, the Benjamini and Hochberg (1995) FDR procedure is applied to the raw p values, and genes for which the adjusted p values are smaller than 0.05 are declared differentially expressed (Ge et al. 2003; Lin et al. 2007). In total, the null hypothesis was rejected for 3,499 out of the 16,998 genes that were tested. We present in Fig. 9.2 examples of two genes with a significant monotonic dose-response relationship. We notice that even though both genes are significant, the range of gene expression in Fig. 9.2a is much larger than that of Fig. 9.2b.

For the analysis presented in this chapter, we used the `ORCME` package. In the first step, we need to create two R objects containing the isotonic means for both upward and downward directions. The function `monotoneDirection()` can be used in order to identify the direction of the trend.

```
## loading ORCME package after installation in R
> library(ORCME)
## loading the data
> data(geneData)
> data(doseData)
## Direction specific isotonic means
> dirData <- monotoneDirection(geneData = geneData,
            doseData = doseData)
```

In the next step, after the determination of the trend direction, we need to create the R objects for genes with upward and downward trends. The function `plotIsomeans()` can be used to produce gene-specific profile plot:

```
## direction of monotone trend
> Direction <- dirData$direction
## Isotonic means  for upward genes
> incData <- as.data.frame(dirData$incData)
## Isotonic means for downward genes
> decData <- as.data.frame(dirData$decData)
## observed data upward genes
> obsIncData <- as.data.frame(dirData$obsincData)
## observed data for downward genes
> obsDecData <- as.data.frame(dirData$obsdecData)
## isotonic means for all genes
> isoMeans <- as.data.frame(dirData$arrayMean)

## gene-specific profile plot

> plotIsomeans(monoData=incData,obsData=obsIncData,doseData=
            doseData,geneIndex=10)
```

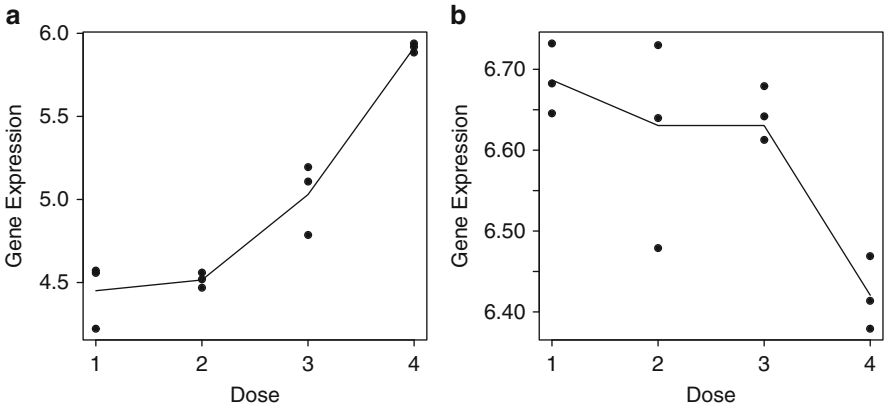

Fig. 9.2 Examples of two significant genes. (**a**) Upward trend. (**b**) Downward trend

9.2.3.2 Step II: Clustering of Order-Restricted Dose-Response Profiles

Based on the penalized within-cluster sum of squares (which will be discussed in the next section), $\lambda = 0.15$ is chosen as the optimum choice of λ for clustering the upward monotone genes. The main function for clustering is `ORCME()`. In our example, for genes with upward trends, we use

```
> ORCMEoutput <- ORCME(DRdata=incData,lambda=0.15,phi=2)
```

The δ-clustering method with $\lambda = 0.15$ results in 26 clusters for the 1,600 upward monotone genes. The first cluster contains 1,278 genes, and the last cluster contains 2 genes. The large size of the first cluster is an inherent feature of the δ-clustering method. The genes in the first cluster are usually the least expressed genes. The first clusters from the δ-clustering method often contain genes that are less expressed than those in the later clusters. Figure 9.3 presents examples of clusters with upward monotone profiles. The upper panel shows the raw gene expression values, and the lower panel shows gene expression values centered around gene-specific means. Genes within a cluster show coherence in terms of similarities between their expression values and trends. The function `plotCluster()` produces the isotonic mean profiles for a specific cluster. The option `zeroMean=TRUE` centered the gene profiles around the gene-specific means, as shown in the lower panels in Fig. 9.3.

```
> plotCluster(DRdata=incData,doseData=doseData, ORCMEoutput=ORCMEoutput,
    clusterID=4,zeroMean=FALSE, xlabel="Dose",ylabel="Gene Expression")
```

The penalized within-cluster sum of squares score suggests $\lambda = 0.3$ as the optimum choice of λ for the downward monotone genes. The application of the δ-clustering method results in 19 clusters for the 1,899 downward monotone genes. The first cluster contains 1,700 genes, and the last cluster contains 2 genes.

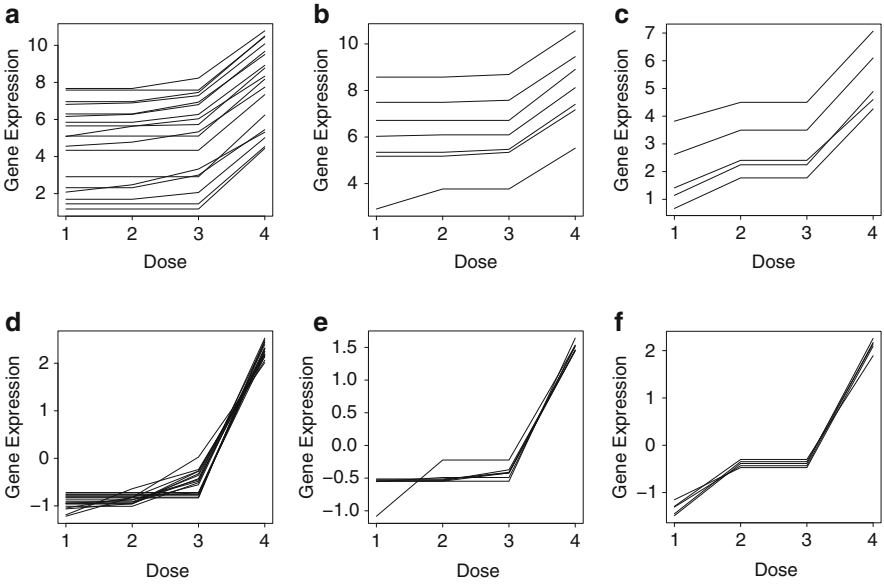

Fig. 9.3 Examples of clusters from upward monotone genes. Panels (**a**) and (**b**): average gene expression profiles. (**a**) Cluster ID = 4; (**b**) cluster ID = 6; (**c**) cluster ID = 13; (**d**) cluster ID = 4 (*centered*); (**e**) cluster ID = 6 (*centered*); (**f**) cluster ID = 13 (*centered*).

Figure 9.4 presents examples of clusters with downward monotone profiles. Similar to the clustering of the upward monotone genes, the clusters contain genes with coherent values. However, there are situations that few members of a cluster show different dose-response trends, but in most cases, the deviation occurs in only one of the four doses in the experiments.

9.3 Choice of the Clustering Parameter λ

A major challenge in cluster analysis is the estimation of the optimum number of clusters, which in most cases determines the quality of the resulting clusters. While the number of clusters is not required for the δ-clustering method, the optimum choice for λ and ϕ is unknown and may be data dependent. We suggest that ϕ will be kept fixed based on the preference of the researcher. The possible choice for λ can be explored based on the within-cluster sum of squares, which can be computed for λ in the range of zero to one. Let us assume that for a specific value of λ the δ-clustering method results in $n(\lambda)$ clusters. Let $R(\lambda)$ denote the within-cluster sum of squares for this value of λ; then

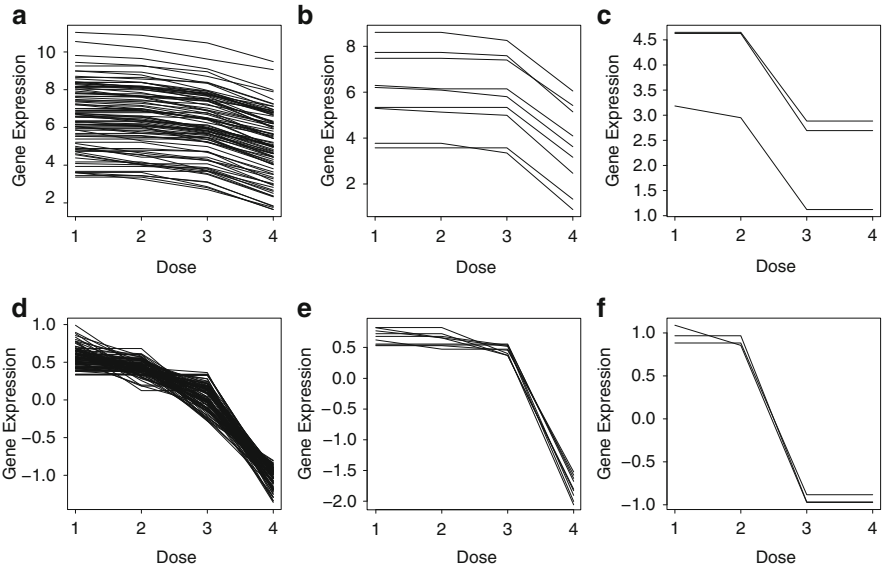

Fig. 9.4 Examples of clusters with downward monotone profiles. (**a**) Cluster ID = 3; (**b**) cluster ID = 6; (**c**) cluster ID = 13; (**d**) Cluster ID = 3 (*centered*); (**e**) cluster ID = 6 (*centered*); (**f**) cluster ID = 13 (*centered*)

$$R(\lambda) = \sum_{q} \sum_{mi} (y_{mi} - \mu_q - \alpha_m - \beta^{\star}_{iq})^2, \quad q = 1, \ldots, n(\lambda). \tag{9.7}$$

Let N be the number of genes to be clustered. The range for $n(\lambda)$ lies between one and the number of genes, i.e., $1 \leq n(\lambda) \leq N$. When $\lambda = 1$, $n(\lambda) = 1$ and $n(\lambda) \leq N$ for $\lambda = 0$. Since $R(\lambda)$ is a decreasing function of $n(\lambda)$ and an increasing function of λ, $R(\lambda)$ will be minimum when $n(\lambda) = N$ and maximum when $n(\lambda) = 1$. Note that when $n(\lambda) = 1$, the within-cluster sum of squares equals the total sum of squares for the gene expression matrix. Our aim is to find the value of λ, taking into account the trade-off between the within-cluster sum of squares and the number of resulting clusters. This criterion is referred to as *penalized within-cluster sum of squares* (pWSS), and it is defined as

$$\text{pWSS}(\lambda) = R(\lambda) + 2n(\lambda). \tag{9.8}$$

Following Tibshirani et al. (2001), other criteria for traditional clustering methods can be considered as well. We can modify the Calinski and Harabasz (1974) index as

$$CH(\lambda) = \frac{B(\lambda)/n(\lambda)}{W(\lambda)/(N - n(\lambda))}, \tag{9.9}$$

where $B(\lambda)$ and $W(\lambda)$ are the between-cluster sum of squares and within-cluster sum of squares, respectively. While the within-cluster sum of squares is expected

to increase with increasing λ, the between-cluster sum of squares is expected to decrease with increasing λ. Another criterion considered is the Hartigan (1975) index, which is also modified as

$$H(\lambda) = \left[\frac{w(\lambda_\ell)}{w(\lambda_{\ell+1})} - 1\right] / \left[N - n\left(\lambda_{\ell+1}\right)\right], \tag{9.10}$$

where ℓ is an index for the unique value of λ. The original definition for the H index is based on the sequential increase in number of clusters. For our proposal, this is not the case, as more than one value of λ may result in the same number of clusters. However, the criterion can still be used to investigate the gain in within-cluster sum of squares when moving from a lower value of λ to an adjacent higher value.

9.3.1 Application to the Data: The Choice of the Clustering Parameter

9.3.1.1 The Trade-Off Between λ and the Number of Clusters

The relative proportion (λ) of the mean squared residue score of the monotonized gene expression matrix is proposed as a clustering parameter for the δ-clustering method. Though λ is bounded between 0 and 1, the choice of the optimum value of λ is unknown. Similar to the resampling approach for random forest (Breiman 1996) and ABC learning (Amaratunga et al. 2008), we propose to generate 100 resampled datasets, with each dataset containing 100 genes randomly sampled with replacement from the reduced expression data. For each of the resampled datasets, the δ-clustering method is applied based on a set of values of λ ranging from 0.05 to 0.95. Note that the minimum number of genes in a cluster is fixed at two. The resampling is done using the function `resampleORCME()`.

```
## exploring optimum choice of lambda (the sequence for lambda)
> lambdaVector <- seq(0.05,0.95,0.05)
## upward trends
> lambdaChoiceOutput <- resampleORCME(clusteringData=incData,
                                      lambdaVector=lambdaVector)
```

Figure 9.5 shows the relationship between the within-cluster sum of square, the number of resulting clusters, and λ. Panels (a) and (c) show the relationship between the within-cluster sum of square and λ for the upward and downward monotone genes, respectively. Panels (c) and (d) show the relationship between the number of resulting clusters and λ for the upward and downward monotone genes, respectively. The within-cluster sum of squares increases with an increase in λ, while the number of clusters decreases with an increase in λ. It shows that a trade-off between the within-cluster sum of squares and number of clusters may be a criterion for an optimum choice of λ. Diagnostic plots were produced using the function `plotLambda()`.

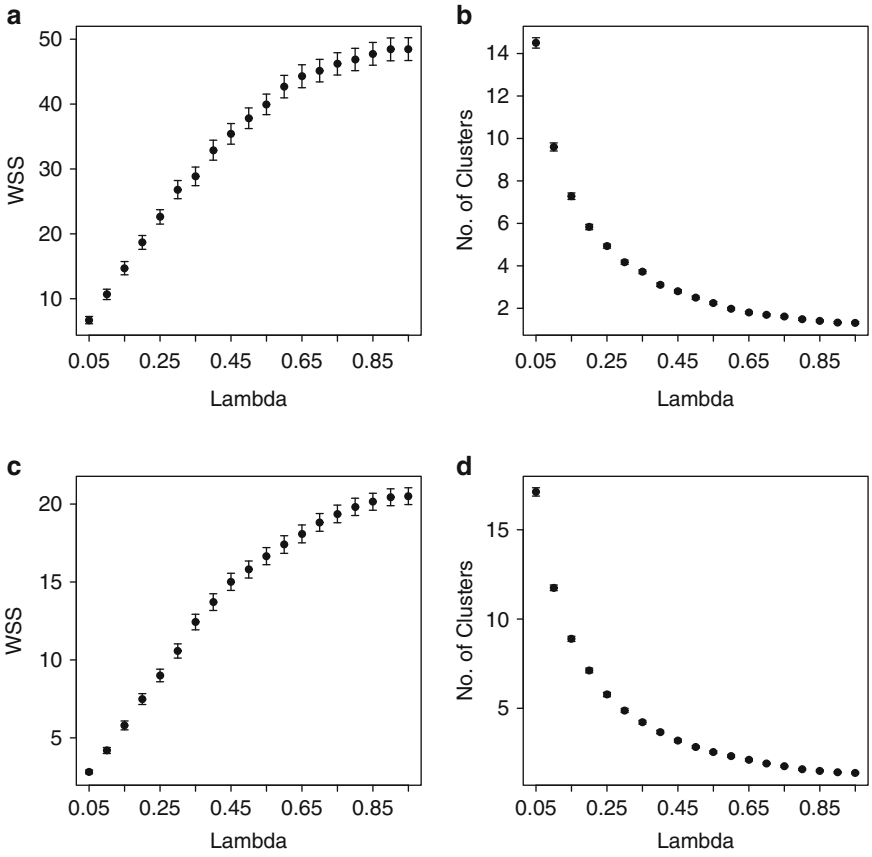

Fig. 9.5 Within-cluster sum of squares and the number of clusters as a function of λ. (**a**) WSS (upward trends); (**b**) number of clusters (upward trends); (**c**) WSS (downwards trends); (**d**) number of clusters (downward trends)

```
> plotLambda(lambdaChoiceOutput,output="wss")
> plotLambda(lambdaChoiceOutput,output="ncluster")
```

9.3.1.2 The Choice of λ

The trade-off between the within-cluster sum of squares and number of clusters based on a pWSS is presented in Fig. 9.6a, d for upward and downward monotone genes, respectively. The pWSS for the upward monotone genes reaches a minimum at $\lambda = 0.15$ and at $\lambda = 0.3$ for the downward monotone genes. Figure 9.6b, e shows the relationship between the *CH* values and λ for upward and downward monotone

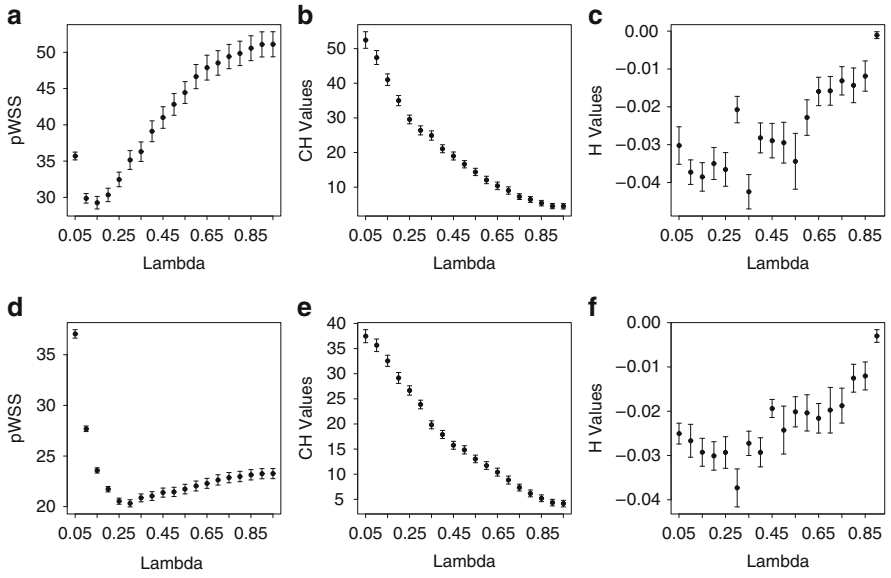

Fig. 9.6 The choice of λ. *Upper panels*: upward trends. *Lower panels*: downward trends. (**a** and **d**) pWSS; (**b** and **e**) CV index; (**c** and **f**) H index

genes, respectively. The maximum value of *CH* is reached at $\lambda = 0.05$ for both the upward and downward monotone genes. It appears for our case study that the variant of the Calsanzik and Harabasx (1974) index is not an informative criterion. It favors the λ value which results in the highest number of clusters. Figure 9.6c, f presents the relationship between the *H* value and λ for the upward and downward monotone genes, respectively. The *H* values do not show a smooth pattern as observed from the *pWSS*. However, it reaches its minimum value at $\lambda = 0.35$ for the upward monotone genes and at $\lambda = 0.3$ for the downward monotone genes. Graphical output can be produced using the function `plotLambda()`. The option `output=` determines which index will be plotted.

```
> plotLambda(lambdaChoiceOutput,output="pwss")
> plotLambda(lambdaChoiceOutput,output="ch")
> plotLambda(lambdaChoiceOutput,output="h")
```

9.4 Discussion

One of the interests in dose-response microarray experiments is to find clusters of genes with similar dose-response relationships under an increasing dose of a therapeutic compound. In this chapter, a δ-clustering method is proposed for dose-response microarray data. The method is motivated by the δ-biclustering

method proposed by Cheng and Church (2000), where they defined a bicluster as a subset of genes and a subset of conditions with a "high similarity score" using the mean squared residue score. For the δ-clustering method, the δ value is modified to be data dependent. It is expressed as a relative proportion (λ) of mean squared error from the direction-dependent monotonized expression matrix. The method shares some features of standard clustering methods in that it partitions genes in a dose-response microarray data into nonoverlapping groups but also benefits from the local structures of the biclustering methods.

The δ-clustering procedure discussed in this chapter was applied to a reduced expression matrix obtained after an initial (inference-based) filtering. After the initial filtering step, the within-gene variability is ignored by the δ-clustering method and the cluster is constructed in order to reduce the between-gene variability (= the within-cluster variability). The optimum choice of λ for the clustering method is explored with a penalized within-cluster sum of squares, which is a trade-off between goodness of fit and complexity of the resulting clusters for different values of λ ranging from zero to one. The goodness-of-fit is captured with the within-cluster sum of squares, and the complexity is captured with the number of clusters. Note that the within-cluster sum of squares increases with an increase in λ and the number of clusters decreases with an increase in λ. Based on the values of λ that correspond to the minimum values of the penalized within-cluster sum of squares for the upward and downward monotone genes, 26 and 19 clusters were obtained for the upward and downward monotone genes, respectively. The first clusters contained the least expressed genes and the last clusters contained the most expressed genes in terms of the raw gene expression values.

References

Amaratunga, D., Cabrera, J., & Kovtun, V. (2008). Microarray learning with ABC. *Biostatistics, 9*, 128–136.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Soceity B, 57*, 289–300.

Breiman, L. (1996) Random forests. *Machine Learning, 24*, 123–140.

Calinski, R. B., & Harabasz, J. A. (1974). Dendrite method for cluster analysis. *Communications in Statistics, 3*, 1–27.

Cheng, Y., & Church, G. M. (2000). Biclustering of expression data. *Proceedings of the Conference on Intelligent Systems for Molecular Biology, 55*, 93–104.

Ge, Y., Dudoit, S., & Speed, P. T. (2003). *Resampling based multiple testing for microarray data analysis* (Technical report, 633). Berkeley: University of Berkeley.

Hartigan, J. A., & Wong, M. A, (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 28*(1), 100–108.

Johnson, R. A., and Wichern, D. W.(2008). *Applied Multivariate statistical analysis*. Pearson.

Kohonen, T. (2001), *Self-Organizing Maps*. 3rd edition, Springer-verlag, Berlin.

Lin, D., Shkedy, Z., Yekutieli, D., Burzykowki, T., Göhlmann, H. W. H., De Bondt, A., et al. (2007). Testing for trend in dose-response microarray experiments: Comparison of several testing procedures, multiplicity, and resampling-based inference. *Statistical Application in Genetics and Molecular Biology, 6*(1). Article 26.

Madeira, S. C., & Oliviera, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 1*(1), 24–45.

Prelic, A., Bleuler, S., Zimmermann, P., Wille, A., Buhlmann, P., Gruissem, W., et al. (2006). Systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics, 22*(9), 1122–1129.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society B, 63*, 411–423.

Chapter 10
Classification of Monotone Gene Profiles Using Information Theory Selection Methods

Dan Lin, Ziv Shkedy, and Marc Aerts

10.1 Introduction

In the previous chapter, we discussed the order-restricted δ-clustering method for clustering subsets of genes with similar monotone dose-response profiles. In this chapter, we focus on a second approach in which clustering is based on information criteria (Lin et al. 2009; Liu et al. 2009a,b). As we mentioned in Chap. 9, for a dose-response experiment with $K+1$ dose levels, after an initial filtering, there are a finite number of monotone dose-response ANOVA models which can be fitted to the data. For example, for a four dose-level experiment, there are, for each direction, seven dose-response profiles (listed in Table 9.1 and shown in Fig. 9.1 can be fitted. The aim is to choose the best model among these possible models with monotonic trends. To this purpose, we propose to use the information model selection theory discussed by Burnham and Anderson (2002) and Claeskens and Hjort (2008). For a given set of candidate models, the order-restricted information criterion (ORIC), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) are used to calculate the posterior probability of each model in the set. The model with the highest posterior probability (or equivalently, the lowest value of the information criterion) is selected. This allows us to identify the shape of the dose-response curve and to classify genes into subsets with similar dose-response curve shapes.

Figure 10.1 shows an example of a gene for which a significant monotone trend was detected in Chaps. 7 and 8. The aim of the analysis presented in this chapter is to select the model with the best trade-off between goodness-to-fit and model

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

Z. Shkedy · M. Aerts
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be; marc.aerts@uhasselt.be

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_10,

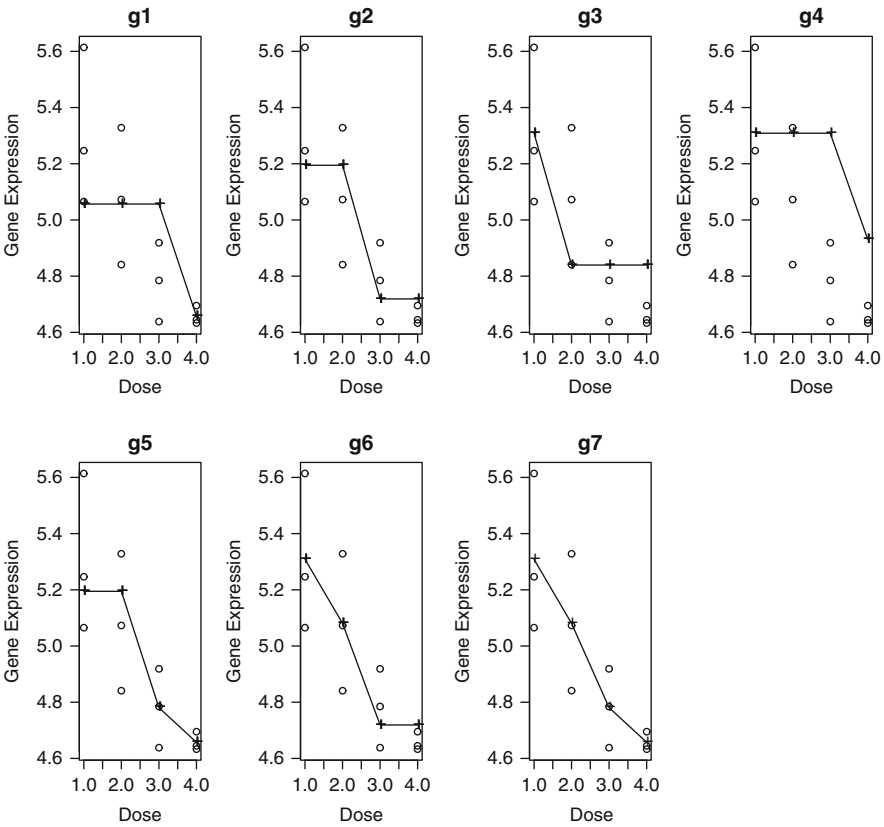

Fig. 10.1 Seven monotone models fitted for gene 3467. The model with the highest likelihood is g_7 while g_6 is the smallest AIC (see discussion in Sect. 10.2.2)

Table 10.1 Comparison of penalty values by the AIC, BIC, and ORIC for models g_1 to g_7. N is the total number of arrays

Models	Parameters	AIC	BIC	ORIC
g_1,g_2,g_3	2	2×2	$2 \times \log(N)$	2×1.16666
g_4,g_5,g_6	3	2×3	$3 \times \log(N)$	2×1.91666
g_7	4	2×4	$4 \times \log(N)$	2×2.08334

complexity and to classify (or cluster) genes with similar dose-response curve shapes. This is in contrast with Chap. 9 in which the δ-clustering procedure was based on the minimization of the residual sum of squares within a cluster.

We use the human epidermal dose-response data (three higher doses and one control dose under the control compound) as a case study to illustrate the methodology discussed in this chapter. After the initial testing procedure (using $\bar{E}^2_{01}$ test statistic), discussed in Sect. 7.6, 3,613 genes were rejected for the null hypothesis of homogeneity of means in gene expression (see Table 7.1). However,

due to a different seed used for the permutations, we now obtained 3,499 genes, which is slightly different from the number of genes reported in Chap. 7.

The content of this chapter is organized as follows. In Sect. 10.2, we address the problem of trend classification within the framework of model selection. Since the set of candidate models is estimated under order restrictions, we discuss, in Sect. 10.3, the ORIC. The one- and two-stage ORICC algorithms are discussed in Sect. 10.4. We apply the proposed method to the case study in Sect. 10.5.

10.2 Classification of Trends Using Information Criteria

10.2.1 Classification of Trends Based on Posterior Probabilities

In this section, we address the problem of trend classification (or the identification of the dose-response curve shape). For each gene declared significant in the first step, the set $\{g_1, g_2, g_3, g_4, g_5, g_6, g_7\}$, given in Table 9.1 and shown in Fig. 9.1, is the set of seven possible models with increasing trend for an experiment with four dose levels. Analogously, a set of seven models with decreasing trend is considered as well (see Fig. 10.1). Note that a model selection procedure that leads to a selection of the best model from the set of all candidate models will allow one to identify both dose-response curve shape and the minimum effective dose (MED). On the other hand, using hypothesis testing for the determination of the MED will not always allow us to identify the shape of the dose-response curve.

To select the best model, we propose to use the posterior probability of the model g_r given the data (Burnham and Anderson 2002), defined by

$$P(g_r|D) = \frac{P(D|g_r)P(g_r)}{\sum_{r=1}^{R} P(D|g_r)P(g_r)}, \quad r = 1, \ldots, R. \tag{10.1}$$

Here, $P(D|g_r)$ and $P(g_r)$ are the likelihood and the prior probabilities of the rth model, respectively, and R is the number of all candidate models (for the order-restricted ANOVA models, $R = 7$ for our case study). Note that if we use a non-informative prior, i.e., $P(g_r) = 1/R$ (Whitney and Ryan 2009), the isotonic regression model has the highest posterior probability since the maximum likelihood estimate is unique. However, the posterior model probabilities (10.1) do not take the complexity of the model into account. In what follows, we focus on the model selection procedures based on information criteria, which take into account both the goodness-of-fit and the model complexity. In other words, we focus on the question whether changing the number of parameters in the model will lead to a better compromise between goodness-of-fit and model complexity.

10.2.2 Akaike Weights and Bayesian Posterior Model Probabilities

10.2.2.1 Akaike Weights

The model selection theory discussed by Burnham and Anderson (2002, 2004) and Claeskens and Hjort (2008) allows one to incorporate the need to balance between goodness-of-fit and model complexity within the model selection procedure. Our starting point for Burnham and Anderson's model selection theory is the Kullback–Leibler (KL) information given by Burnham and Anderson (2002) and Claeskens and Hjort (2008):

$$I(f, g) = \int f(x) \log \frac{f(x)}{g(x|\boldsymbol{\theta})} \mathrm{d}x. \tag{10.2}$$

Here, f represents the density function of the true and unknown model, g represents the density function of the model that is used to approximate f, and $\boldsymbol{\theta}$ is the unknown parameter to be estimated. The KL information (or the KL distance) is interpreted as the loss of information when the true model f is approximated by the model $g(x|\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the parameter estimate for the unknown parameter $\boldsymbol{\theta}$. For a given set of candidate models $\{g_1, g_2, \ldots, g_R\}$, one can compare the KL information for each model and select the model that minimizes the information loss across the considered set of models (Burnham and Anderson 2002, 2004; Poeter and Anderson 2005). However, in practice, $I(f, g)$ cannot be computed since the true model f is unknown.

Akaike (1973, 1974) and Burnham and Anderson (2002) made the link between the KL information and likelihood theory and showed that the expected Kullback–Leibler information can be expressed as

$$\hat{E}(\mathrm{KL}) = -\log L(\boldsymbol{\theta}|D) + M, \tag{10.3}$$

where $L(\boldsymbol{\theta}|D)$ is the likelihood and M is the number of parameters in the model. The well-known Akaike's information criterion (AIC) is given by

$$\mathrm{AIC} = -2 \log L(\boldsymbol{\theta}|D) + 2M. \tag{10.4}$$

Akaike's approach allows for model selection that takes into account both goodness-of-fit and model complexity. Because the individual AIC values are not interpretable, as they contain arbitrary constants and are much affected by sample size, for a given set of R models, Burnham and Anderson (2004) proposed to rescale the AIC to

$$\Delta\mathrm{AIC}_r = \mathrm{AIC}_r - \mathrm{AIC}_{\min}, \tag{10.5}$$

with $\mathrm{AIC}_{\min}$ being the smallest AIC value across the set of R models. The AIC difference, $\Delta\mathrm{AIC}_r$, is interpreted as the information loss when model g_r, rather than the best model $g_{\min}$, is used to approximate f. Some simple rules of thumb are used

in assessing the relative merits of the models in the set (Burnham and Anderson 2002): models with $\Delta\text{AIC}_r \leq 2$ have substantial support (evidence); those with $4 \leq \Delta\text{AIC}_r \leq 7$ have considerably less support; and models with $\Delta\text{AIC}_r > 10$ have essentially no support.

Akaike (1981) and Burnham and Anderson (2002) advocated the use of $\exp(-\frac{1}{2}\Delta\text{AIC}_i)$ for the relative likelihood of the model given the data, defined by

$$L(g_r|D) \propto \mathrm{e}^{-\frac{1}{2}\Delta\text{AIC}_r}. \tag{10.6}$$

Note that the model likelihood $L(g_r|D)$ takes into account both goodness-of-fit and model complexity, while $P(g_r|D)$ takes into account only goodness-of-fit. Similar to the posterior probabilities in (10.1), Akaike (Burnham and Anderson 2002) defined Akaike's weights by

$$w_r = P_A(g_r|D) = \frac{\exp(-\frac{1}{2}\Delta\text{AIC}_r)P(g_r)}{\sum_{r=1}^{R}\exp(-\frac{1}{2}\Delta\text{AIC}_r)P(g_r)}. \tag{10.7}$$

Akaike's weight $P_A(g_r|D)$ can be interpreted as the weight of evidence that model g_r is the best KL model given a set of R models and given that one of the models in the set must be the best KL model. Claeskens and Hjort (2008) referred to $P_A(g_r|D)$ as the smooth AIC weight. Note that Akaike's weights can be interpreted as the posterior probabilities of the models.

10.2.2.2 Bayesian Posterior Model Probabilities

The BIC, proposed by Schwarz (1978), is given by

$$\text{BIC} = -2\log L(\boldsymbol{\theta}|D) + M\log(N), \tag{10.8}$$

where N is the number of arrays in the data and M is the number of parameters in the model. The BIC uses a higher penalty on the number of observations than the AIC, which penalizes on the number of parameters in the model. Therefore, the BIC leads to the selection of less complex models. The posterior model probabilities are given by

$$P_B(g_r|D) = \frac{\exp(-\frac{1}{2}\Delta\text{BIC}_r)P(g_r)}{\sum_{r=1}^{R}\exp(-\frac{1}{2}\Delta\text{BIC}_r)P(g_r)}. \tag{10.9}$$

Whitney and Ryan (2009) referred to $P_B(g_r|D)$ as the approximation to the model posterior probability. Burnham and Anderson (2002) showed that for prior probabilities

$$P(g_r) = B\exp(\frac{1}{2}\Delta\text{BIC}_r)\exp(-\frac{1}{2}\Delta\text{AIC}_r), \tag{10.10}$$

it follows that $P_B(g_r|D) = P_A(g_r|D)$ where B is a constant.

10.3 Order-Restricted Information Criterion

Anraku (1999) proposed an information criterion for parameters under an order restriction, the ORIC. In contrast to the AIC, the ORIC method is particularly suitable for detecting the configuration of the isotonic means under order restriction.

In contrast to the AIC and BIC, which penalize on the number of parameters in the model and/or on the sample size, the ORIC takes into account the level probability for the number of parameters under each order-restricted model in set R. The ORIC proposed by Anraku (1999) is given by,

$$\text{ORIC} = -2\log L(\boldsymbol{\theta}|D) + 2\sum_{\ell=1}^{K} \ell P(\ell, K, \boldsymbol{w}),$$

with $P(\ell, K, \boldsymbol{w})$ the level probability discussed in Chap. 2 (Robertson et al. 1988), $w_i = n_i/\tau_i$ where n_i is the number of arrays at dose i ($i = 0, \ldots, K$) and τ_i is the variance at dose i. For the case that $w_i = \cdots = w_K$, or $K = 2$, it follows that $\text{ORIC} = -2\log L(\boldsymbol{\theta}|D) + \sum 1/\ell$ (Robertson et al. 1988; Anraku 1999).

The posterior model probabilities (with non-informative priors) are given by

$$P_{OR}(g_r|D) = \frac{\exp(-\frac{1}{2}\Delta\text{ORIC}_r)}{\sum_{r=1}^{R}\exp(-\frac{1}{2}\Delta\text{ORIC}_r)}. \tag{10.11}$$

Under simple order alternatives, i.e., H_1^{Up} or H_1^{Down} [see (3.3] and (3.4), respectively], in the setting of four doses (control and three higher doses) and equal number of arrays per dose, the level probabilities $P(\ell, K, \boldsymbol{w})$, given by Robertson et al. (1988), are equal to 0.25, 0.45833, 0.25, and 0.04167 for $\ell = 1, 2, 3$, and 4, respectively. The smallest value of the ORIC for the model in the set of R possible models indicates the best configuration of the parameters under order restriction in terms of the compromise between goodness-to-fit and model complexity.

Table 10.1 lists the values of the penalty used by the AIC, BIC, and ORIC. Note that the relative magnitude of the penalty used by the ORIC for models g_1, g_2, g_3 (two parameters) and models g_4, g_5, g_6 (three parameters) is larger than that used for model g_7 (four parameters). Thus, due to the small penalty difference between models g_4, g_5, g_6, and model g_7, the latter is more likely to be classified as the best model in the set of R possible models.

10.4 Classification of Dose-Response Trends Using the One- and Two-Stage ORICC Algorithms

The one-stage ORICC was proposed by Liu et al. (2009a, 2009b) in the context of time-course microarray data and a similar algorithm, in the context of dose-response microarray data, was proposed by Lin et al. (2009) in order to classify monotone

gene profiles (with respect to dose). Liu's and Lin's algorithms are similar in the sense that they both used the ORIC for clustering. However, they are different in their focus. Lin's algorithm proposes for clustering of monotone (increasing or decreasing) profiles into their subset profiles while Liu's algorithm treats all monotone profiles as one cluster but allows for clustering of any order-restricted profiles (including monotone, umbrella, and cyclical profiles). In addition, the algorithm of Lin et al. (2009) is based on Anraku's information criterion while Liu et al. (2009b) developed an information criterion for the setting of non-monotone profiles. The two-stage ORICC algorithm is similar to the one-stage ORICC algorithm but includes an initial filtering stage in order to reduce computation time. Liu et al. (2009b) proposed a filtering procedure based on model selection while Lin et al. (2009) advocated the use of the likelihood ratio test for initial filtering. The ORICC of Liu et al. (2009b) will be discussed further in Chap. 11.

The two-stage ORICC algorithm (in Lin et al. 2009) is defined as follows:

1. Initial filtering. Use the LRT to select genes with significant monotone dose-response relationship.
2. Prespecify a collection of all possible order-restricted ANOVA models, $g_1, \ldots, g_R$.
3. Compute for each model

$$\text{ORIC}(g_1), \ldots, \text{ORIC}(g_R).$$

4. A gene is assigned to the rth cluster if

$$\text{ORIC}(g_r) = \min\{\text{ORIC}(g_1), \ldots, \text{ORIC}(g_R)\}.$$

In Lin et al. (2009), the initial filtering is done based on the likelihood ratio test. Genes for which the null hypothesis cannot be rejected are excluded from the analysis. The one-stage ORICC algorithm is similar to the two-stage algorithm, but it does not have the filtering stage, i.e., all genes are included for the analysis and the null model g_0 is fitted for each gene as well. Based on a simulation study, Lin et al. (2009) advocated the use of the two-stage ORICC algorithm for three reasons:

1. The initial filtering reduces the misclassification error in the model selection step.
2. The model selection procedure is applied to those genes for which there is an evidence of a monotone relationship between gene expression and dose. Hence, the best monotone model must be in the model set, as required.
3. The initial testing step reduces the computation time for the model selection, as the selection is applied for a relatively small number of genes and since the direction of the trend (upward or downward trend) is known from the initial step.

The method discussed above can be used to identify the MED (i.e., the first dose for which we observed a different response from the control dose) for the order-restricted ANOVA models. An alternative approach to determine the MED is the use of a hypothesis testing procedure, such as Williams' (1971, 1972) or Marcus'

(1976) procedure, that identifies the MED based on hypothesis testing in a step-down fashion. However, identification of the MED using these procedures does not imply the shape of the dose-response curve. On the other hand, model selection procedures based on information theory can address two objectives at the same time—the determination of the MED and the classification of trends. Once the best model from the candidate set is selected, one can identify the MED from the selected model.

10.5 Application to the Data

10.5.1 Classification of the Trends

Based on the 3,499 genes found to be significant in the initial inference step, the classification of mean profiles for these genes is to be obtained. For each gene, we fit seven models to the data. The log-likelihood of these models are obtained by using the following R code:

```
> dim(data.sign) ## 3499 significant genes rejected by the LRT
3499 12
> g1 <- as.factor(c(rep(1,9),rep(2,3)))
> loglik.g1 <- apply(data.sign, 1, function(genei) logLik(aov(genei~g1-1)))
> g2 <- as.factor(c(rep(1,6),rep(2,6)))
> loglik.g2 <- apply(data.sign, 1, function(genei) logLik(aov(genei~g2-1)))
> g3 <- as.factor(c(rep(1,3),rep(2,9)))
> loglik.g3 <- apply(data.sign, 1, function(genei) logLik(aov(genei~g3-1)))
> g4 <- as.factor(c(rep(1,3),rep(2,6),rep(3,3)))
> loglik.g4 <- apply(data.sign, 1, function(genei) logLik(aov(genei~g4-1)))
> g5 <- as.factor(c(rep(1,6),rep(2,3),rep(3,3)))
> loglik.g5 <- apply(data.sign, 1, function(genei) logLik(aov(genei~g5-1)))
> g6 <- as.factor(c(rep(1,3),rep(2,3),rep(3,6)))
> loglik.g6 <- apply(data.sign, 1, function(genei) logLik(aov(genei~g6-1)))
```

In the above code, `g1, g2, g3, g4, g5, g6,` and `g7` are R objects which define the design matrices. For example, `g1` and `g6` imply the following design matrices, respectively,

$$\mathbf{X}_{(g_1)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{X}_{(g_6)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Table 10.2 Numbers of genes classified for each model

Model	Likelihood	AIC	BIC	ORIC
g_1	344	1,528	1,648	1,348
g_2	25	307	369	221
g_3	14	106	126	86
g_4	343	370	337	253
g_5	885	823	715	655
g_6	178	170	149	120
g_7	1,710	195	155	816

Note that the log-likelihood of the first six models can be easily obtained after fitting the data with predefined constraints. For model g_7, the isotonic means are obtained by using the function `monoreg()` by specifying the increasing or decreasing trend as option.

```
> g7m <- matrix(0, nrow(data.sign), ncol(data.sign))
> dir <- IsoGenem(x.res,data.sign)[[11]]
> g7m[dir=="u",] <- t(apply(data.sign[dir=="u",],1,
+function(genei)monoreg(unique(x.res),tapply(genei,as.factor(x.res),mean)
                        ,type="isotonic")$yf))[,x.res]
> g7m[dir=="d",] <- t(apply(data.sign[dir=="d",],1,
+function(genei)monoreg(unique(x.res),tapply(genei,as.factor(x.res),mean)
                         ,type="antitonic")$yf))[,x.res]
> residuals <- data.sign-g7m
> w <- rep.int(1, ncol(data.sign))
> N <- ncol(data.sign)
> loglik.g7 <- apply(residuals, 1, function(res)
+0.5*(sum(log(w)) - N * (log(2 * pi) + 1 - log(N) + log(sum(w * res^2)))))
> m2loglik.mat <- -2* cbind(loglik.g1,loglik.g2,loglik.g3,
+loglik.g4,loglik.g5,loglik.g6,loglik.g7)
```

Using the four information criteria (the likelihood, AIC, BIC, and ORIC) the genes are classified into the seven curve shapes for each direction (see Table 10.2 and Fig. 10.2a). For the likelihood-based posterior probabilities defined in (10.1), 1,710 genes (48.85%) are classified as g_7—the isotonic regression model with four parameters. As shown in Table 10.2, when the AIC and BIC criteria are used to calculate the posterior probabilities, the number of genes that are classified as g_1 (isotonic regression with two parameters) increases from 344 to 1,528 and 1,648, respectively. The same pattern is observed for models g_2 and g_3 (both are isotonic regression models with two parameters). For the ORIC, the number of genes classified to models from g_1 to g_6 decreases as compared to the AIC and BIC. On the other hand, 816 genes are classified as g_7.

Figure 10.2b shows the data for gene 3467. Based on the likelihood and ORIC, the best model for the gene is g_7 (dashed line). Using the AIC, the gene is classified as g_6 (solid line). For both models, the second dose level is estimated to be the MED level. Using the BIC, the model is further reduced and the gene is classified as g_2 (dotted line), and the MED level is estimated to be the third dose level. Figure 10.3 shows the clusters for genes which are classified as g_3 and g_6 according to the ORIC. We notice that there is a large variability between these genes. This is expected since

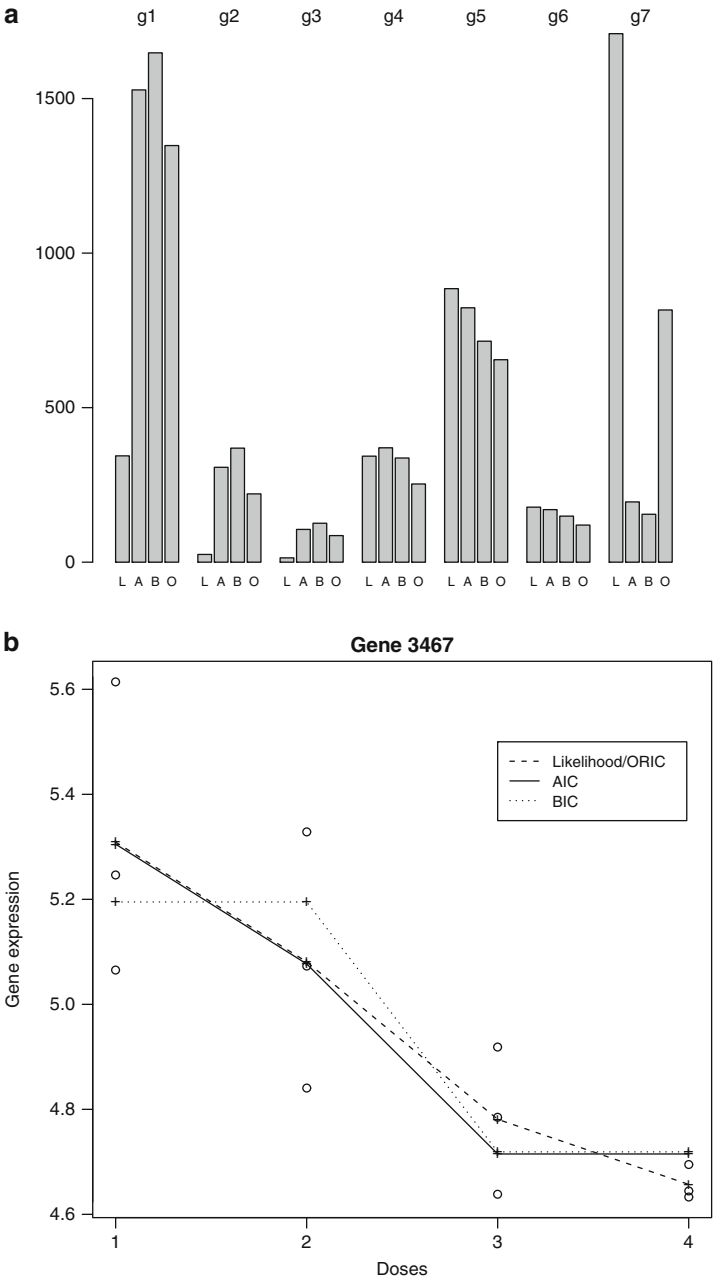

Fig. 10.2 Classification based on information criteria. (**a**) Classification of trends based on likelihood (L), AIC (A), BIC (B), and ORIC (O). (**b**) The best model according to the likelihood, AIC, BIC, and ORIC for gene 3467. *Dashed line* is the isotonic regression model g_7 (four parameters) selected by the likelihood and ORIC, *Solid line* is g_6 (three parameters) selected by the AIC, and *dotted line* is g_2 (two parameters) selected by the BIC

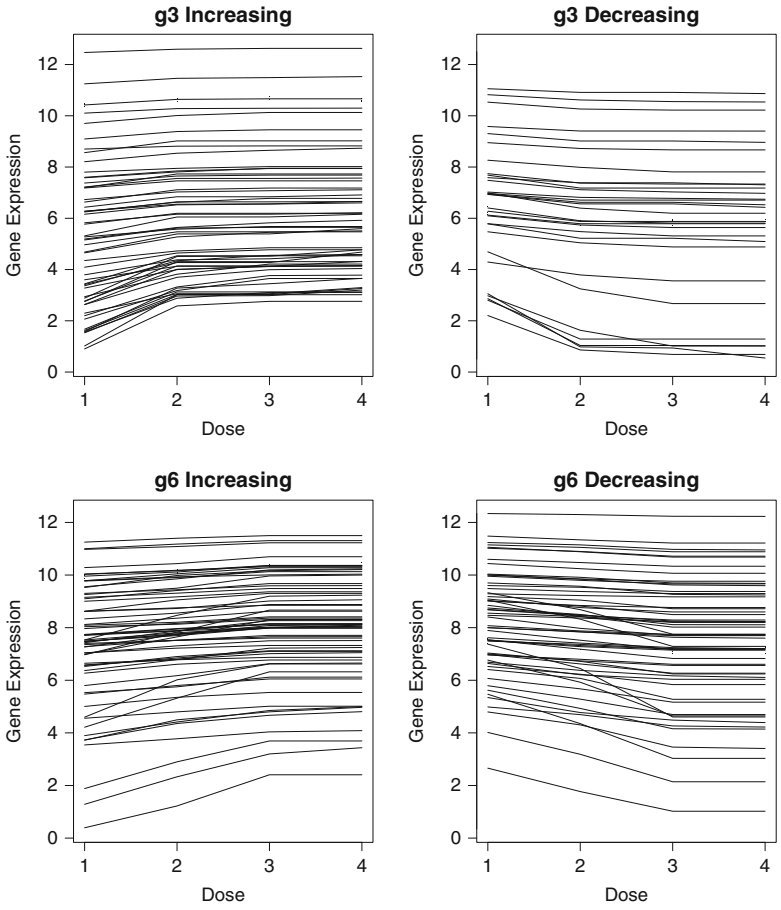

Fig. 10.3 Classification of trends based on the AIC for the clusters of g_3 and g_6

the classification is based on information criteria and does not aim to reduce the variability between genes.

Note that, although the three models (g_1, g_2, and g_3) have the same number of parameters, the effective dose levels are not the same. The MED for model g_1 is the highest dose level, while the effective dose levels for g_2 and g_3 are the third and the second dose levels, respectively. In general, the AIC and BIC criteria favor the two-parameter models as compared to the three-parameter models g_4, g_5, g_6, and the four-parameter model g_7, which is favored by the likelihood. Table 10.2 shows that the AIC and BIC classify most of the genes as g_1 and g_5, the ORIC criterion classifies most of the genes as g_1 and g_7, while the likelihood criterion classifies most of the genes as g_7.

10.6 Discussion

In this chapter, we have investigated the issue of classifying dose-response curve shapes of the genes for which a monotone dose-response relationship was detected. After the initial inference step, the set of all possible models is considered and a non-informative prior is given to these models. The likelihood, AIC, BIC, and ORIC are used to assess the goodness-of-fit and complexity of the models. The Akaike weights, representing the posterior probabilities of the models, are calculated from the AIC of the models; Bayesian posterior probabilities are obtained from the BIC; the posterior probability obtained from the ORIC takes order restriction into account. We have shown that using the likelihood to calculate the posterior probabilities of the models leads to the selection of the isotonic regression model that maximizes the likelihood under the order constraints. In our example , for the setting of four doses, the ORIC tends to classify genes as model g_7 as apposed to models g_4, g_5, and g_6 that have one parameter less. And, as expected, using the AIC and BIC leads to a selection of simpler models with less parameters (compared with the model selection based on the likelihood). Moreover, we have shown that after a classification is made, the mean structure of the selected model can be used to estimate the MED level.

One can rightly argue that posterior probabilities of the model discussed in Sects. 10.2.2 and 10.3 are not needed for the model selection procedure discussed in this chapter. Selection based on the information criterion or the posterior probability based on the information criterion is equivalent. The importance of the posterior probability of the model will be discussed in Chap. 14 where the posterior probabilities are used in order to calculate a weighted average for the parameters of primary interest. In this chapter, the classification procedure should be seen as a classification post-selection since a gene is assigned to a cluster after the model selection procedure. Thus, only one model, the one with the smallest information criterion, is used for the classification step. In Chap. 14, we discuss the model averaging methodology for the estimation of the dose-response curve using parametric dose-response models. Within this modeling approach, the dose-response curve for each gene is estimated using weighted information (according to the posterior probabilities) obtained from all candidate models.

The clustering procedures discussed in this chapter were performed under the assumption of monotone dose-response relationship. In some experiments, this assumption may not be appropriate. In the next chapter, we discuss two clustering approaches for order-restricted dose-response (but not necessary monotone) gene profiles.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov, & B. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Academiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics, 16*, 3–14.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika, 86*(1), 141–152(12).

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information—Theoretic approach*. New York: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods Research, 33*, 261–304.

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging.* Cambridge Series in Statistical and Probabilistic Mathematics.

Lin, D., Shkedy, Z., Burzykowki, T., Aerts, M., Göhlmann, H. W. H., De Bondt, A., et al. (2009). Classification of trends in dose-response microarray experiments using information theory selection methods. *The Open Applied Informatics Journal, 3*, 34–43.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009a). *Order-restricted information criterion-based clustering algorithm. Reference manual*. http://cran.r-project.org/web/packages/ORIClust/.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009b). Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments. *BMC Bioinformatics, 10*, 146.

Marcus, R. (1976). The powers of some tests of the quality of normal means against an ordered alternative. *Biometrika, 63*, 177–83.

Poeter, E., & Anderson, D. (2005). Multimodel ranking and inference in ground water modeling. *Ground Water, 43*(4), 597–605.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Schwarz, M. J. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics, 27*, 103–117.

Whitney, M., & Ryan, L. (2009). Quantifying dose-response uncertainty using bayesian model averaging. In Cooke, R. M. (Ed.), *Uncertainty modeling in dose-response*. New York: Wiley.

Chapter 11
Beyond the Simple Order Alternatives

Dan Lin and Ziv Shkedy

11.1 Introduction

Up to this point in the book, we focused on estimation and inference for monotone mean profiles. In Chaps. 7 and 8, the null hypothesis of no dose effect was tested against order alternatives of $\mu(d_0) \leq \mu(d_1) \leq \mu(d_2) \leq \mu(d_3)$ or $\mu(d_0) \geq \mu(d_1) \geq \mu(d_2) \geq \mu(d_3)$ (with at least one strict inequality), and in Chaps. 9 and 10, we discussed two-stage clustering procedures for the subgroup of genes which were found to be significant. The ordered alternatives discussed in the previous chapters are called *simple order alternatives*, and the underlying assumption is that there is a monotone relationship between the dose and the mean gene expression. Typical mean profiles which satisfy a simple order are shown in Fig. 11.1a, d. A second assumption that was made in the previous chapters is that the variance is equal across all dose levels, i.e., $Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$. In this chapter, we relax these assumptions and discuss the case of testing the null hypothesis against order-restricted, but not necessarily monotone, alternatives assuming heteroscedastic variances. The order-restricted alternatives we consider in this chapter are the *unimodal partial order (umbrella profiles)* alternatives (Robertson et al. 1988; Bretz and Hothorn 2003; Peddada et al. 2003, 2005, 2009, 2010; Simmons and Peddada 2007; Liu et al. 2009b). Typical mean profiles which follow an umbrella dose-response curve shape are shown in Fig. 11.1b, c, e, f. We discuss two algorithms for inference and clustering of ordered-restricted gene expression data. The first, the ORIOGEN, proposed by Peddada et al. (2003), can be used for inference and clustering. The algorithm

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

Z. Shkedy
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_11,

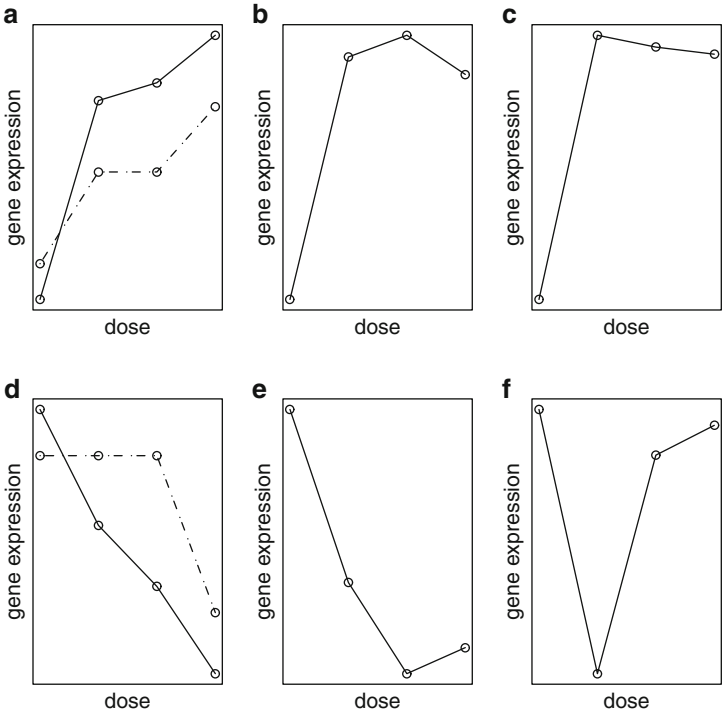

Fig. 11.1 Examples of order-restricted profiles in a dose-response experiment with four dose levels. Panel (**a**): increasing profiles. Panel (**b**): an umbrella profile with downturn at the third dose level. Panel (**c**): an umbrella profile with downturn at the second dose level. Panel (**d**): decreasing profiles. Panel (**e**): an inverted umbrella profile with upturn at the third dose level. Panel (**f**): an inverted umbrella profile with upturn at the second dose level

is discussed in Sect. 11.2 and the `ORIOGEN` 3.0 package (Peddada et al. 2005) is used for the analysis. An elaborate discussion about the `ORIOGEN` 3.0 package is given in Chap. 18. In Sect. 11.2.1, we test the null hypothesis against a simple order alternative under heteroscedastic variances and discuss the similarities to the analyses presented in Chaps. 7 and 8. In Sect. 11.2.2, the null hypothesis is tested against order-restricted alternatives (either simple order or partial order alternatives). The second algorithm we discuss in this chapter is the ORICC algorithm, proposed by Liu et al. (2009b). Similar to the analysis discussed in Chap. 10, the ORICC algorithm clusters order-restricted gene profiles using information criterion. The ORICC algorithm is implemented in the R package `ORIClust` and discussed in Sect. 11.3.

11.2 The ORIOGEN Algorithm

The ORIOGEN algorithm (order-restricted inference for ordered gene expression) was proposed by Peddada and colleagues in a series of manuscripts from 2003 onward. The ORIOGEN method can be used for both inference and clustering

under order restrictions, and in contrast with the analysis presented in the previous chapters, the ORIOGEN algorithm is not focused only on simple order alternatives. In Sect. 11.2.1, we present an analysis under simple order alternatives for the case of heteroscedastic variances, while the analysis under partial order alternatives is discussed in Sect. 11.2.2.

11.2.1 Testing H_0 Versus a Simple Order Alternative

Similar to Chap. 3, we consider a gene-specific one-way ANOVA model and assume heteroscedastic variances:

$$Y_{ij} \sim \mathrm{N}(\mu(d_i), \sigma_i^2), \quad i = 0, 1, \ldots, K, \ j = 1, 2, \ldots, n_i. \tag{11.1}$$

Our primary interest is to test the null hypothesis of no dose effect $H_0 : \mu(d_0) = \mu(d_1) = \cdots = \mu(d_K)$ against the simple order alternatives $H_1^{\mathrm{Up}} : \mu(d_0) \leq \mu(d_1) \leq \cdots \leq \mu(d_K)$ or $H_1^{\mathrm{Down}} : \mu(d_0) \geq \mu(d_1) \geq \cdots \geq \mu(d_K)$ with at least one strict inequality. To test the null hypothesis against simple order alternatives, the `ORIOGEN` 3.0 package performs a resampling-based SAM analysis using a SAM t-type test statistics given, respectively, for increasing and decreasing profiles by

$$l_{\mathrm{Up}}^{\infty} = \frac{\hat{\mu}_K^* - \hat{\mu}_0^*}{(s_0 + s)\sqrt{\frac{1}{n_K} + \frac{1}{n_0}}} \quad \text{and} \quad l_{\mathrm{Down}}^{\infty} = \frac{\hat{\mu}_0^* - \hat{\mu}_K^*}{(s_0 + s)\sqrt{\frac{1}{n_0} + \frac{1}{n_K}}}.$$

Note that for a given value of the fudge factor s_0, this specific analysis in `ORIOGEN` is similar to the SAM analysis in `IsoGene` discussed in Chap. 8. The main difference is that `IsoGene` uses permutation-based inference of the actual expression data in which the columns of the expression matrix are permuted (and therefore assumes variance homogeneity across the dose levels) while `ORIOGEN` approximates the distribution of the test statistic by bootstraping residuals. The bootstrap algorithm is discussed in Simmons and Peddada (2007), Gou and Peddada (2008) and Peddada et al. (2010).

In order to perform the analysis using the `ORIOGEN` package, we need to specify that the mean profiles of primary interest for the analysis are increasing or decreasing. For the FDR level equal to 5% and $s_0 = s_{10\%}$ (i.e., s_0 equals to tenth percentile of the standard errors of the test statistics), we obtained 3,459 significant genes from which 1,187 and 2,272 are clustered as increasing and decreasing profiles, respectively. We consider a second analysis (with 1,000 bootstraps) without the fudge factor, i.e., $s_0 = 0$. The number of significant genes is equal to 3,151 from which 1,716 and 1,435 are clustered as increasing and decreasing profiles, respectively. In Chap. 7, for Marcus' test statistic, 3,533 genes were found to be significant (using permutation-based inference with FDR-BH = 5%). In total, 2,957 significant genes are found to be common for the two approaches. An example of

three genes found to be significant is shown in Fig. 11.2, and the partial output of the ORIOGEN package is presented in the panel below. Note that the predicted means for a simple order profile are obtained from the isotonic regression on the observed means.

```
#PARTIAL OUTPUT OF THE ORIOGEN.

FDR Level   0.05
S0 Percentile   0.1
Longitudinal sampling   FALSE
Bootstrap random seed   Automatic
Profile Selections:
1   Decreasing profile
2   Increasing profile


Results:
Gene ID Profile #   P Value Q Value  Fit.1  Fit.2    Fit.3  Fit.4

 g1      2          0,001    0,00358  7,039  7,234    7,403  7,795
 g3      2          0,002    0,00607  5,567  5,567    5,567  6,911
g10      2          0,001    0,00358  6,604  6,604    6,969  8,993
```

11.2.2 Testing H_0 Against Partial Order (Umbrella) Alternatives

The main advantage of the ORIOGEN package is that it allows to test the null hypothesis against any ordered alternative of interest. To keep notation in line with Peddada et al. (2003), we denote C_r as a possible order-restricted profile. For a dose-response experiment with four dose levels ($K+1=4$), there are six noncyclical order-restricted profiles given by

$$
\begin{aligned}
C_1 &= \left\{\mu \in R^{K+1} : \mu_0 \le \mu_1 \le \mu_2 \le \mu_3\right\}, && \text{increasing profile,}\\
C_2 &= \left\{\mu \in R^{K+1} : \mu_0 \le \mu_1 \ge \mu_2 \ge \mu_3\right\}, && \text{umbrella profile, downturn at 2,}\\
C_3 &= \left\{\mu \in R^{K+1} : \mu_0 \le \mu_1 \le \mu_2 \ge \mu_3\right\}, && \text{umbrella profile, downturn at 3,}\\
C_4 &= \left\{\mu \in R^{K+1} : \mu_0 \ge \mu_1 \ge \mu_2 \ge \mu_3\right\}, && \text{decreasing profile,}\\
C_5 &= \left\{\mu \in R^{K+1} : \mu_0 \ge \mu_1 \le \mu_2 \le \mu_3\right\}, && \text{inverted umbrella profile, upturn at 2,}\\
C_6 &= \left\{\mu \in R^{K+1} : \mu_0 \ge \mu_1 \ge \mu_2 \le \mu_3\right\}, && \text{inverted umbrella profile, upturn at 3.}
\end{aligned}
\tag{11.2}
$$

The profiles C_1 and C_4 are the simple order profiles that were tested in the previous section. The algorithm implemented in the ORIOGEN package consists of the following steps:

- Specify the set of candidate profiles of primary interest $C_1, C_2, \ldots, C_R$.
- For each gene, estimate the mean under each candidate profile.
- Calculate the goodness of fit statistic, $l_\infty^1, \ldots, l_\infty^R$, for each candidate model.
- Use a bootstrap algorithm to approximate the distribution of the test statistic (the goodness of fit statistic) $l_\infty^r = \max(l_\infty^1, \ldots, l_\infty^R)$ under the null hypothesis.

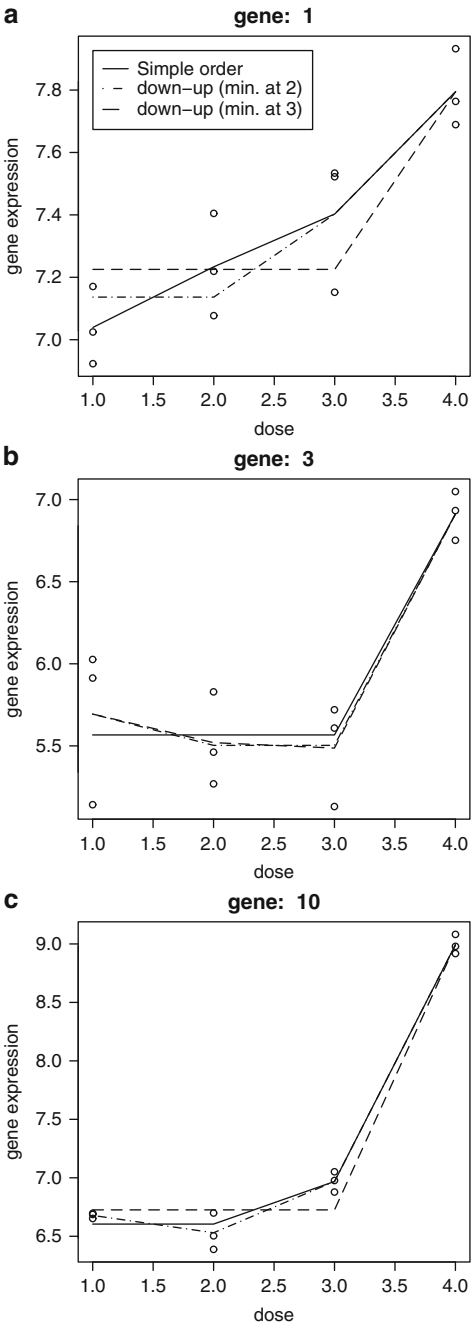

Fig. 11.2 Testing the null hypothesis against simple order alternative. Example of three genes with significant dose-response relationship. *Solid line*: the estimated mean expression under simple order restriction. *Dashed line*: inverted umbrella with turn point at the third dose level. *Dotted-dashed line*: inverted umbrella with turn point at the second dose level. (**a**) Gene 1; (**b**) Gene 3; (**c**) Gene 10

- Once a gene is declared significant, assign a gene to the profile which has the largest goodness of fit statistic.

The methodology used in each step was discussed in details by Peddada et al. (2003), Simmons and Peddada (2007), and Peddada et al. (2010). In what follows, we discuss briefly the calculation of the test statistic and the testing (and clustering) procedures.

11.2.2.1 Calculation of the Test Statistic

Let us consider a dose-response experiment with $K + 1$ dose levels and assume that there are two candidate profiles of primary interest. A simple order profile $C_1 = \{\mu \in R^{K+1} : \mu_0 \leq \mu_1 \leq \cdots \leq \mu_K\}$ and an umbrella-shaped profile with downturn at dose level s, $C_2 = \{\mu \in R^{K+1} : \mu_0 \leq \mu_1 \leq \cdots \leq \mu_s \geq \mu_{s+1} \geq \ldots \geq \mu_K\}$. In order to calculate the goodness of fit statistic for each candidate profile, Peddada et al. (2003) and Simmons and Peddada (2007) used the following definitions:

1. *Linked parameters*: two parameters μ_i and μ_j are linked if the inequality between them is specified by the order restrictions of the profile (Peddada et al. 2010). For example, for the simple order profile C_1, all the parameters are linked. However, for an umbrella-shaped profile, μ_i and μ_j are linked for $0 \leq i < j \leq s$ and for $s \leq i < j \leq K$ but not for $0 \leq i < s$ and $s < j \leq K$. In other words, we can define two subvectors:

$$C_{21} = \{\mu \in R^s : \mu_0 \leq \mu_1 \leq \cdots \leq \mu_s\},$$
$$C_{22} = \{\mu \in R^{K+2-s} : \mu_K \leq \mu_{K-1} \leq \cdots \leq \mu_s\},$$

   for which all the parameters in C_{21} are linked and all the parameters in C_{22} are linked, but the parameters in C_{21}, except μ_s, are not linked with the parameters in C_{22}. The parameter μ_s is linked with all the parameters in the profile, and it is called a *nodal* parameter (Peddada et al. 2003).
2. A *subgraph* is formed by a subvector of the profile (such as C_{21} and C_{22}). A subgraph is a *linked subgraph* if all the parameters in the subvectors are linked (Simmons and Peddada 2007). A linked subgraph is a *maximal linked subgraph* if all other linked subgraphs are subvectors of the subgraph. Hence, C_{21} and C_{22} are maximal linked subgraphs for the umbrella-shaped profile and C_1 is a maximal linked subgraph of a simple order profile.
3. The *goodness of fit statistic* of a profile is the maximum standardized difference between the parameter estimates of the *farthest* linked parameters of all the maximal linked subgraphs of the profile (Simmons and Peddada 2007). For example

   - For the simple order profile, C_1 is a maximal subgraph, and therefore, the goodness of fit statistic is the standardized difference between the parameter estimates of the two farthest parameters of C_1, that is,

$$l_1^{\infty} = \frac{\hat{\mu}_K^* - \hat{\mu}_0^*}{\sqrt{\frac{\hat{\sigma}_K^2}{n_K} + \frac{\hat{\sigma}_0^2}{n_0}}}.$$

We notice that for the simple order profiles, the goodness of fit statistic is Marcus' statistic for the case of heteroscedasticity.

- For the umbrella-shaped profile, the goodness of fit statistic is the maximum between the standardized difference of the farthest parameter estimates of C_{21} and C_{22}, that is,

$$l_2^{\infty} = \max\left(\frac{\hat{\mu}_s^* - \hat{\mu}_0^*}{\sqrt{\frac{\hat{\sigma}_s^2}{n_s} + \frac{\hat{\sigma}_0^2}{n_0}}}, \frac{\hat{\mu}_s^* - \hat{\mu}_K^*}{\sqrt{\frac{\hat{\sigma}_s^2}{n_s} + \frac{\hat{\sigma}_K^2}{n_K}}} \right).$$

Hence, for an umbrella with a downturn point at dose level s, the goodness of fit statistic is the maximum of Marcus' statistics calculated for each subgraph.

Under the assumption of homoscedastic variance the parameter estimates for dose-specific variance can be replaced by the pooled sample variance in both l_1^{∞} and l_2^{∞}. Both test statistics can be modified as SAM test statistics as discussed in the previous section.

11.2.2.2 Testing and Clustering

Let $C_1, \ldots, C_R$ be a set of candidate order-restricted profiles, $l_1^{\infty}, \ldots, l_R^{\infty}$ the corresponding test statistics, and $l_r^{\infty} = \max(l_1^{\infty}, \ldots, l_R^{\infty})$. A bootstrap algorithm is performed in order to approximate the distribution of the test statistic under the null hypothesis. A gene is declared significant if $l_r^{\infty} > z_{\alpha}^{\star}$ (the α percentile of the bootstrap distribution). A gene is assigned to profile C_r whenever it is found to be significant.

11.2.2.3 Analysis of the Human Epidermal Squamous Carcinoma Experiment

In order to carry out the analysis for simple and partial order alternatives, we need to specify all the noncyclical profiles as candidate profiles. As we mentioned above, for $K + 1 = 4$, there are six possible profiles.

```
### PARTIAL OUTPUT OF ORIOGEN ###

Profile Selections:
1   Decreasing profile
2   Umbrella profile, downturn at 2
3   Umbrella profile, downturn at 3
4   Increasing profile
5   Inverted umbrella profile, upturn at 2
6   Inverted umbrella profile, upturn at 3
```

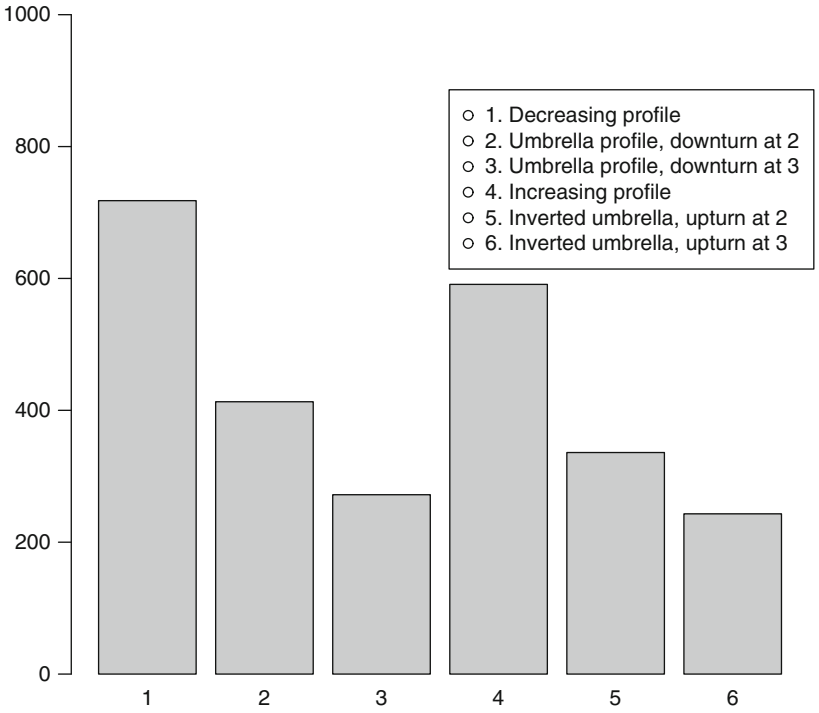

Fig. 11.3 Number of genes per cluster. Seven hundred and eighteen genes are clustered into C_1, 314 genes are clustered into C_2, 272 genes are clustered into C_3, 591 genes are clustered into C_4, 336 genes are clustered into C_5, and 243 genes are clustered into C_6

Let us focus again on genes 1, 3, and 10 presented in Fig. 11.2. Recall that these genes were found to be significant for the analysis discussed in Sect. 11.2.1 when only simple order alternatives were considered. As can be seen in the panel below, in the current analysis, the three genes are found be significant as well. Gene 1 is clustered into C_1 (an increasing profile), gene 3 is clustered into C_6 (an inverted umbrella profile with upturn at the third dose level), and gene 10 was clustered into C_5 (an inverted umbrella profile with upturn at the second dose level).

```
### PARTIAL OUTPUT OF ORIOGEN ###

Results:
Gene ID Profile #   P Value Q Value  Fit.1  Fit.2    Fit.3  Fit.4

 g1      4          0,001   0,00458  7,039  7,234    7,403  7,795
 g3      6          0,002   0,00767  5,694  5,52     5,487  6,911
g10      5          0,001   0,00458  6,678  6,53     6,969  8,993
```

For an analysis with $s_0 = 0$ and FDR $= 5\%$, we obtained 2,573 significant genes from which 718, 413, 272, 591, 336, and 243 genes are clustered to profiles C_1–C_6, respectively (see Fig. 11.3). Figure 11.4a shows the genes that are clustered based

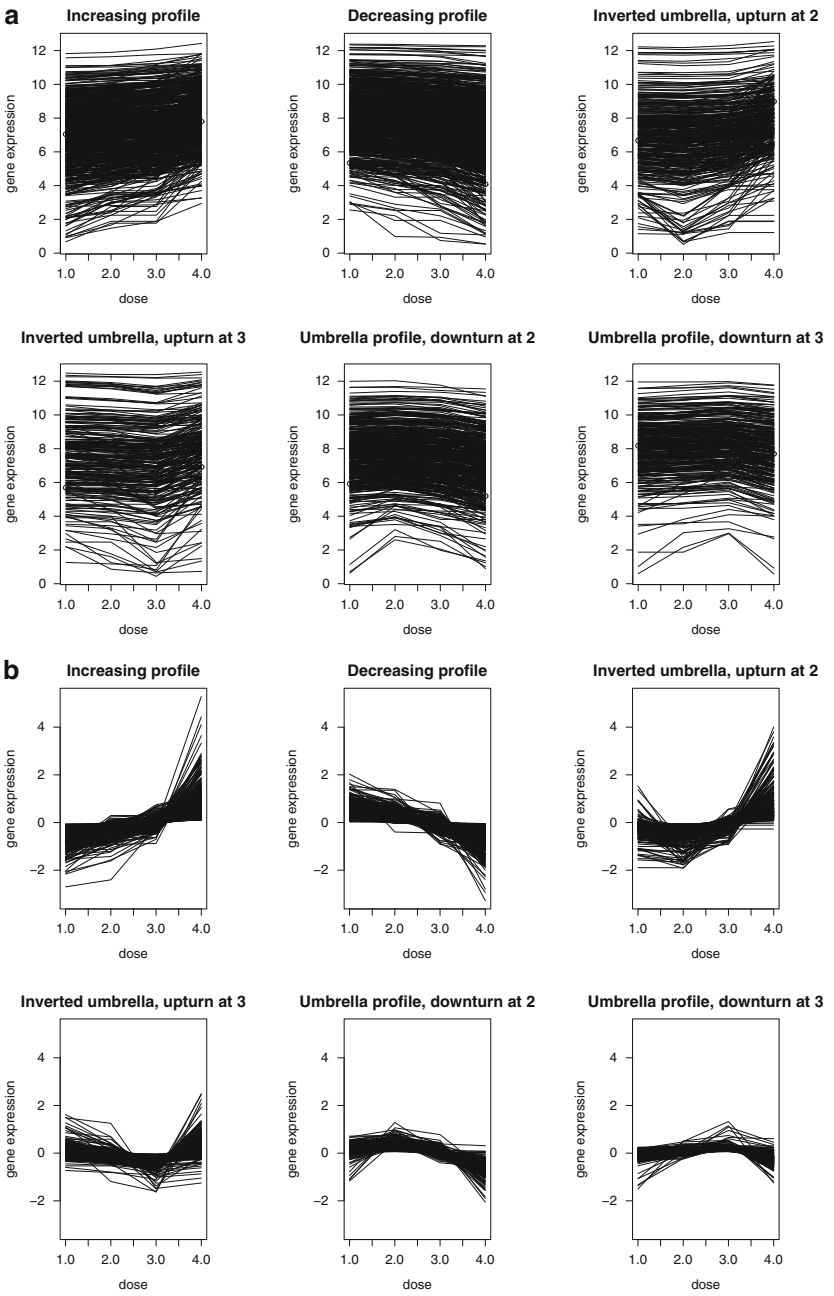

Fig. 11.4 Graphical output of the ORIOGEN 3.0 package. (**a**) Mean profiles for the genes that were found to be significant clustered into the six profiles of interest. (**b**) Six clusters with overall mean gene expression subtracted

on the six order-restricted profiles defined above. Figure 11.4b shows the genes in the clusters with the overall mean gene expression subtracted. After subtracting the means, it becomes evident to identify the dose-response trends for the six clusters.

11.3 The ORICC Algorithm

In the previous section, the ORIOGEN algorithm has two components: inference and clustering. Genes were clustered into a candidate profile only if they were found to have a significant dose-response relationship. This is similar to the two-stage approaches that were discussed in Chaps. 9 and 10. In both chapters, the initial filtering step was an inference step, and a clustering method was applied in the second step. Recall that the clustering procedure of monotone genes profiles, discussed in Chap. 10, was applied to simple order profiles:

$$\begin{aligned} C_1 &= \left\{\mu \in R^{K+1} : \mu_0 \leq \mu_1 \leq \mu_2 \leq \mu_3\right\}, \quad \text{increasing profile,} \\ C_4 &= \left\{\mu \in R^{K+1} : \mu_0 \geq \mu_1 \geq \mu_2 \geq \mu_3\right\}, \quad \text{decreasing profile.} \end{aligned}$$

In Chap. 10, we decomposed the simple order profile into all possible sub profiles, given by (for increasing profiles and $K+1=4$)

$$\begin{aligned} g_1 &= \left\{\mu \in R^{K+1} : \mu_0 < \mu_1 < \mu_2 < \mu_3\right\}, \\ g_2 &= \left\{\mu \in R^{K+1} : \mu_0 < \mu_1 = \mu_2 = \mu_3\right\}, \\ g_3 &= \left\{\mu \in R^{K+1} : \mu_0 < \mu_1 < \mu_2 = \mu_3\right\}, \\ g_4 &= \left\{\mu \in R^{K+1} : \mu_0 < \mu_1 = \mu_2 < \mu_3\right\}, \\ g_5 &= \left\{\mu \in R^{K+1} : \mu_0 = \mu_1 = \mu_2 < \mu_3\right\}, \\ g_6 &= \left\{\mu \in R^{K+1} : \mu_0 = \mu_1 < \mu_2 < \mu_3\right\}, \\ g_7 &= \left\{\mu \in R^{K+1} : \mu_0 = \mu_1 < \mu_2 = \mu_3\right\}. \end{aligned}$$

The ORIC (Anraku 1999) was used for clustering after an initial inference-based filtering step.

Similar to the previous section, the focus of the ORICC algorithm is not only on monotone profiles. Note that the main difference between the initial filtering step implemented in the ORIOGEN algorithm and the ORICC algorithm is that the former uses an inference-based initial filtering step while the latter is a model selection-based algorithm. Note that the order-restricted information criterion implemented in the ORICC algorithm is different from Anraku's ORIC, and it was discussed by Liu et al. (2009b). For the analysis presented in this section, we use the `ORIClust` R package (Liu et al. 2009a).

11.3.1 The One- and Two-Stage ORICC Algorithm of Liu et al. (2009b)

Let us consider a dose-response experiment with $K+1$ dose levels for which the ordered-restricted profiles of interest are given in (11.2). The one-stage ORICC algorithm proposed by Liu et al. (2009b) is as follows:

1. Specify a set of candidate order-restricted profiles $C_1, \ldots, C_R$.
2. For each gene, calculate the ORIC(r) and assign a gene to the candidate profile from which ORIC(r) is minimum.

The two-stage ORICC algorithm is similar to the one-stage algorithm but, in order to reduce computation time, has a filtering stage in which the ORIC is calculated for the null model ($C_0 = \{\mu \in R^{K+1} : \mu_0 = \mu_1, \ldots, = \mu_K\}$) and for the unrestricted model ($C_{\text{UR}} = \{\mu \in R^{K+1} : \mu_0 \neq \mu_1 \neq, \ldots, \neq \mu_K\}$). A gene is considered for clustering if ORIC(C_0) > ORIC(C_{UR}).

11.3.2 Example: Fitting Order-Restricted Profiles Using the `ORIClust` Package

The ORICC algorithm is implemented in the R package `ORIClust`. The package consists of several functions to calculate order-restricted profiles. In our example, increasing and down-up umbrella profiles [C_1, C_2, and C_3, in (11.2)] can be estimated using the following code:

```
> #order restricted models for gene 537
> #dose and number of replicates per dose
> dose<-c(1,1,1,2,2,2,3,3,3,4,4,4)
> n.rep<-c(3,3,3,3)
> i<-573 # gene number
> #expression levels and unrestricted means
> Y573<-as.numeric(as.vector(data[i,]))
> my573<-tapply(Y573,as.factor(dose),mean)

> library(ORIClust)
> #order restricted models
> fit.1<-increasing(Y573,my573,n.rep) #increasing
> fit.3<-down.up(Y573,my573,n.rep,2)  #umbrella with up turn at 2
> fit.4<-down.up(Y573,my573,n.rep,3)  #umbrella with up turn at 3
```

Parameter estimates under each profile for gene 573 are shown in the panel below and in Fig. 11.5a. Note that for the down-up profile with minimum at the second dose level, the profile should be monotone from the second dose level and onward, and therefore, the observed means in the second and the third dose levels were pooled together. Figure 11.5b shows mean profiles estimated for gene 64 under decreasing and up-down profiles assumptions.

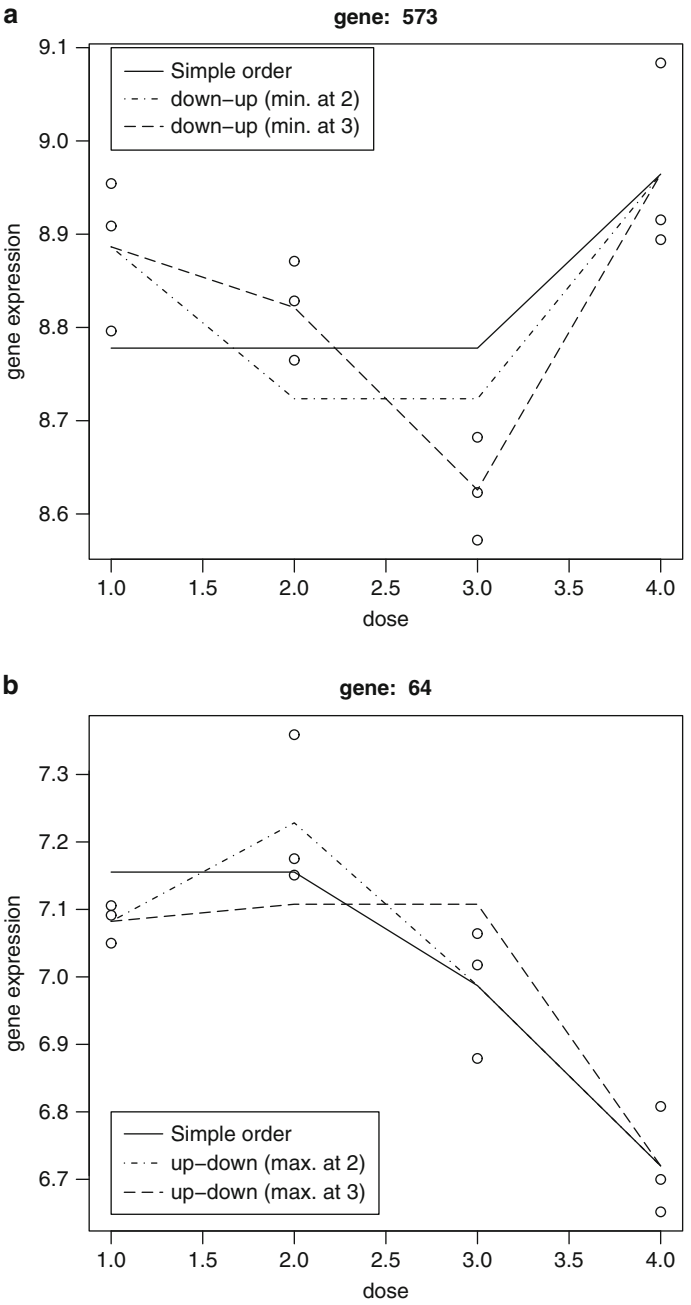

Fig. 11.5 Ordered-restricted profiles for genes 573 (**a**) and 64 (**b**)

```
> #fit.1 : increasing
> #fit.3 : umbrella with up turn at 2
> #fit.4 : umbrella with up turn at 3
> data.frame(my573,fit.1$mu,fit.3$mu,fit.4$mu)
      my573 fit.1.mu fit.3.mu fit.4.mu
1 8.886433 8.777842 8.886433 8.886433
2 8.821386 8.777842 8.723547 8.821386
3 8.625707 8.777842 8.723547 8.625707
4 8.964358 8.964358 8.964358 8.964358
```

11.3.3 Example: Clustering Genes into Order-Restricted Profiles

An analysis using the one-stage ORICC algorithm can be carried out using the function `ORICC1()`.

```
> data1 <- data.frame(paste("g",1:16998,sep=""),data)
> fit.clust<-ORICC1(data1,data.col=2:13,id.col=1,n.rep=rep(3,4),
+         n.top=250,transform=0,name.profile="all",plot.format="eps")
```

The R object `data1` is the data frame containing the gene expression with the first column as the gene names. The `ORIClust` ranked genes based on the variability across the dose levels (Peddada et al. 2003 and Liu et al. 2009b) measured by

$$v_g = \frac{1}{K+1}\sum_{i=0}^{K}\left(\hat{\mu}_i^\star - \bar{\hat{\mu}}\right)^2, \quad \bar{\hat{\mu}} = \frac{\sum_{i=0}^{K}\hat{\mu}_i^\star}{K+1}.$$

The option `n.top=G` implies that only the top G genes will be reported. For the analysis above, for the top 250 genes, 59, 16, 9, 81, 60, and 25 genes were clustered into $C_1, \ldots, C_6$, respectively. Mean profiles of these genes are shown in Fig. 11.6.

```
> table(fit.clust$cluster)

 1  2  3  4  5  6
59 16  9 81 60 25
```

The option `name.profile="all"` implies that all order-restricted profiles will be used as candidate profiles. Alternatively, we can define the candidate profiles of interest as

```
> ## note that profile3 & profile4 define two profiles each
> profile1<-"decreasing"
> profile2<-"increasing"
> profile3<-paste("up down max at",c(2,3),sep=" ")
> profile4<-paste("down up min at",c(2,3),sep=" ")
> name.profileK=c(profile1,profile2,profile3,profile4)
```

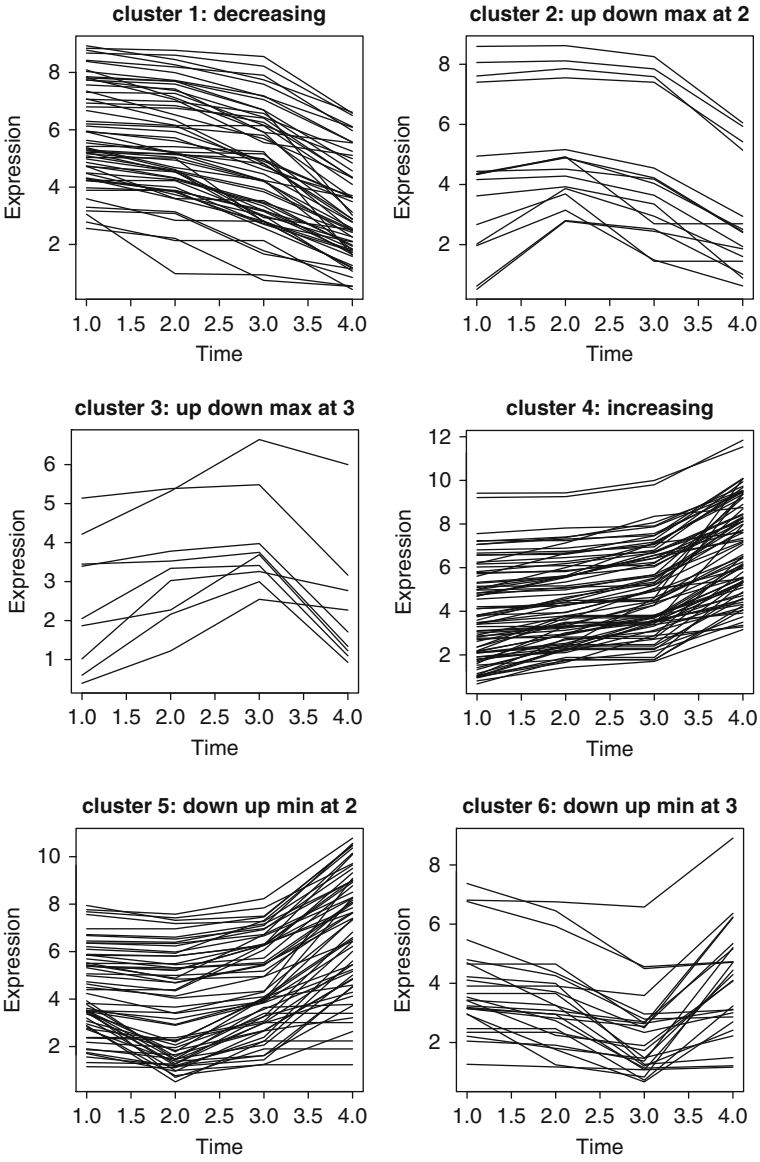

Fig. 11.6 The graphical output of the `ORIClust` for the top 250 genes

The vector `name.profileK` contains the names of all profiles of interest:

```
> name.profileK
[1] "decreasing"       "increasing"       "up down max at 2"
[4] "up down max at 3"  "down up min at 2" "down up min at 2"
```

A similar analysis can be produced with the option `name.profile=name.profileK`. A two-stage ORICC algorithm can be carried out by using the function `ORICC2()`:

```
> fit.clust<-ORICC2(data1,data.col=2:13,id.col=1,n.rep=rep(3,4),
+  n.top=250,transform=0,name.profile=name.profileK,plot.format="eps")
```

Note that only 7,254 genes out of the 16,988 were considered for clustering.

```
[1] "ORICC select 7254 genes"
[1] "retain the top 250 genes"
```

However, it is important to mention that these genes were selected since $\text{ORIC}(C_0) > \text{ORIC}(\text{UR})$. It does not imply that these genes have a significant order-restricted mean profile with respect to dose.

11.4 Discussion

In some experiments, the underlying assumptions of monotone dose-response realtionship and homoscedastic variances may not be appropriate. In this chapter, we relaxed both assumptions and discussed two different algorithms for an analysis of order-restricted gene profiles. The ORIOGEN algorithm, implemented in `ORIOGEN` 3.0, provides both inference and clustering tools. The inference procedure implemented in the `ORIOGEN` package is a resampling-based SAM procedure, and therefore, the FDR can be controlled. The clustering step in `ORIOGEN` is applied only for genes for which a significant dose-response relationship was detected. This is similar to the two-stage clustering approaches discussed in Chaps. 9 and 10 in which an initial filtering based on a significance test is applied to the microarray data in order to filter out genes with nonsignificant dose-response relationship. The ORICC algorithm, implemented in the `ORIClust` package, is a model selection-based algorithm and therefore it provides only a clustering tool but not an inference tool. For that reason, one cannot interpret order-restricted profiles which were selected by the package as "significant" profiles. As mentioned in Sects. 11.2 and 11.3.1, both the `ORIOGEN` 3.0 and the `ORIClust` do not decompose the simple order profiles into their sub-(monotone) profiles. If such an analysis is of primary interest, the researcher can first perform an analysis for all order-restricted profiles and thereafter use the methodology discussed in Chaps. 9 and 10 to further cluster the increasing and decreasing profiles into their sub-profiles.

As we pointed out in Sect. 11.2, the test statistics implemented in the `ORIOGEN` 3.0 are Marcus' t-type SAM test statistic. Therefore, if only simple order alternatives are of interest, one can produce a similar analysis in the `IsoGene` (homoscedastic variances) and the `ORIOGEN` (heteroscedastic variances). Non-monotone alternatives such as the simple tree alternative and the umbrella alternative will be discussed further in Chap. 15.

References

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika, 86*(1), 141–152(12).

Bretz, F., & Hothorn, L. A. (2003). Statistical analysis of monotone and non-monotone dose-response data from in vitro toxicological assays. *ALTA 31*, (Suppl. 1), 81–96, 2003, http://ecvam.jrc.it/publication/Bretz-Hothorn.pdf.

Gou, W. and Peddada, S. (2008), Adaptive choice of the number of bootstrap samples in large scale multiple testing, *Statistical Applications in Genetics and Molecular Biology, 7*(1).

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009a). *Order-restricted information criterion-based clustering algorithm. Reference manual*. http://cran.r-project.org/web/packages/ORIClust/.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009b). Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments. *BMC Bioinformatics, 10*, 146.

Peddada, S., Lobenhofer, E. K., Li, L., Afshari, C. A., Weinberg, C. R., & Umbach, D. M. (2003). Gene selection and clustering for time-course and dose-response microarray experimants using order-restricted inference. *Bioinformatics, 19*(7), 834–841.

Peddada, S., Harris, S., & Harvey E. (2005). ORIOGEN: order restricted inference for ordered gene expression data. *Bioinformatics, 21*(20), 3933–3934.

Peddada, D. S., Umbach, M.D., & Harris, F.S. (2009). A response to information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments. *BMC Bioinformatics, 10*, 438. http://www.biomedcentral.com/1471-2105/10/438

Peddada, S., Harris, S., & Davidov, O. (2010). Analysis of correlated gene expression data on ordered categories. *Journal of Indian Society of Agricaltural Statistics, 64*(1), 45–60.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Simmons, S. J., & Peddada, S. (2007). Order-restricted inference for ordered gene expresion (ORIOGEN) data under heteroscedastic variances. *Bioinformation, 1*(10), 414–419.

Chapter 12
Gene Set Analysis as a Means of Facilitating the Interpretation of Microarray Results

Nandini Raghavan, An De Bondt, Tobias Verbeke, and Dhammika Amaratunga

12.1 Introduction

We have, thus far, been discussing gene-specific methods for identifying statistically significant associations between gene expression profiles and a response variable. These methods involved fitting, for each gene, an isotonic regression model to relate the gene expression levels to the levels of the explanatory variable (dose). An appropriate test statistic was then calculated for each gene and assigned a p value. The intention was to produce a ranked list of genes with small p values and to examine the genes in that list for biological significance.

While it is indeed important to identify individual genes associated with response, interpretation of the resultant findings is facilitated by taking into account the fact that biological phenomena occur through the actions and interactions of multiple genes, via signaling pathways or other functional relationships. It is therefore possible to categorize genes into gene sets according to these relationships. This is enabled by the availability of databases that provide biological annotation for known genes. For example, the Gene Ontology Consortium (2000) has developed a comprehensive taxonomy of gene annotations for three ontologies: biological process, cellular component, and molecular function. Each ontology is structured as a directed acyclic graph, with a hierarchy of terms that vary from broad levels of classification down to more narrowly defined levels.

N. Raghavan (✉) · D. Amaratunga
Janssen Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ, USA
e-mail: nraghava@its.jnj.com; DAMARATU@its.jnj.com

A. De Bondt
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: ADBONDT@its.jnj.com

T. Verbeke
OpenAnalytics BVBA, Heist-op-den-Berg, Belgium
e-mail: tobias.verbeke@openanalytics.be

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_12,

This chapter presents a method that incorporates known biological information available in gene ontologies to identify potentially important gene sets. Gene set analysis often elicits a more comprehensive view on the biology than the gene-specific analysis that we have been discussing so far. In fact, the individual genes in a significant gene set may not even have particularly significant individual p values. However, these genes may act in concert with each other to regulate certain molecular mechanisms and biological pathways and thus play an important role in the cellular event of interest. Thus, when interpreting experimental results, gene set analysis offers an alternative and perhaps more interesting and useful approach, not only to gene selection, but conceivably to the discovery of new mechanisms as well.

The starting point of the method is a list of p values (or similar statistics) that quantify the degree of differential expression for each gene probed by the experiment. The idea is to examine the set of p values associated with a particular gene set (e.g., a particular GO term) to determine whether it exhibits a tendency toward small p values compared to the global set of p values for all the genes. This involves the use of (a) a test statistic to quantify the extent of the difference and (b) a resampling or analytical scheme to judge whether the difference is possibly real or attributable to chance. This process can be applied to all gene sets of interest.

The content of this chapter is organized as follows: Sect. 12.2 describes the method used to score genes in a gene set. Section 12.3 illustrates the use of R package `MLP` for the analysis of the case study data.

12.2 Gene Function Scoring Method

12.2.1 The General Approach

Existing methods for evaluating the significance of gene sets can be classified into two broad classes:

(1) Overrepresentation analysis: this approach thresholds all the p values and labels each gene as being "significant" (e.g., $p < 0.05$) or "not significant" (e.g., $p > 0.05$). The proportion of significant genes in the gene set is then compared to the corresponding proportion for the population of genes being studied using Fisher's exact test (Fisher 1934) or a variant thereof to determine whether there is over-representation of significant genes in the gene set. Analogous methods have been described by a number of authors (see Raghavan et al. 2006 for a list of references) and is the basis for a number of software packages for doing gene set analysis and its sibling, pathway analysis.

(2) Functional class scoring : this approach computes a statistic that summarizes the p values of the genes in a gene set. In particular, we use the statistic

$$\text{MLP} = \sum(-\log(p_i)), \tag{12.1}$$

where p_i refers to the p value of the ith gene in the gene set (Pavlidis et al. 2004, Raghavan et al. 2006). Here, the acronym "MLP" stands for "mean log p." Other summary statistics which have been proposed for functional class scoring include the Kolmogorov–Smirnov statistic which compares the distribution of the p values in a gene set to the distribution of all the p values in the study. A widely used version of this approach is considered by Mootha et al. (2003), who use the term *gene set enrichment analysis* (GSEA).

Of these methods, we will only consider the method based on the MLP statistic as per the arguments presented in Raghavan et al. (2006). If a certain gene set corresponds to a biological process that is implicated with the response, it is likely that many of the genes comprising that gene set will have relatively small p values, so that the value of the MLP statistic for that gene set will be relatively large. Thus, gene sets with relatively large MLP statistics are the gene sets of interest.

12.2.2 Gene Set Analysis Using the MLP Statistic

Once the MLP statistic is calculated for each gene set in the dataset, a permutation procedure can be applied to determine whether or not a particular gene set is significant. Each permutation randomly permutes the p values across the genes and maps them to the respective gene sets. This permutation scheme ensures that the p values within a gene set are random but with the correlation structure among the gene sets being preserved. The latter is important because if multiple gene sets share a gene, that gene will have the same p value in all these gene sets in any given permutation. Preserving the correlation among the gene sets in this way is important as it maintains the interpretability of the results (e.g., across the GO terms along a branch of an ontology).

The significance of a gene set is determined by comparing the observed value of the MLP test statistic for that gene set (labeled as MLP) to the values of the MLP test statistic for that same gene set across multiple random permutations of the p values as described above (labeled as MLP*). If MLP is larger than MLP* in most permutations, the gene set would be declared significant. This means that MLP must exceed a threshold quantile of the MLP*s to be declared significant. This threshold quantile is the empirically determined critical value for that gene set.

For example, suppose that a gene set G5 comprises five genes whose p values are 0.0004, 0.1700, 0.0002, 0.0028, and 0.0011. Four of the genes in G5 are highly significant, and the value of the MLP statistic for G5 is 6.161. Now suppose that we randomly permute all the p values in the entire collection of genes in the experiment and the five genes in G5 are assigned p values of 0.7944, 0.2310, 0.0001, 0.4141, and 0.7395. Now only one gene is significant and the value of the MLP statistic for G5 is 2.418, which is much smaller than the observed value of 6.161. Suppose that over 1,000 such random permutations, the MLP values are 2.418, 3.178, . . ., and

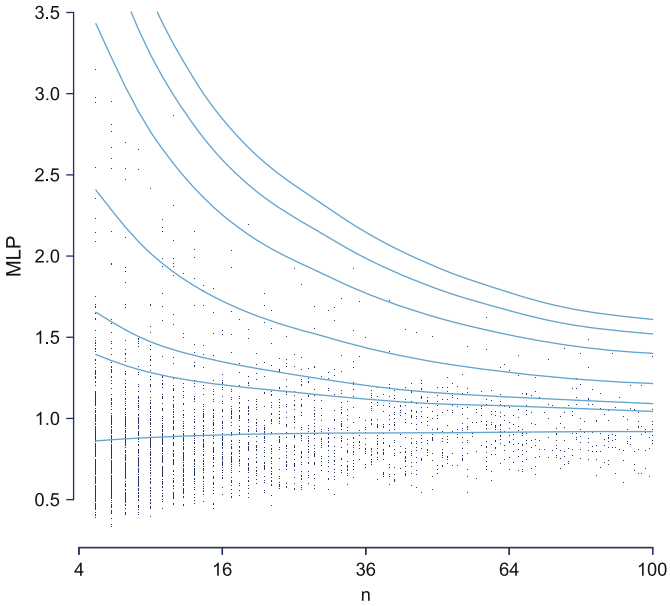

Fig. 12.1 Example of a quantile curves plot for the MLP results showing $geneSetStatistic$ versus $testedGeneSetSize$. *Every dot* represents a gene set. *Every line* represents a smoothing of the null quantile per gene set size. The indicated levels of significance are 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 5×10^{-2}, 10^{-1} and 0.5 (*top* to *bottom*)

that the 95th percentile of these random MLP values is 2.945. This value is the 5% critical value. Since the observed value of 6.161 is larger than this, we would claim that G5 is significant at the 5% level.

The determination of the critical value can be made more efficient by recognizing that the distribution of the MLP test statistic varies systematically as a function of the gene set size as illustrated in Fig. 12.1. In order to determine the appropriate critical value for a gene set of size n, we empirically calibrate the critical values using the following procedure. Randomly generate "null" gene sets of size n, using the permutation scheme described above. For each n, calculate the test statistic MLP_n^*. Run a quantile smoother, $q(\alpha, n)$, through the MLP_n^* versus n relationship so that a proportion α of the observations lie above the smoother (see Amaratunga and Cabrera 2004). The $q(\alpha, n)$ are the α level critical values. The advantage of this procedure is that we borrow strength across all gene sets of the same size, as well as across gene sets of similar sizes. As a result, this procedure will yield a uniform critical value for all gene sets of a given size, and gene sets of similar sizes will have critical values close to each other. More details are provided by Raghavan et al. (2006).

12.3 Implementation of the MLP Package to the Case Study

We now illustrate the MLP method on the case study, i.e. the Affymetrix gene expression dataset of subjects treated with a series of doses of an antipsychotic compound (the harmacological activity of antipsychotics case study). We illustrate this for one of the statistics considered earlier, the $\hat{\tilde{E}}_{01}^2$ statistic. Please note that the MLP package, available from Bioconductor, makes use of various annotation packages also available at the Bioconductor website http://www.bioconductor.org. The package GO.db contains a set of annotation maps of the Gene Ontology, in particular information describing each of the gene sets GOTERM in GO for each of the three ontologies: BP (biological processes), MF (molecular function) and CC (cellular component). It also provides maps for linking each GOTERM to its ancestors and children. The package org.Rn.eg.db provides genome-wide annotation for species *rat*. Among other things, it provides mappings between each GOTERM and the EntrezGene identifiers in that GOTERM. This is a species-specific library and for each experiment, the library corresponding to the species of the subjects used in the experiment, in this case *rat*, must be used. Corresponding species packages for *mouse* and *human* are org.Mm.eg.db and org.Hs.eg.db, respectively.

The first step is to load the MLP package in the R-workspace. This automatically loads various additional required packages, such as GO.db and the appropriate species package (e.g. org.Rn.eg.db).

```
> library(MLP)
```

The next step is to set up the MLP package, as shown below. The inputs that the user needs to specify include the set of p values corresponding to the gene-by-gene analysis, and the annotation information mapping genes to gene sets. The first input to the MLP package is geneStatistic, a named numeric vector of p values where the names correspond to the EntrezGene identifiers. We use the p values from the likelihood ratio test in Chapter 7, where 3,613 genes were found to have a significant monotone dose-response relationship. The p values for first ten genes are shown below.

```
> #e2 is the vector of the two-sided p-values for the LRT from Chapter 7
> pvalues <- e2
> pvalues[1:10]
    112400     113882     113892     113893     113894     113898
0.50169961 0.37989980 0.08577877 0.10567617 0.17950223 0.04372303
113900     113901     113902     113906
0.43165756 0.42154820 0.37882345 0.18401486
```

The second input to the MLP package is geneSet, a mapping of EntrezGene identifiers to the GO terms. This can be generated by the function getGeneSets in the MLP package. The user will need to specify the various parameters, including the *species*; the *geneSetSource* of interest, e.g., "GOBP"; and the entrezIdentifiers for the probe sets used in the experiment. Note that this needs to be a unique list of EntrezGene identifiers.

There are 11,562 unique EntrezGene identifiers in the case study. The output of function getGeneSets gives a list of 7,156 GO terms with a set of corresponding EntrezGene identifiers. This number can change depending on

the version of GO as the information in the GO is regularly updated. Three GOBP terms are shown below.

```
> pathwaySource <- "GOBP"
> geneSets <- getGeneSets(species = "Rat",
+       geneSetSource = pathwaySource,
+       entrezIdentifiers = names(pvalues)
+ )

> geneSets[3:5]
$`GO:0000012`
[1] "24839"  "64573"  "84495"  "259271" "290907" "314380"

$`GO:0000018`
 [1] "24699"  "25660"  "25712"  "59086"  "64012"  "81685"  "81709"  "81816"
 [9] "116562""171369" "287287" "287437" "288905" "290803" "291609" "303496"
[17] "303836""308755" "312398" "317382" "362288" "362412" "362896" "499505"
[25] "690237"

$`GO:0000019`
[1] "64012"  "81685"  "308755"
```

The core of the `MLP` package is the `MLP` function , which is used as shown below. The two main inputs are `geneStatistic` and `geneSet`, obtained above. In addition, the user can specify various other parameters, for which default settings are preset in the package. These include specification of the minimum ($minGenes$) and maximum ($maxGenes$) gene set size to be considered within the analysis, default 5 and 100, respectively. Another parameter is `nPermutations`, which specifies the number of permutations to be run. An additional option, `smoothPValues`, is whether or not to calculate probability values by smoothing. Smoothing calculates smoothed monotonically decreasing probability values by leveraging information across gene sets of similar sizes and is the default setting for the package. The user can also specify the quantiles at which probability values are desired.

```
> mlpOut <- MLP(
+       geneStatistic = pvalues,
+       geneSet = geneSets,
+       minGenes = 5,
+       maxGenes = 100,
+       nPermutations = 50,
+       smoothPValues = TRUE,
+       probabilityVector = c(0.5, 0.9, 0.95, 0.99, 0.999, 0.9999, 0.99999)
+ )
>   write.csv(mlpOut, "summaryMLP.csv")
```

The results of the analysis for the case study are as shown below. The output is a data frame of class "MLP" listing the GO terms, the number of genes corresponding to the GO term, in total ($totalGeneSetSize$) as well as for which a p value has been provided ($testedGeneSetSize$), the value of the `MLP` statistic ($geneSetStatistic$), the p value of the GO term ($geneSetPValue$), and the description of the GO term ($geneSetDescription$). The results are ordered based on the significance of the $geneSetStatistic$. The top six GO terms are shown below, for illustration.

```
> class(mlpOut)
[1] "MLP"         "data.frame"
>   head(mlpOut)
           totalGeneSetSize testedGeneSetSize geneSetStatistic geneSetPValue
GO:0014037               11                10        3.116157 0.0001625596
GO:0051789               91                78        1.554874 0.0003010118
GO:0006986               35                30        2.013559 0.0003787278
GO:0006470              111                79        1.529384 0.0003975008
GO:0051591               63                54        1.715411 0.0004080530
GO:0009612               77                64        1.564498 0.0007417438
                              geneSetDescription
GO:0014037         Schwann cell differentiation
GO:0051789         response to protein stimulus
GO:0006986         response to unfolded protein
GO:0006470 protein amino acid dephosphorylation
GO:0051591                      response to cAMP
GO:0009612      response to mechanical stimulus
```

The table below gives an overview of the significance of the tested GO terms (3,098 passed the *minGenes* and *maxGenes* size criteria) in this analysis. Since the algorithm is permutation based, these numbers can be slightly different.

```
Significance Level        Number of Significant Gene sets
p-Value in (0.00001,0.0001]          1
p-Value in (0.0001,0.001]            7
p-Value in (0.001,0.01]             37
p-Value in (0.01,0.1]              193
p-Value in (0.1,1]                2860
```

The `MLP` package, also provides several plot functions. One of them is the probability quantile curves, with *testedGeneSetSize* on the x-axis and the *geneSetStatistic* on the y-axis, as shown in Fig. 12.1.

```
>   pdf("mlpQuantileCurves.pdf", width = 10, height = 10)
>   plot(mlpOut, type = "quantileCurves")
>   dev.off()
```

Figure 12.2 shows the significance of the 20 most significant gene sets.

```
>   pdf("mlpBarplot.pdf", width = 10, height = 10)
>   op <- par(mar = c(30, 10, 6, 2))
>   plot(mlpOut, type = "barplot")
>   par(op)
>   dev.off()
```

In addition, one can also plot, the top set of significant GO terms according to their structure in the ontology, as shown in Fig. 12.3, with the biggest and least specific GO terms shown at the bottom. The most specific "leaves" of the tree are shown at the top. The ovals are colored by the level of significance, with darker shades indicating more significant GO terms. In that sense, the genes in `GO:0014037` are a subset of the genes in `GO:0007422`.

```
>   pdf("mlpGOgraph.pdf", width = 8, height = 6)
>   op <- par(mar = c(0, 0, 0, 0))
>   plot(mlpOut, type = "GOgraph")
>   par(op)
>   dev.off()
```

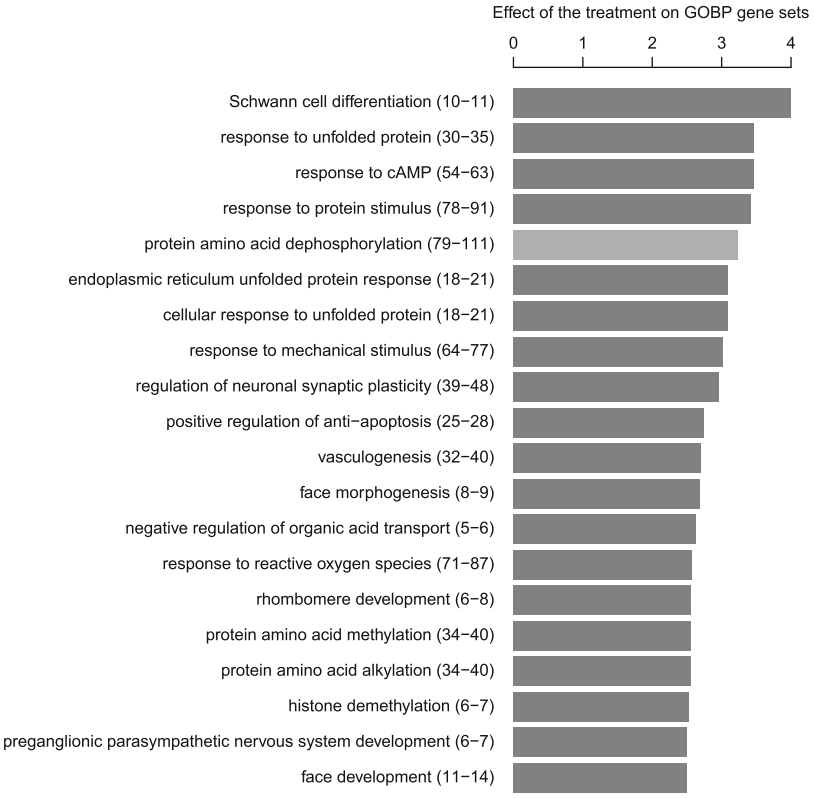

Fig. 12.2 Example of a barplot for the MLP results. The length of a *bar* represents the significance ($-\log_{10}$(`geneSetPValue`)) of the gene set indicated *horizontally*. The *number between brackets* represents the number of genes within that gene set (number of genes for which a gene statistic has been submitted as well as the total number of genes)

The genes contributing most to the significance of a certain gene set can easily be visualized as shown in Fig. 12.4.

```
> geneSetID <- rownames(mlpOut)[1]
> pdf("geneSignificance.pdf", width = 10, height = 10)
> par(mar = c(25, 10, 6, 2))
> plotGeneSetSignificance(
        geneSet = geneSet,
        geneSetIdentifier = geneSetID,
        geneStatistic = pvalues,
        annotationPackage = "rat2302rnentrezg",
)
> dev.off()
```

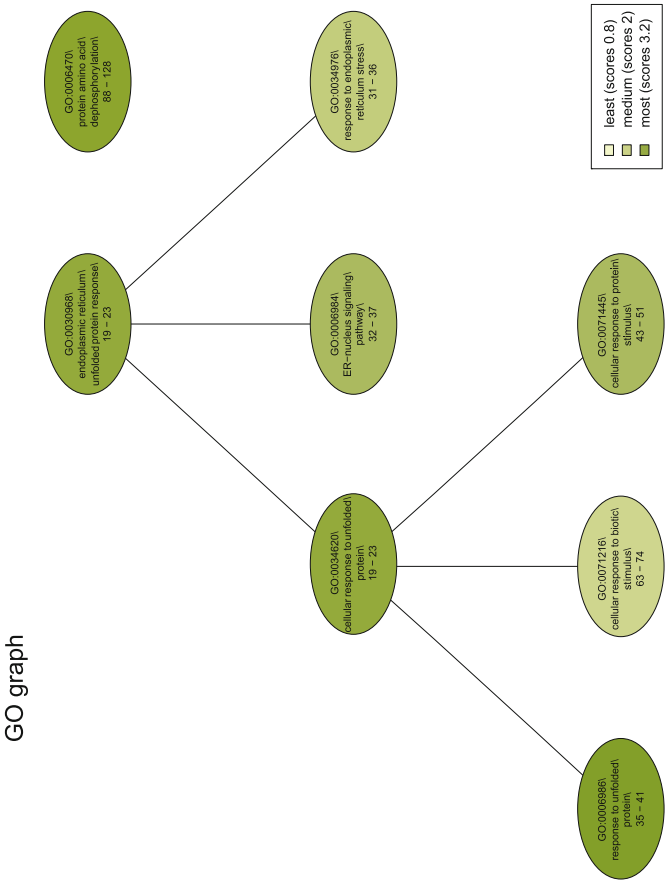

Fig. 12.3 Example of a GOgraph for the MLP results. *Every elipse* represents a gene set. The *color* indicates the significance, the darker, the more significant. The *connectors* indicate the parent–child relationship. The *number at the bottom of the elipses* represent the number of genes within that gene set (number of genes for which a gene statistic has been submitted as well as the total number of genes)

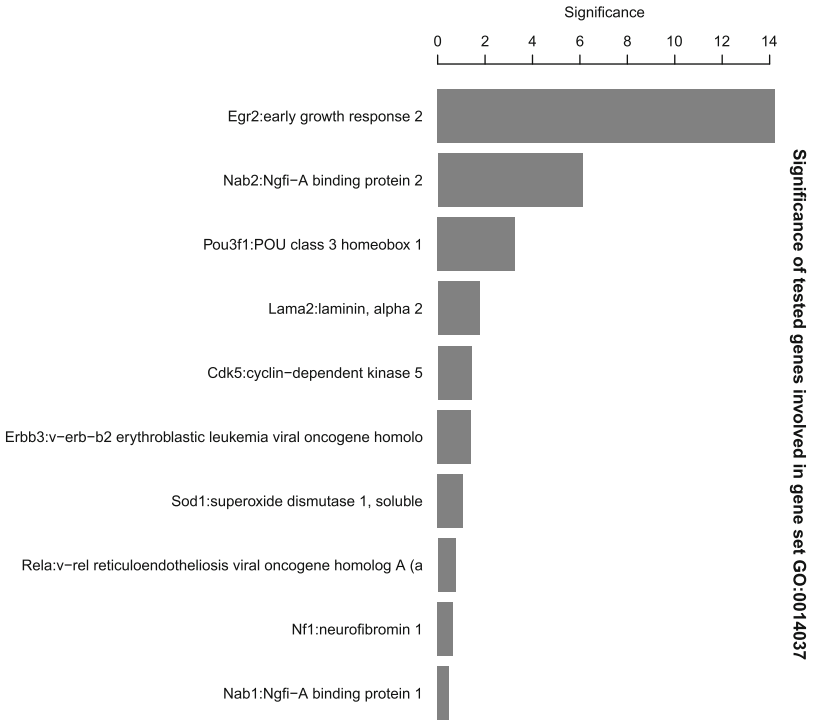

Fig. 12.4 Example of a gene significance plot for a gene set of interest. The length of a *bar* represents the significance ($-\log_{10}$(`geneStatistic`)) of the gene indicated *horizontally*

12.4 Discussion

By its very nature, false-positive findings are prodigious when we interrogate tens of thousands of genes simultaneously and individually. While we can attempt to quantify the degree of such findings using various approaches, it can be difficult, especially in case of a subtle signal, to pinpoint the specific genes that correspond to real findings in an individual gene-by-gene analysis. Approaches based on gene set analysis are an attempt to elicit biologically more meaningful results from high-content genomic experiments. Typically, the output from such an analysis, like the MLP-based analysis above, will need to be interpreted in conjunction with the scientist performing the experiments. This would include follow-up analyses on the specific genes involved in the significant gene sets. However, the user must be cautioned that gene set analysis results are of course dependent on the inputs to the analysis, in this case, the p values from the preceding individual gene-by-gene analysis. Practical experience suggests that a limited number of significant cascading gene sets tend to be more credible than significant gene sets scattered across the spectrum of the hierarchical ontology.

References

Amaratunga, D., & Cabrera, J. (2004). *Exploration and analysis of DNA microarray and protein array data*. Hoboken: Wiley.

Gene Ontology Consortium. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics, 25*, 25–29.

Fisher, R. A. (1934) Statistical Methods for Research Workers. Oxford University Press.

Mootha, V. K. Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics, 34*(3), 267–273.

Pavlidis, P. Qin, J., Arango, V., Mann, J. J., & Sibille, E. (2004). Using the gene ontology for microarray data mining: A comparison of methods and application to age effects in human prefrontal cortex. *Nuerochemical Research 29*, 1213–1222.

Raghavan, N., Amaratunga, D., Cabrera, J., Nie, A., Qin, J., & McMillian, M, (2006). On methods for gene function scoring as a means of facilitating the interpretation of microarray results. *Journal of Computational Biology, 13*(3), 798–809.

Chapter 13
Estimation and Inference Under Simple Order Restrictions: Hierarchical Bayesian Approach

Adetayo Kasim, Ziv Shkedy, and Bernet S. Kato

13.1 Introduction

In Chap. 10, we classified genes into different subsets according to the order-restricted model with the best goodness of fit for each gene. The classification was based on information criteria such as the AIC, BIC, and ORIC. In this chapter, we focus on hierarchical Bayesian modeling of dose-response microarray data. The materials presented in the first part of the chapter are closely related to the classification procedure discussed in Chap. 10 in the sense that we fit several order-restricted models for each gene. However, in contrast with Chap. 10, we do not aim to select the model with the best goodness of fit but to test the null hypothesis of no dose effect. In Sect. 13.2, we formulate order-restricted hierarchical Bayesian model for dose-response data and present gene-specific examples to illustrate the estimation procedures. Within the hierarchical Bayesian framework, one of the major challenges is related to the question of how to perform Bayesian inference and in particular how to adjust for multiplicity. In Sect. 13.3, we discuss the Bayesian variable selection

A. Kasim (✉)
Wolfson Research Institute, Durham University, Durham, UK
e-mail: a.s.kasim@durham.ac.uk

Z. Shkedy
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be

B.S. Kato
Respiratory Epidemiology and Public Health, Imperial College London, London, UK
e-mail: b.kato@imperial.ac.uk

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_13,

(BVS) method (O'Hara and Sillanpää 2009) which we use in order to calculate the posterior probability of a specific model given the data and the model parameters. In Sect. 13.4, following Newton et al. (2004, 2007), we use the posterior probability of the null model to control for multiplicity using the *direct posterior probability* method for multiplicity adjustment. Throughout this chapter, we use the index i for dose, j for replicates within a dose, and r for a candidate model for a specific gene.

13.2 Hierarchical Bayesian Model for Dose–Response Microarray Data

13.2.1 Bayesian Inference: A Short Introduction

Within the Bayesian framework, inference is based on the posterior distribution denoted by $P(\boldsymbol{\mu}|\mathbf{y})$, where $\boldsymbol{\mu}$ is the parameter vector of primary interest and $\mathbf{y}$ is the observed data. The posterior distribution is based on two components:

1. The likelihood function $P(\mathbf{y}|\boldsymbol{\mu})$, which defines the probability distribution of the observed data conditional on the unknown parameters $\boldsymbol{\mu}$
2. The prior distribution $P(\boldsymbol{\mu})$ of the unknown model parameters

The posterior distribution is constructed when the likelihood function and the prior distribution are combined,

$$P(\boldsymbol{\mu}|\mathbf{y}) = \frac{P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu})}{\int P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu})\,\mathrm{d}\boldsymbol{\mu}} \propto P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu}).$$

The integral $\int P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu})\,\mathrm{d}\boldsymbol{\mu}$ is a normalizing constant, and therefore, the posterior distribution is proportional to the product of the likelihood function and the prior distribution. Often, the distribution of $\boldsymbol{\mu}$ depends on unknown hyperparameters $\boldsymbol{\tau}$ for which the hyprerprior distribution is denoted by $P(\boldsymbol{\tau})$. For this case, the hierarchical Bayesian model has three levels:

$$\begin{array}{ll} \text{First level; likelihood} & P(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\tau}), \\ \text{Second level; prior distribution} & P(\boldsymbol{\mu}|\boldsymbol{\tau}), \\ \text{Third level; hyperprior distribution} & P(\boldsymbol{\tau}), \end{array} \tag{13.1}$$

and subsequently, the posterior distribution is given by $P(\boldsymbol{\mu}|\mathbf{y}) \propto P(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\tau})P(\boldsymbol{\mu}|\boldsymbol{\tau})$ $P(\boldsymbol{\tau})$. The posterior distribution can be derived analytically or by drawing a sample from it using Markov Chain Monte Carlo (MCMC) simulation. For detailed discussion about Bayesian inference in general and hierarchical Bayesian modeling in

particular, we refer to Gilks et al. (1996) and Gelman et al. (2004). For an elaborate discussion about order-restricted Bayesian models, we refer to Hoijtink et al. (2008).

13.2.2 Hierarchical Order-Restricted Model

The first challenge in Bayesian analysis of dose-response microarray experiments is the estimation of the model parameters under order restrictions. In contrast with the Bayesian inequality models (Klugkist et al. 2005a,b; Kato 2006; Hoijtink et al. 2008), both the null and alternative models for dose-response microarray experiments require estimation of some of the parameters under equality constraints. This can be achieved if we decompose the alternative hypothesis into the basic profiles as shown in Table 9.1 for an experiment with four dose levels. Note that any two or more parameters with an equality constraint can be replaced by a single unknown parameter. For example, the null model $g_0 : \mu_0 = \mu_1 = \cdots = \mu_{K-1} = \mu_K$ can be modeled as a model with one parameter μ, i.e, $g_0 : \mu_0 = \mu_1 = \cdots = \mu_{K-1} = \mu_K = \mu$. In this case, μ is the overall mean gene expression across all dose levels. Similarly, a monotone profile of the form $g_1 : \mu_0 = \mu_1 = \cdots = \mu_{K-1} < \mu_K$ can be modeled as $g_1 : \mu_{012,\ldots,K-1} < \mu_K$, where $\mu_{012,\ldots,K-1}$ is the mean gene expression across all dose levels except the maximum dose. This approach requires that all the possible monotone models are considered per gene.

Similar to the previous chapters, we formulate a gene-specific linear model as

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2),\ i = 0, \ldots, K,\ j = 0, 1, 2, \ldots, n_i, \qquad (13.2)$$

where Y_{ij} is the gene expression at the ith dose level for array j and μ_i is the mean gene expression level at dose level i. Note that since we assume that the dose-response relationship is monotone, we wish to estimate a model in which $\mu_0 \leq \mu_1 \leq \mu_2 \leq \cdots \leq \mu_K$ for a monotone upward profile and $\mu_0 \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$ for a monotone downward profile, where μ_0 is the mean gene expression for the control dose. The monotone constraints are achieved by constraining the parameter space of $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_K)$, whereby the order restrictions are imposed on the prior distributions. For a monotone upward profile, we assume that $\boldsymbol{\mu}$ is a right-continuous nondecreasing function defined on $[0, K]$. We do not assume any deterministic relationship between μ_i and the dose levels, but instead, we specify a probabilistic model for μ_i at each distinct dose level.

The problem is to estimate $\boldsymbol{\mu}$ under the order restrictions, $\mu_0 \leq \mu_1 \leq \cdots \leq \mu_K$. Thus, the $K + 1$ dimensional parameter vector is constrained to lie in a subset S^{K+1} of R^{K+1}. The constrained set S^{K+1} is determined by the order among the components of $\boldsymbol{\mu}$. In this case, it is natural to incorporate the constraints into the specification of the prior distribution (Klugkist and Mulder 2008). Let $\mathbf{y} = (Y_{11}, Y_{12}, \ldots, Y_{Kn_k})$ be the expression levels for a specific gene and $\boldsymbol{\eta}$ and $\boldsymbol{\tau}$ the hyperparameters for $\boldsymbol{\mu}$ which will be discussed later. Gelfand, Smith, and

Lee (1992) showed that the posterior distribution of $\boldsymbol{\mu}$, given the constraints, is the unconstrained posterior distribution normalized such that

$$P(\boldsymbol{\mu}|\mathbf{y}) \propto \frac{P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu}|\boldsymbol{\eta},\boldsymbol{\tau})}{\int_{S^k} P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu}|\boldsymbol{\eta},\boldsymbol{\tau})d\boldsymbol{\mu}}, \quad \boldsymbol{\mu} \in S^{K+1}. \tag{13.3}$$

Let $S_l^{K+1}(\mu_l, l \neq i)$ be a cross section of S^{K+1} defined by the constraints for the component μ_i at a specified set of μ_i, $l \neq i$, and $l = 0, 1, 2, \ldots, K$. In our setting, $S_l^{K+1}(\mu_l, l \neq i)$ is in the interval $[\mu_{i-1}, \mu_{i+1}]$. It follows from (13.3) that the posterior distribution for μ_i is given by

$$\begin{cases} P(\mu_i|\mathbf{y},\boldsymbol{\eta},\boldsymbol{\tau},\boldsymbol{\mu}_{-i}) \propto P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu}|\boldsymbol{\eta},\boldsymbol{\tau}), & \mu_i \in S_l^{K+1}(\mu_l, l \neq i), \\ 0, & \mu_i \notin S_l^{K+1}(\mu_l, l \neq i). \end{cases} \tag{13.4}$$

Here, $\boldsymbol{\mu}_{-i} = (\mu_0, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_K)$. Hence, when the likelihood and the prior distribution are combined, the posterior conditional distribution of $\mu_i|\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\mu}_{-i}$ is the standard posterior distribution restricted to $S_l^{K+1}(\mu_l, l \neq i)$, i.e., restricted to the interval $[\mu_{i-1}, \mu_{i+1}]$ (Gelfand, Smith, and Lee 1992). This means that during the MCMC simulations, the sampling from the full conditional distribution can be reduced to interval restricted sampling from the standard posterior distribution. For an elaborate discussion of constrained analysis of variance, we refer to Sect. 3.2.4 in Klugkist and Mulder (2008).

The hierarchical model we consider is given by

$$\begin{array}{ll} Y_{ij} \sim \mathrm{N}(\mu_i, \tau_1^2) & \text{likelihood} \\ \mu_i \sim \mathrm{N}(\eta_{\mu_i}, \tau_{\mu_i}^2) I(\mu_{i-1}, \mu_{i+1}) & \text{prior,} \end{array} \tag{13.5}$$

where $I(\mu_{i-1}, \mu_{i+1})$ is an indicator variable, which takes the value of 1 if $\mu_{i-1} \leq \mu_i \leq \mu_{i+1}$ and zero elsewhere. In order to complete the specification of the hierarchical model in (13.5), we need to specify hyperprior distributions for η_{μ_i}, $\tau_{\mu_i}^{-2}$ and τ_1^{-2}. We assume $K + 1$ hyperprior distributions at the third level of the model, $\eta_{\mu_i} \sim N(0, 10,000)$ and $\tau_{\mu_i}^{-2}, \tau_1^{-2} \sim \text{gamma}(0.001, 0.001)$.

Alternatively, the order constraints can be incorporated into the model using the following parameterization for the mean structure (for an increasing trend):

$$E(Y_{ij}) = \mu_i = \begin{cases} \mu_0, & i = 0, \\ \mu_0 + \sum_{\ell=1}^{i} \delta_\ell, & i = 1, \ldots, K. \end{cases} \tag{13.6}$$

In matrix notation, the mean gene expression can be expressed as $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is a direction-dependent known design matrix, i.e., for an upward and downward trend the design matrices are given, respectively, by

$$\mathbf{X}_{\text{up}} = \begin{pmatrix} 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,1 \\ 1\,1\,1\,1 \\ 1\,1\,1\,1 \end{pmatrix} \quad \text{and} \quad \mathbf{X}_{\text{dn}} = \begin{pmatrix} 1\,1\,1\,1 \\ 1\,1\,1\,1 \\ 1\,1\,1\,1 \\ 1\,0\,1\,1 \\ 1\,0\,1\,1 \\ 1\,0\,1\,1 \\ 1\,0\,0\,1 \\ 1\,0\,0\,1 \\ 1\,0\,0\,1 \\ 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 1\,0\,0\,0 \end{pmatrix},$$

and $\boldsymbol{\beta}$ is a direction-dependent parameter vector, in particular $\boldsymbol{\beta}'_{\text{up}} = (\mu_0, \delta_1, \ldots, \delta_K)$ and $\boldsymbol{\beta}'_{\text{dn}} = (\mu_K, \delta_0, \ldots, \delta_{K-1})$. The constraints on the parameter space can be incorporated in the hierarchical model by assuming truncated normal distributions as the prior for the components of $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_K)$,

$$\delta_\ell \sim \text{truncated } N(\eta_{\delta_\ell}, \tau^2_{\delta\ell}), \quad \ell = 1, 2, \ldots, K. \tag{13.7}$$

Here, the normal distribution for the priors of δ_ℓ is left truncated at 0 to ensure that $\delta_\ell \geq 0$. To complete the formulation of the model, we assume vague hyperprior distributions at the third level of the model, i.e., $\eta_{\delta_\ell} \sim N(0, 10,000)$ and $\tau^{-2}_{\delta_\ell} \sim$ gamma$(0.001, 0.001)$.

As pointed out by Dunson and Neelon (2003), since the prior of the components of $\boldsymbol{\delta}$ is truncated normal distribution, the mean structure $\mu_i = \mu_0 + \sum \delta_\ell$ implies an order constraints mean structure with strict inequalities $\mu_0 < \mu_1 <, \ldots, < \mu_K$. Equality constraints can be incorporated in the model by setting some of the components in $\boldsymbol{\delta}$ to be equal to zero. Indeed, $\delta_i = 0$ implies $\mu_i = \mu_{i-1}$.

13.2.2.1 Example 1: Unrestricted and Order-Constrained Models

For the analyses presented in this chapter, we use human epidermal squamous carcinoma cell line A431 experiment as a case study. The expression matrix consists of 12 arrays and 16,988 genes. We focus on a single gene (1,095) as an example to illustrate the difference between the posterior estimates of the constrained and unconstrained models. The likelihood and the parameterization of the mean structure for the two models are identical and given in (13.5) and (13.6). For the unconstrained model, vague priors are considered for parameter δ_i. Note that for this model, δ_i is not restricted to be nonnegative. In Winbugs 1.4, the likelihood and the linear predictor for the mean can be implemented by

```
model
{
for(i in 1:N)
{
Y[i] ~dnorm(mu[i],tau)
mu[i] <- mu0*X1[i]+ delta1*X2[i]+ delta2*X3[i]+ delta3*X4[i]
}
```

Next, we use the truncation function in Winbugs `I()` for the constrained parameters. For the unconstrained model, the same priors are used without truncation function `I()`.

```
mu0 ~dnorm(mu.mean,mu.tau)I(0.00,)
delta1 ~dnorm(delta1.mean,delta1.tau)I(0.00,)
delta2 ~dnorm(delta2.mean,delta2.tau)I(0.00,)
delta3 ~dnorm(delta3.mean,delta3.tau)I(0.00,)
```

As mentioned above, we use independent flat hyperprior distributions for all hyperparameters in the model.

```
mu.mean ~dnorm(0,0.000001)
mu.tau~ dgamma(0.001,0.001)
delta1.mean ~dnorm(0,0.000001)
delta1.tau~ dgamma(0.001,0.001)
delta2.mean ~dnorm(0,0.000001)
delta2.tau~ dgamma(0.001,0.001)
delta3.mean ~dnorm(0,0.000001)
delta3.tau~ dgamma(0.001,0.001)
tau~dgamma(0.001,0.001)
sigma <-1/ sqrt(tau)
```

We define two lists, one for the gene expression data and the other for the initial values for the model.

```
#DATA (for gene 3310)
list(a=1.5,N=12,Y=c(6.662642,6.548664,6.893789,6.568433,6.483458
    ,6.340048,6.814083,6.748551,6.803282,7.193578,7.340334,7.047279),
   X1=c(1,1,1,1,1,1,1,1,1,1,1,1),
   X2=c(0,0,0,1,1,1,1,1,1,1,1,1),
   X3=c(0,0,0,0,0,0,1,1,1,1,1,1),
   X4=c(0,0,0,0,0,0,0,0,0,1,1,1))

#INITIAL VALUES
list(mu0=0.5,delta1=0.05,delta2=0.05,delta3=0.05,tau=1,mu.mean=0.1,mu.tau=1,
     delta1.tau=1,delta2.tau=1,delta3.tau=1,delta1.mean=0.05,
     delta2.mean=0.05,delta3.mean=0.05)
```

Note that the design matrix $\mathbf{X}_{\text{up}}$ in (13.2.2) is defined by the variables `X1`, `X2`, `X3`, and `X4`.

```
> X1=c(1,1,1,1,1,1,1,1,1,1,1,1)
> X2=c(0,0,0,1,1,1,1,1,1,1,1,1)
> X3=c(0,0,0,0,0,0,1,1,1,1,1,1)
> X4=c(0,0,0,0,0,0,0,0,0,1,1,1)
> data.frame(X1,X2,X3,X4)
   X1 X2 X3 X4
1   1  0  0  0
2   1  0  0  0
3   1  0  0  0
4   1  1  0  0
5   1  1  0  0
6   1  1  0  0
7   1  1  1  0
8   1  1  1  0
9   1  1  1  0
10  1  1  1  1
11  1  1  1  1
12  1  1  1  1
```

The hierarchical model was fitted in Winbugs 1.4. We used a chain of 30,000 iterations from which the first 10,000 were treated as burn in period and discarded from the analysis. Figure 13.1a shows the observed data, isotonic means, and the parameter estimates for the posterior means obtained from the constrained and unconstrained models. The parameter estimates for the posterior means obtained from the unconstrained reveal a violation in the order at the second dose level ($\bar{\mu}_0 = 9.672$ and $\bar{\mu}_1 = 9.608$, see the panel below). Indeed, we notice that the pool adjacent violators algorithm (PAVA), discussed in Chap. 2, pools together the means of the first two dose levels. Similar to isotonic regression, the parameter estimates for posterior means for the constrained model are monotone. However, in contrast with isotonic regression, the parameter estimates for the posterior means obtained from the constrained model are equal to 9.625 and 9.655 for the first two dose levels, respectively (compared with the isotonic mean $\hat{\mu}^{\star}_{12} = 9.64$).

```
## EDITED OUTPUT FOR GENE 1095 ###
## isotonic regression ##
## mu[1]: posterior mean in dose 0, mu[4]: mu[1]: posterior mean in dose 1,
   mu[7]: posterior mean in dose 2, mu[10]: posterior mean in dose 3

> whichgene<-1095
> Y<-as.numeric(as.vector(dat[whichgene,]))
> #mY: mean gene expression at each dose
> mY<-tapply(Y,as.factor(dose),mean)
> iso.r1<-pava(mY,w=c(3,3,3,3))
> iso.r1
        1         2         3         4
 9.640501  9.640501  9.842729 10.001414



## edited output from WINBUGS for the unconstrained model
node       mean    sd      MC error    2.5%    median  97.5%   start   sample
mu[1]     9.672   0.03807 5.453E-4     9.595   9.672    9.747   20000   30001
mu[4]     9.608   0.03809 4.567E-4     9.532   9.608    9.684   20000   30001
mu[7]     9.843   0.03742 3.869E-4     9.77    9.843    9.918   20000   30001
mu[10]   10.0     0.03806 2.513E-4     9.925  10.0      10.08    20000   30001

## edited output from WINBUGS for the constrained  model
node       mean    sd      MC error    2.5%    median  97.5%   start   sample
mu[1]     9.625 0.03439 3.774E-4     9.55    9.626   9.688   20000   30001
```

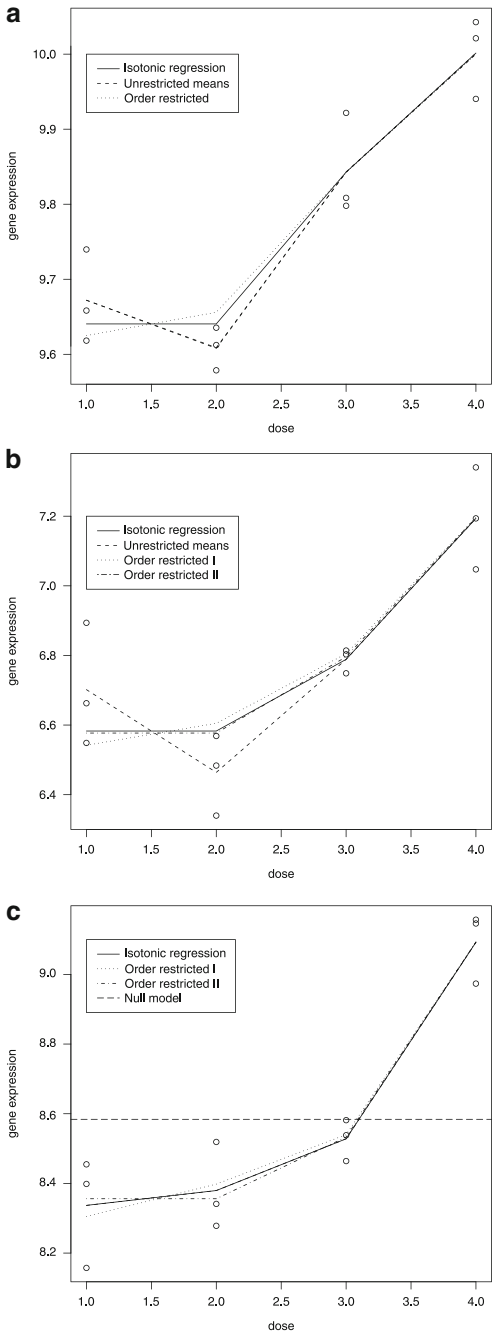

Fig. 13.1 Point estimate for the posterior means and isotonic regression. (**a**) Gene 1095: isotonic regression and posterior means obtained from the unrestricted and the restricted model with four parameters. (**b**) Gene 3310: isotonic regression and posterior means obtained from the unrestricted model, and from the constrained model with four parameters (order restricted I) and from a constrained model with three parameters (order restricted II). (**c**) Gene 13386: the null model, isotonic regression, and posterior means obtained from order-restricted model with three parameters (order restricted II) and order-restricted model with four parameters (order restricted I)

```
mu[4]      9.655 0.03388 3.608E-4    9.592   9.653    9.728    20000    30001
mu[7]      9.843 0.04264 4.402E-4    9.757   9.843    9.928    20000    30001
mu[10]    10.0   0.04284 2.522E-4    9.919  10.0     10.09     20000    30001
```

13.2.2.2 Example 2: Isotonic Regression and the Order-Constrained Bayesian Model

As mentioned in Chap. 2, isotonic regression reproduces the means of the observed data whenever the unrestricted means are monotone. This is not necessarily the case for the order-restricted Bayesian model which depends on the specific constrained model that we specify. Figure 13.1b shows the expression data for gene 3310 for which we consider two different constrained Bayesian models with mean structure given by $g_7 : (\mu_0 < \mu_1 < \mu_2 < \mu_3)$ and $g_5 : (\mu_0 = \mu_1 < \mu_2 < \mu_3)$, respectively. The models can be parameterized as

Model 1 (g_7) :		Model 2 (g_5) :	
$\mu_0 = \mu_0$	control,	$\mu_0 = \mu_0$	control,
$\mu_1 = \mu_0 + \delta_1$	dose level 1,	$\mu_1 = \mu_0$	dose level 1,
$\mu_2 = \mu_0 + \delta_1 + \delta_2$	dose level 2,	$\mu_2 = \mu_0 + \delta_2$	dose level 2,
$\mu_3 = \mu_0 + \delta_1 + \delta_2 + \delta_3$	dose level 3.	$\mu_3 = \mu_0 + \delta_2 + \delta_3$	dose level 3.

The first model consists of four constrained parameters, while the latter consists of three constrained parameters, and the means of dose levels 0 and 1 are constrained to be equal (similar to the pooling of the PAVA). Note that for model 2 $\delta_1 = 0$ so $\mu_0 = \mu_1$. The output of the two models are given below.

```
## OUTPUT FOR GENE 3110 ###

## Isotonic regression ##
> whichgene<-3110
> Y<-as.numeric(as.vector(dat[whichgene,]))
> mY<-tapply(Y,as.factor(dose),mean)
> iso.r1<-pava(mY,w=c(3,3,3,3))
> iso.r1
       1        2        3        4
6.582839 6.582839 6.788639 7.193730


## edited output from WINBUGS for model 1(g7, 4 levels)
node        mean    sd     MC error    2.5%    median  97.5%    start    sample
mu[1]      6.542 0.08516 9.363E-4    6.355    6.547    6.692    20000    30001
mu[4]      6.605 0.07953 7.425E-4    6.449    6.604    6.765    20000    30001
mu[7]      6.805 0.09824 9.086E-4    6.626    6.8      7.012    20000    30001
mu[10]     7.196 0.1095  6.943E-4    6.984    7.196    7.416    20000    30001


## edited output from WINBUGS for model 2(g5, 3 levels)
node      mean      sd     MC error    2.5%    median  97.5%    start    sample
mu[1]    6.577    0.06958 7.474E-4    6.434    6.579    6.712    20000    30001
```

```
mu[4]   6.577   0.06958 7.474E-4    6.434   6.579    6.712   20000   30001
mu[7]   6.798   0.09224 9.003E-4    6.624   6.796    6.99    20000   30001
mu[10]  7.194   0.09876 6.626E-4    6.999   7.193    7.394   20000   30001
```

For the second model, we need to redefine the design matrix; this can be done by excluding the second column in the design matrix in the data list in the following way:

```
# data for model 1 (g_7: 4 levels)
list(a=1.5,N=12,Y=c(6.662642,6.548664,6.893789,6.568433,6.483458,
                    6.340048,6.814083,6.748551,6.803282,7.193578,
                    7.340334 ,7.047279),
   X1=c(1,1,1,1,1,1,1,1,1,1,1,1),
   X2=c(0,0,0,1,1,1,1,1,1,1,1,1),
   X3=c(0,0,0,0,0,0,1,1,1,1,1,1),
   X4=c(0,0,0,0,0,0,0,0,0,1,1,1))



# data for model 2 (g_5: 3 levels)
list(a=1.5,N=12,Y=c(6.662642,6.548664,6.893789,6.568433,6.483458,
                    6.340048,6.814083,6.748551,6.803282,7.193578,
                    7.340334 ,7.047279),
   X1=c(1,1,1,1,1,1,1,1,1,1,1,1),
   X3=c(0,0,0,0,0,0,1,1,1,1,1,1),
   X4=c(0,0,0,0,0,0,0,0,0,1,1,1))
```

13.2.2.3 Example 3: Model Selection

Within the hierarchical Bayesian framework, the goodness of fit and complexity for the gene-specific models can be assessed using the *deviance information criterion* (DIC), proposed by Spiegelhalter et al. (2002). The effective number of parameters (the complexity) in the hierarchical model can be measured by the difference between the posterior expectation of the deviance and the deviance evaluated at the posterior expectation of $\boldsymbol{\mu}$ (Spiegelhalter et al. 2002; Hoijtink et al. 2008), that is,

$$P_D = E_{\boldsymbol{\mu}|\mathbf{y}}(D) - D(E_{\boldsymbol{\mu}|\mathbf{y}}(\boldsymbol{\mu})) = \bar{D} - D(\bar{\boldsymbol{\mu}}).$$

Here, deviance is given by $D(\boldsymbol{\mu}) = -2 \log P(\mathbf{y}|\boldsymbol{\mu}) + 2 \log(f(\mathbf{y}))$. The second term in the deviance is a standardizing factor which does not depend on $\boldsymbol{\mu}$. The deviance information criterion for model selection is given by

$$\text{DIC} = \bar{D} + P_D = D(\bar{\boldsymbol{\mu}}) + 2P_D.$$

A small value of DIC indicates a better goodness of fit. The use of the DIC in the context of order-restricted Bayesian model is discussed in Myung et al. (2008) and Chen and Kim (2008). Three models were fitted to the data, g_0, g_5, and g_7. Figure 13.1c shows the data and posterior means for gene 13386. The panel below

shows that g_5 has the smallest DIC value (-13.074) and therefore should be prefered.

```
### OUTPUT FOR GENE 13386 ###

#DIC PANEL FOR THE NULL MODEL (g_0)
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
          Dbar      Dhat   pD       DIC
Y         8.616    6.489   2.128   10.744
total     8.616    6.489   2.128   10.744


#DIC PANEL FOR THE MODEL 1 (g_7: 4 levels)
          Dbar      Dhat       pD      DIC
Y       -16.690   -21.264    4.574   -12.117
delta1    7.243     0.000    7.243    14.485
delta2    7.082     0.143    6.938    14.020
delta3    6.838     0.000    6.838    13.677
mu0       0.860     0.000    0.860     1.721
total     5.332   -21.121   26.453    31.786


#DIC PANEL FOR THE MODEL 2 (g_5: 3 levels)
         Dbar      Dhat     pD      DIC
Y      -17.313  -21.552   4.239  -13.074
delta2   7.166   11.871  -4.705    2.461
delta3   7.049   63.439 -56.390  -49.342
mu0      0.857    0.000   0.857    1.714
total   -2.241   53.759 -56.000  -58.241
```

13.3 Dose–Response Modeling Using Bayesian Variable Selection Methods

13.3.1 Introduction

In the previous section, we discussed the hierarchical Bayesian approach for dose-response microarray experiments with monotone constraints. If our goal is to select the model with the best goodness of fit for each gene, we can fit all possible order-constrained models and select the one with the smallest DIC (Spiegelhalter et al. 2002) among the fitted models, similar to the model selection procedure discussed in Chap. 10. However, this approach is computationally intensive and may become impractical for a large number of dose levels, since the number of all possible monotone models increases rapidly. In this section, in order to overcome the need to fit all possible monotone models separately, we focus on BVS methods (George and McCulloch 1993; O'Hara and Sillanpää 2009) which allow us to estimate the mean gene expression under order restrictions taking into account all possible monotone models. For an application of BVS models within the framework of dose-response modelling we refer to Whitney and Ryan (2009). Let us consider a dose-response experiment with a control and 3 dose levels. Together with the null model, $Y_{ij} = \mu_0 + \varepsilon_{ij}$, there are eight possible models that can be fitted for which the mean structures are given by

$$
\begin{aligned}
g_0 &: \ \mu_0 = \mu_1 = \mu_2 = \mu_3,\\
g_1 &: \ \mu_0 = \mu_1 = \mu_2 < \mu_3,\\
g_2 &: \ \mu_0 = \mu_1 < \mu_2 = \mu_3,\\
g_3 &: \ \mu_0 < \mu_1 = \mu_2 = \mu_3,\\
g_4 &: \ \mu_0 < \mu_1 = \mu_2 < \mu_3,\\
g_5 &: \ \mu_0 = \mu_1 < \mu_2 < \mu_3,\\
g_6 &: \ \mu_0 < \mu_1 < \mu_2 = \mu_3,\\
g_7 &: \ \mu_0 < \mu_1 < \mu_2 < \mu_3.
\end{aligned}
$$

As we mentioned in the previous section, if we parameterize the mean structure at each dose level as $\mu_i = \mu_0 + \sum \delta_\ell$, the difference in the mean structures of the different models depends on which components in $\boldsymbol{\delta}$ are set to be equal to zero or equivalently which column in the design matrix $\mathbf{X}$ is excluded. The design matrices for all models are given, respectively, by

$$
\mathbf{X}_{(g_0)} = \begin{pmatrix} 1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1 \end{pmatrix}, \quad \mathbf{X}_{(g_1)} = \begin{pmatrix} 1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 1\\1\ 1\\1\ 1 \end{pmatrix}, \quad \mathbf{X}_{(g_2)} = \begin{pmatrix} 1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 0\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1 \end{pmatrix}, \quad \mathbf{X}_{(g_3)} = \begin{pmatrix} 1\ 0\\1\ 0\\1\ 0\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1\\1\ 1 \end{pmatrix},
$$

$$
\mathbf{X}_{(g_4)} = \begin{pmatrix} 1\ 0\ 0\\1\ 0\ 0\\1\ 0\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 1\\1\ 1\ 1\\1\ 1\ 1 \end{pmatrix}, \quad \mathbf{X}_{(g_5)} = \begin{pmatrix} 1\ 0\ 0\\1\ 0\ 0\\1\ 0\ 0\\1\ 0\ 0\\1\ 0\ 0\\1\ 0\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 1\\1\ 1\ 1\\1\ 1\ 1 \end{pmatrix}, \quad \mathbf{X}_{(g_6)} = \begin{pmatrix} 1\ 0\ 0\\1\ 0\ 0\\1\ 0\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 0\\1\ 1\ 1\\1\ 1\ 1\\1\ 1\ 1\\1\ 1\ 1\\1\ 1\ 1\\1\ 1\ 1 \end{pmatrix}, \quad \mathbf{X}_{(g_7)} = \begin{pmatrix} 1\ 0\ 0\ 0\\1\ 0\ 0\ 0\\1\ 0\ 0\ 0\\1\ 1\ 0\ 0\\1\ 1\ 0\ 0\\1\ 1\ 0\ 0\\1\ 1\ 1\ 0\\1\ 1\ 1\ 0\\1\ 1\ 1\ 0\\1\ 1\ 1\ 1\\1\ 1\ 1\ 1\\1\ 1\ 1\ 1 \end{pmatrix}.
$$

The mean gene expression, for each model, is given, respectively, by

$$E(Y_{ij}|g_r) = \mathbf{X}_{g_r}\boldsymbol{\beta}_r', \quad r = 1, \ldots, R,$$

where $\boldsymbol{\beta}_r$ is the parameter vector for each model given by

$$\boldsymbol{\beta}_r' = \begin{cases} \mu_0, & \text{model } g_0, \\ (\mu_0, \delta_3)', & \text{model } g_1, \\ (\mu_0, \delta_2)', & \text{model } g_2, \\ (\mu_0, \delta_1)', & \text{model } g_3, \\ (\mu_0, \delta_1, \delta_3)', & \text{model } g_4, \\ (\mu_0, \delta_2, \delta_3)', & \text{model } g_5, \\ (\mu_0, \delta_1, \delta_2)', & \text{model } g_6, \\ (\mu_0, \delta_1, \delta_2, \delta_3)' & \text{model } g_7. \end{cases}$$

For example, the mean gene expression for the three models fitted in Sect. 13.2.2.3 is given by $\mathbf{X}_{g_0}\boldsymbol{\mu}_0$, $\mathbf{X}_{g_7}\boldsymbol{\beta}_7'$, and $\mathbf{X}_{g_5}\boldsymbol{\beta}_5'$, respectively. In fact, all the design matrices above are submatrices of $\mathbf{X}_{g_7}$, and our aim is to select, for each gene, the most appropriate design matrix and parameter vector.

13.3.2 Bayesian Variable Selection

The BVS approach (George and McCulloch 1993) is related to the problem of the choice of an optimal model from a priori set of R known plausible models. As pointed out by O'Hara and Sillanpää (2009), the choice of an optimal model reduces to the choice of a subset of variables which will be included in the model (i.e., model selection) or the choice of which parameters in the parameter vector are different from zero (i.e., inference).

In our setting, the BVS model allows us to calculate the posterior probability of each model, $p(g_r|\text{data})$, and in particular the posterior probability of the null model, $p(g_0|\text{data})$. Let z_i, $i = 1, \ldots, K$ be an indicator variable such that

$$z_i = \begin{cases} 1, & \delta_i \text{ is included in the model,} \\ 0, & \delta_i \text{ is not included in the model,} \end{cases}$$

and let $\theta_i = \delta_i \times z_i$, $\boldsymbol{\beta} = (\mu_0, \theta_1, \theta_2, \theta_3)$ and $\mathbf{y}$ the gene expression vector. Hence, we can reformulate the mean structure in (13.6) (O'Hara and Sillanpää 2009) in terms of θ_i and z_i as

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}' = \begin{pmatrix} 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,0 \\ 1\,1\,1\,1 \\ 1\,1\,1\,1 \\ 1\,1\,1\,1 \end{pmatrix} \begin{pmatrix} \mu_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \quad \text{or} \quad E(Y_{ij}) = \mu_0 + \sum_{\ell=1}^{K} \theta_\ell = \mu_0 + \sum_{\ell=1}^{K} z_\ell \delta_\ell.$$

For the three dose level experiment discussed above, the triplet $\mathbf{z} = (z_1, z_2, z_3)$ defines uniquely each one of the eight plausible models. For example, for $\mathbf{z} = (z_1 = 0, z_2 = 0, z_3 = 0)$, we have that $E(Y_{ij}|g_r, \mathbf{z}) = (\mu_0, \mu_0, \mu_0, \mu_0)$ (which corresponds to the mean of model g_0), and for $\mathbf{z} = (z_1 = 1, z_2 = 0, z_3 = 0)$, we obtain $E(Y_{ij}|g_r, \mathbf{z}) = (\mu_0, \mu_0 + \delta_1, \mu_0 + \delta_1, \mu_0 + \delta_1)$ (which corresponds to the mean of model g_3) etc. In order to complete the specification of the hierarchical model defined in (13.5)–(13.7) we need to specify prior and hyperprior distributions for the indicator variables. We assume that z_i and δ_i are independent. As before, we use truncated normal prior distribution (13.7) for δ_i and

$$\begin{aligned} z_i &\sim \text{Bernouli}(\pi_i), \\ \pi_i &\sim U(0, 1). \end{aligned}$$

As pointed out by O'Hara and Sillanpää (2009), the posterior inclusion probability of δ_i in the model is the posterior mean of z_i. Further, using the indicator variable z_i, we specify a transformation function that uniquely defines each one of the plausible models (Ntzoufras 2002 and Dellaportas 2002). Let $\mathbf{C} = (1, 2, 4)$ and let $\mathbf{Z}$ be a $K \times 2^K$ matrix (3×8 in our example) given by

$$\mathbf{Z} = \begin{pmatrix} 0\,0\,0 \\ \hline 0\,0\,1 \\ 0\,1\,0 \\ 1\,0\,0 \\ \hline 1\,0\,1 \\ 0\,1\,1 \\ 1\,1\,0 \\ \hline 1\,1\,1 \end{pmatrix}.$$

It follows that for $M_r = 1 + \mathbf{ZC}'$

$$M_r = \begin{cases} 1, & \text{for } \mathbf{z} = (z_1 = 0, z_2 = 0, z_3 = 0), \quad \text{model } g_0, \\ 5, & \text{for } \mathbf{z} = (z_1 = 0, z_2 = 0, z_3 = 1), \quad \text{model } g_1, \\ 3, & \text{for } \mathbf{z} = (z_1 = 0, z_2 = 1, z_3 = 0), \quad \text{model } g_2, \\ 2, & \text{for } \mathbf{z} = (z_1 = 1, z_2 = 0, z_3 = 0), \quad \text{model } g_3, \\ 6, & \text{for } \mathbf{z} = (z_1 = 1, z_2 = 0, z_3 = 1), \quad \text{model } g_4, \\ 7, & \text{for } \mathbf{z} = (z_1 = 0, z_2 = 1, z_3 = 1), \quad \text{model } g_5, \\ 4, & \text{for } \mathbf{z} = (z_1 = 1, z_2 = 1, z_3 = 0), \quad \text{model } g_6, \\ 8, & \text{for } \mathbf{z} = (z_1 = 1, z_2 = 1, z_3 = 1), \quad \text{model } g_7. \end{cases}$$

Thus, the posterior probability of M_r defines uniquely the posterior probability of a specific model and in particular, the posterior probability of the null model is given by,

$$p(M_r = 1|\mathbf{z}, \text{data}) = p(E(Y_{ij}) = \mu_0|\mathbf{z}, \text{data}) = p(g_0|\mathbf{z}, \text{data}).$$

13.3.2.1 Example: BVS For Gene 13386

Recall that in Sect. 13.2.2.3, we fitted two order-constrained models (g_5 and g_7) and the null model (g_0) to gene 13386 and we selected model g_5 since it had the lowest DIC value. In this section, we apply the BVS model for the same gene. In Winbugs, the BVS model can be implemented in a similar way to the model discussed in Sect. 13.2.2.1 with minor modifications. First, the mean structure needs to be defined in terms of the parameter vector $\boldsymbol{\beta}$, the design matrix $\mathbf{X}$, and the indicator variables $\mathbf{z}$.

```
for(i in 1:N)
{
    Y[i] ~dnorm(mu[i],tau)
    mu[i] <- mu0*X1[i]+ delta1*X2[i]*Z1+ delta2*X3[i]*Z2+ delta3*X4[i]*Z3
}
```

Next, we need to specify a prior and hyperprior for $\mathbf{z}$.

```
Z1~dbern(pi1)
Z2~dbern(pi2)
Z3~dbern(pi3)
pi1 ~ dunif(0,1)
pi2 ~ dunif(0,1)
pi3 ~ dunif(0,1)
```

Finally, the transformation function $1+\mathbf{ZC}'$ for the variable M_r (the object `modi` in the panel below) is defined by

```
modi<-1+Z1+Z2*2+Z3*4
for(r in 1:8)
{
pmod[r]<-equals(modi,r)
}
```

The object `pmod[r]` is a vector of eight binary variables.

Note that in each MCMC iteration one and only one variable in `pmod[r]` is equal to 1 and the rest are equal to zero. This implies that, for example, the posterior mean of the object `pmod[1]` $= p(M_r = 1|\mathbf{z}, \text{data}) = p(g_0|\mathbf{z}, \text{data})$ and the posterior mean of the object `pmod[2]` $= p(M_r = 2|\mathbf{z}, \text{data}) = p(g_3|\mathbf{z}, \text{data})$ (see also Ntzoufras 2002). We run the model for 50,000 iterations from which the first 20,000 were considered as the burn-in period and discarded from the analysis. As can be seen from the output panel below, the posterior means for the mean gene expression are monotone as expected. Figure 13.2a shows the expression data for gene 13386 with the parameter estimates for the posterior means obtained for models, g_5 and g_7 and the BVS model. Figure 13.2b shows the posterior probabilities for the eight models. The model with the highest posterior probability is g_5, `pmod[7]` $= \bar{p}(g_5|\mathbf{z}, \text{ data}) = 0.4186$, much higher than the posterior probability of g_7 (0.0457). Notice, however, that the posterior probability of g_1 (`pmod[5]`) is equal to 0.4059. This implies that the data support two different models; the first assumes that the mean gene expression is equal for the first two doses (g_5), while the second assumes that the mean gene expression is equal for the first three doses (g_1). A second pattern which is revealed in Fig. 13.2a is that the posterior means obtained for the BVS model, $\bar{\boldsymbol{\mu}}_{\text{BVS}} = (8.372, 8.395, 8.477, 9.092)$, are not equal to the posterior means obtained for model g_5 in Sect. 13.2.2.3, $\bar{\boldsymbol{\mu}}_{g_5} = (8.356, 8.356, 8.533, 9.093)$. This expected, but somewhat surprising, result will be clarified in the next section in which we focus on two main issues: (1) gene-specific posterior probability of the null model, $p_g(g_0|\mathbf{z}, \text{data})$ and (2) the posterior mean for gene expression.

```
# posterior means (BVS model)
node      mean    sd  MC error       2.5%    median  97.5%    start   sample
mu[1]    8.372   0.06535 0.001186    8.23    8.376    8.49    20000   30001
mu[4]    8.395   0.06303 0.001153    8.275   8.394    8.522   20000   30001
mu[7]    8.477   0.08304 0.001639    8.338   8.468    8.651   20000   30001
mu[10]   9.092   0.07977 5.758E-4    8.933   9.093    9.251   20000   3000


# posterior model probabilities
# pmod[1]=posterior probability of the null model
node      mean        sd       MC error  2.5%50%97.5%  start  sample
pmod[1] 1.0E-4    0.009999    5.747E-5  0.0 0.0 0.0    20000  30001
pmod[2] 5.0E-4    0.02235     3.308E-4  0.0 0.0 0.0    20000  30001
pmod[3] 6.333E-4  0.02516     2.764E-4  0.0 0.0 0.0    20000  30001
pmod[4] 2.0E-4    0.01414     1.48E-4   0.0 0.0 0.0    20000  30001
pmod[5] 0.4059    0.4911      0.01164   0.0 0.0 1.0    20000  30001
pmod[6] 0.1284    0.3345      0.008457  0.0 0.0 1.0    20000  30001
pmod[7] 0.4186    0.4933      0.0126    0.0 0.0 1.0    20000  30001
pmod[8] 0.0457    0.2088      0.003034  0.0 0.0 1.0    20000  30001
```

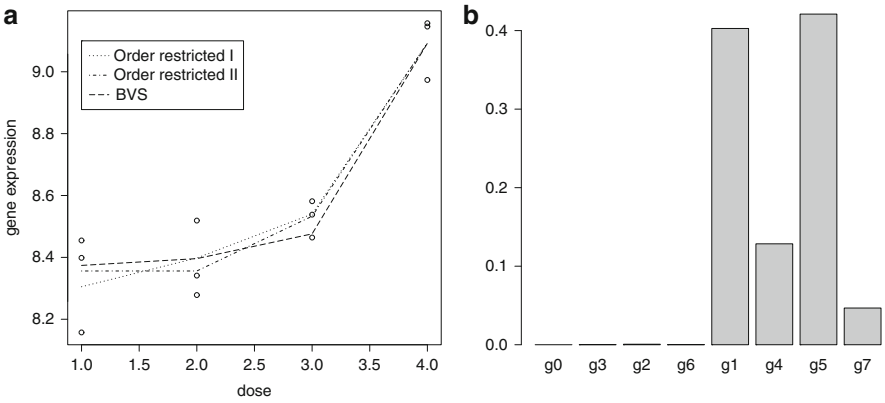

Fig. 13.2 Bayesian variable selection for gene 13386. Panel (**a**): posterior means obtained from the restricted model with four parameters (g_7: *dotted line*), restricted model with three parameters (g_5: *dotted-dashed line*), and the posterior means for the BVS model (*dashed line*). Panel (**b**): posterior probabilities of the models, $\bar{p}(g_5|\mathbf{z},\ \text{data}) = 0.4186$

13.3.2.2 Example: BVS for Gene 3413

Figure 13.3a shows the expression levels and fitted model for gene 3413. Note that the isotonic regression (solid line) predicts a g_5 model, but compared with the previous example, the mean profile is relatively flat. The posterior means obtained from the BVS model $\bar{\boldsymbol{\mu}}_{\text{BVS}} = (5.288, 5.316, 5.346, 5.394)$ (dashed line) is monotone as required, but similar to the isotonic regression, rather flat. Figure 13.3b shows the posterior probabilities of the models. The model with the highest posterior probability is the null model with $\bar{p}(g_0|\mathbf{z}, \text{data}) = 0.514$ (see output below). Hence, for this gene, the data clearly support the null model, i.e., a model with a flat mean profile.

```
# Order restricted model for gene 3413 (4 parameters)
node     mean     sd       MC error    2.5%    median    97.5%   start    sample
mu[1]    5.124    0.1377   0.001702    4.813   5.139     5.354   20000    30001
mu[4]    5.264    0.1124   7.712E-4    5.029   5.268     5.475   20000    30001
mu[7]    5.39     0.114    6.911E-4    5.175   5.386     5.629   20000    30001
mu[10]   5.56     0.144    8.539E-4    5.319   5.544     5.89    20000    30001


# BVS model for gene 3413
# posterior means
node     mean     sd       MC error    2.5%    median    97.5%   start    sample
mu[1]    5.288    0.112    0.00133     5.03    5.3       5.479   20000    30001
mu[4]    5.316    0.09891  8.884E-4    5.106   5.32      5.501   20000    30001
mu[7]    5.346    0.09962  7.837E-4    5.154   5.344     5.556   20000    30001
mu[10]   5.394    0.1241   9.833E-4    5.194   5.378     5.69    20000    30001

#posterior model probabilities
#pmod[1] is the posterior probability of the null model
node     mean     sd       MC error    2.5     %50%      97.5%   start    sample
pmod[1]  0.514    0.4998   0.006476    0.0     1.0       1.0     20000    30001
pmod[2]  0.1074   0.3096   0.00424     0.0     0.0       1.0     20000    30001
pmod[3]  0.1299   0.3362   0.004264    0.0     0.0       1.0     20000    30001
```

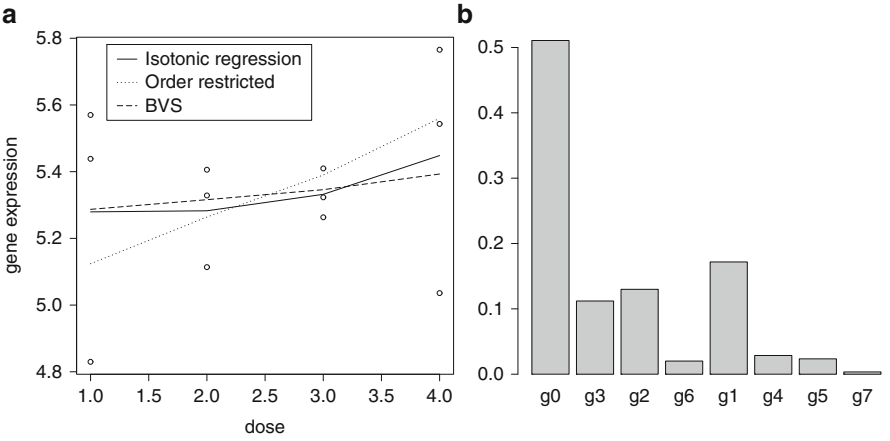

Fig. 13.3 Bayesian variable selection for gene 3413. Panel (**a**): posterior means obtained from the restricted model with four parameters (*dotted line*), isotonic regression (*solid line*), and the posterior means for the BVS model (*dashed line*). Panel (**b**): posterior probabilities of the models, $\bar{p}(g_0|\mathbf{z}, \text{data}) = 0.514$

```
pmod[4] 0.01637     0.1269     0.001417    0.0 0.0 0.0  20000  30001
pmod[5] 0.1743      0.3794     0.004662    0.0 0.0 1.0  20000  30001
pmod[6] 0.0282      0.1655     0.001607    0.0 0.0 1.0  20000  30001
pmod[7] 0.02627     0.1599     0.001639    0.0 0.0 1.0  20000  30001
pmod[8] 0.003567    0.05961    4.734E-4    0.0 0.0 0.0  20000  30001
```

Three questions arise now: (1) How come that the BVS model did not predict a flat mean profile? (2) Although $\bar{\boldsymbol{\mu}}_{\text{BVS}}$ is monotone, how come that it is not equal to the posterior means obtained for model g_7 (dotted line)? and (3) Why is $\bar{\boldsymbol{\mu}}_{\text{BVS}}$ not equal to the isotonic regression?

The isotonic regression obtained by the PAVA pools together, the means in the doses for which the order in the unconstrained means is violated. For gene 3413, the isotonic regression predicts a pattern of model g_5, but it does not mean that we reject the null hypothesis of no dose effect. The answers of the questions above are related to the interpretation of the posterior means for the BVS model.

Let us consider an MCMC simulation of T iterations and let $\boldsymbol{\beta}^{(t)} = (\mu_0^{(t)}, z_1^{(t)}\delta_1^{(t)}, z_2^{(t)}\delta_2^{(t)}, z_3^{(t)}\delta_3^{(t)})$ be the parameter vector at iteration $t, t = 1, \ldots, T$ and $\boldsymbol{\mu}^{(t)} = \mathbf{X}\boldsymbol{\beta}^{(t)}$ be the mean gene expression in iteration t, respectively. The posterior mean $\bar{\boldsymbol{\mu}}_{\text{BVS}}$ is calculated using MCMC integration, i.e.,

$$\bar{\boldsymbol{\mu}}_{\text{BVS}} = \frac{1}{T}\Sigma_{t=1}^{T}\boldsymbol{\mu}^{(t)}.$$

Hence, $\bar{\boldsymbol{\mu}}_{\text{BVS}}$ is the average of all means which were calculated during the MCMC simulation. In our example, in 51.4% of the iterations, $\boldsymbol{\mu}^{(t)}$ was equal to $(\mu_0^{(t)}, \mu_0^{(t)}, \mu_0^{(t)}, \mu_0^{(t)})$ (the mean structure of g_0); in 12.99% of the iterations, it was

equal to $(\mu_0^{(t)}, \mu_0^{(t)}, \mu_0^{(t)}+\delta_2^{(t)}, \mu_0^{(t)}+\delta_2^{(t)})$ (the mean structure of g_2); in 10.74% of the iterations, it was equal to $(\mu_0^{(t)}, \mu_0^{(t)}+\delta_1^{(t)}, \mu_0^{(t)}+\delta_1^{(t)}, \mu_0^{(t)}+\delta_1^{(t)})$ (the mean structure of g_3), etc. The posterior means obtained from the BVS model should be interpreted as the model average for the eight possible models which were fitted during the MCMC simulations. Since at each iteration $\boldsymbol{\mu}^{(t)}$ is monotone, $\bar{\boldsymbol{\mu}}_{\text{BVS}}$ is monotone as well. However, we can see clearly the shrinkage through the overall mean since the weight of g_0 is the highest. As we argued above, in this example, the data support the null model for which the posterior probability $\bar{p}(g_0|\mathbf{z}, \text{data}) = 0.514$. Up to this point in this chapter, we did not address the question of how to select the subset of genes which are differentially expressed. In Chaps. 7 and 8, we used either BH-FDR or SAM in order to select a subset of genes which were declared differentially expressed. The posterior probability of the null model will be a key concept in the next section in which we discuss the issue of inference and multiplicity adjustment within the hierarchical Bayesian framework.

13.4 Inference and Adjustment for Multiplicity

13.4.1 The Direct Posterior Probability Approach for Multiplicity Adjustment

Newton, Wang and Kendziorski (2007) discussed a setting of a microarray experiment with m genes and defined an equally expressed gene (EE_g, $g = 1, \ldots, m$) as a gene which is not differentially expressed across the condition of the experiment and $p(\text{EE}_g|\text{data})$ as the posterior probability of the gth gene of being EE. In the previous section, we defined the posterior probability of the rth model and in particular the posterior probability of the null model. In Newton et al. (2007) notation

$$p(\text{EE}_g|\text{data}) = p_g(g_0|\mathbf{z}, \text{data}).$$

In this section, we use the posterior probability of the null model for multiplicity adjustment. Similar to the example of gene 13368, a gene for which there is a strong evidence for a monotone dose-response relationship will have small value of $p_g(g_0|\mathbf{z}, \text{data})$ and vice versa, and a gene with a weak evidence of a monotone dose-response relationship (such as gene 3413) will have a high value of $p_g(g_0|\mathbf{z}, \text{data})$. The *direct posterior probability approach*, proposed by Newton et al. (2004, 2007), uses the dual characteristic of $p_g(g_0|\mathbf{z}, \text{data})$ in order to adjust for multiplicity. Let α be a threshold value and let N be a discovery list in which all genes satisfying $p_g(g_0|\mathbf{z}, \text{data}) \leq \alpha$ are included. Hence, the discovery list contains all genes with posterior probability of the null model smaller than α, and at the same time, $p_g(g_0|\mathbf{z}, \text{data})$ is the probability that we commit an error when we assign the gth gene to the discovery list. Newton et al. (2007) term this the dual characteristic of $p_g(g_0|\mathbf{z}, \text{data})$.

Let I_g be an indicator variable such that

$$I_g = \begin{cases} 1 & \text{if } p_g(g_0|\mathbf{z}, \text{data}) \leq \alpha, \\ 0 & \text{if } p_g(g_0|\mathbf{z}, \text{data}) > \alpha. \end{cases}$$

Since $p_g(g_0|\mathbf{z}, \text{data})$ is also the probability that the assignment of the gth gene to the discovery list is incorrect, the expected number of false discoveries (FD) is

$$E(\text{FD}) = \sum_{g=1}^{m} p_g(g_0|\mathbf{z}, \text{data}) I_g = \text{cFD}(\alpha).$$

Newton et al. (2007) defined the conditional (on the data) false discovery rate as

$$\text{cFDR}(\alpha) = \frac{1}{N(\alpha)} \text{cFD}(\alpha) = \frac{\text{Expected number of false discoveries}}{\text{Number of discoveries}},$$

where $N(\alpha)$ is the number of genes declared EE for a given threshold α. Newton et al. (2004) terms this approach the *direct posterior probability approach* for multiplicity adjustment. Note that cFDR(α) is interpreted as the average error that we make when we assign a gene to the discovery list. The value of α is selected in such a way that cFDR(α) will not exceed a pre specified threshold τ.

13.4.2 Application to the Data

In this section, we apply the *direct posterior probability approach* discussed above for multiplicity adjustment. As we mentioned above, the framework enables adjustment for false discovery rates among the significant genes. We use the `R2WINBUGS` package to fit a gene-specific model and to obtain the posterior probability of the null model. For each gene, an MCMC simulation of 20,000 iterations (from which 5,000 are used as burn-in period) was used to fit the BVS model. Figure 13.4 shows the relationship between false discovery rate (cFDR), number of significant genes, and cutoff values. Panel a shows that an increase in cutoff values results in an increase in false discovery rate. However, the false discovery rate reaches its maximum of 0.2 at the cutoff of about 0.5. Panel b also shows an increase in the number of significant genes with an increase in cutoff values. The implication of the finding is that, as expected, the higher the cutoff value, the larger the number of significant genes and consequently, the higher the proportion of false positives among the significant genes. Similar to the frequentists practice, one may wish to control for false discovery rate at 1% or 5%, which corresponds to cutoff values of 0.029 and 0.102, respectively. Based on these cutoff values, the corresponding number of significant genes is 609 and 3,295 genes, respectively.

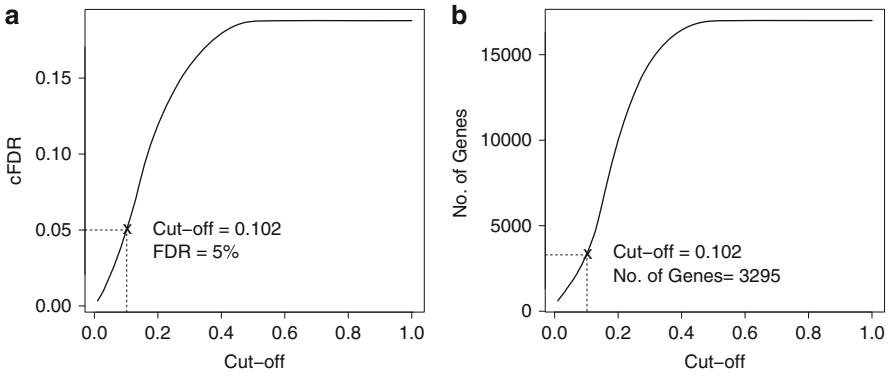

Fig. 13.4 Adjustment for multiplicity. Panel (**a**): relationship between the estimated false discovery rate (cFDR) and the cutoff values. Panel (**b**): the relationship between number of significant genes and the cutoff values

13.5 Discussion

There are two main challenges in Bayesian analysis of dose-response microarray data. The first is the presence of strictly equality relationship between differences in gene expressions at different doses of a therapeutic compound, and the second is the question of how to adjust for multiplicity. In this chapter, we discuss the BVS method as an approach to circumvent the first problem by replacing strict equality between doses by a common parameter. We have shown that the BVS model estimates equal mean for two successive dose levels, i and $i-1$, whenever the corresponding binary variable for the ith dose level, $z_i = 0$. Further, we have shown that the posterior probability of the null model can be estimated and can be used for multiplicity adjustment. Similar approach for multiplicity adjustment within the hierarchical Bayesian framework is discussed by Lewin et al. (2007) and Broet et al. (2004). The materials presented in this chapter resulted from an ongoing research; we are currently investigating Bayesian isotonic regression (Dunson and Neelon 2003) approach for dose-response microarray data, and our findings will be reported in the near future through a scientific publication.

References

Broet, P., Lewin, A., Richardson, S., Dalmasso, C., & Magdelenat, H. (2004). A mixture model based strategy for selecting sets of genes in multiclass response microarray experiments. *Bioinformatics, 20*(16), 2562–2571.

Chen, M. H., & Kim, S. (2008). The Bayes factor versus other model selection criteria for the selection of constrained models. In H. Hoijtink, I. Klugkist, & P. Boelen (Eds.), *Bayesian evaluation of informative hypotheses*. Berlin: Springer.

Dellaportas, P., Forster, J. J., & Ntzouras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing, 12*, 27–36.

Dunson, D. B., & Neelon, B. (2003). Bayesian inference on order constrained parameters in generalized linear models. *Biometrics, 59*, 286–295.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed). Boca Raton: Chapman & Hall/CRC.

George, E. I., & McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association, 88*, 881–889.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996) *Markov chain Monte Carlo in practice*. London: Chapman & Hall.

Gelfand, A. E., Smith, A. F. M., & Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems. *Journal of the American Statistical Association, 87*, 523–532.

Hoijtink, H., Klugkist, I., & Boelen, P. (Eds.). (2008). *Bayesian evaluation of informative hypotheses*. Berlin: Springer.

Kato, B. S., & Hoijtink, H. (2006). A bayesian approach to inequality constrained linear mixed models: estimation and model selection. *Statistical Modelling, 6*, 231–249.

Klugkist, I., Kato, B., & Hoijtink, H. (2005a) Bayesian model selection using en- compassing priors. *Statistica Neerlandica, 59*(1), 57–69.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005b) Inequality constrained analysis of variance: A bayesian approach. *Pyschological Methods, 10*(4), 477–493.

Klugkist, I., & Mulder, J. (2008). Bayesian estimation for inequality constrained analysis of variance. In H. Hoijtink, I. Klugkist, & P. Boelen (Eds.), *Bayesian evaluation of informative hypotheses*. Berlin: Springer.

Lewin, A., Bochkina, N., & Richardson, S. (2007). Fully bayesian mixture model for differential gene expression: simulations and model checks. *Statistical Applications in Genetics and Molecular Biology, 6*(1). Article 36.

Myung, J. I., Karabatsos, G., & Iverson, G. J. (2008). A statistician's view on Bayesuan evaluation of informative hypotheses. In H. Hoijtink, I. Klugkist, , & P. Boelen (Eds.), *Bayesian evaluation of informative hypotheses*. Berlin: Springer.

Newton, M. A., Noueiry, D., Sarkar, D., & Ahlquist, P. (2004). Detecting differential gene expression with a semi-parametric hierarchical mixture method. *Biostatistics, 5*(2), 155–176.

Newton, M. A., Wang, P., & Kendziorski, C. (2007). Hierarchical mixture models for expression profiles. In K. M. Do, P. Muller, & M. Vannucci (Eds.), *Bayesian Inference for gene expression and proteomics*. Cambridge: Cambridge University Press.

Ntzoufras, I. (2002). Gibbs variable selection using BUGS. *Journal of Statistical Software, 7*(7), 1–19.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 64*(4), 583–639.

O'Hara, R. B., & Sillanpaa, M. J. (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis, 4*(1), 85–118.

Whitney, M., & Ryan, L. (2009). Quantifying dose-response uncertainty using bayesian model averaging. In R. M. Cooke (Ed.), *Uncertainty modeling in dose-response*. New York: Wiley.

Chapter 14
Model-Based Approaches

Setia Pramana, Ziv Shkedy, Hinrich W. H. Göhlmann, Willem Talloen, An De Bondt, Roel Straetemans, Dan Lin, and José Pinheiro

14.1 Introduction

In the previous chapters, we have discussed different methods to identify genes for which a monotone trend with respect to doses can be detected. The aim of the analysis presented in this chapter is not to detect genes with a significant dose-response relationship but to use parametric dose-response models in order to compare between several compounds or between the characteristics of the dose-response curves for several genes in a single or multiple compound experiment. The basic methodology for parametric dose-response models presented in this chapter

S. Pramana (✉)
Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden
e-mail: setia.pramana@ki.se

Z. Shkedy
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be

D. Lin
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

H.W.H. Göhlmann · W. Talloen · An De Bondt
Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
e-mail: HGOEHLMA@its.jnj.com; WTALLOEN@its.jnj.com; ADBONDT@its.jnj.com

R. Straetemans
Ablynx NV, Zwijnaarde, Belgium
e-mail: Roel.Straetemans@ablynx.com

J. Pinheiro
Janssen Pharmaceutical Companies of Johnson & Johnson, Titusville, NJ, USA
e-mail: JPinhei1@its.jnj.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_14,

Table 14.1 Number of significant genes for the six compounds. Resampling-based inference using the LRT and the BH-FDR method for multiplicity adjustment ($\alpha = 0.05$)

JNJa	JNJb	JNJc	CompB	CompA	CompC
211	251	164	332	72	242

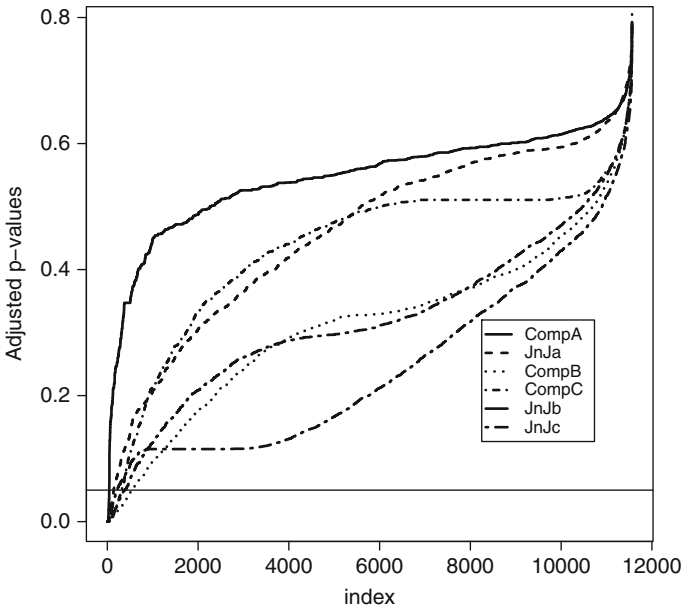

Fig. 14.1 BH-FDR adjusted p values for the LRT for the six compounds

was introduced in Chap. 4. In contrast with Chap. 10 in which the classification post-selection procedure was applied to order-restricted ANOVA models, in this chapter we focus on the case in which several parametric models are fitted to the gene expression data and we discuss model averaging techniques for the estimation of the ED_{50} parameter. The case study we use for illustration is the pharmacological activity of antipsychotic (EPS) study introduced in Chap. 5. The study consists of a dose-response microarray experiment, in which gene expression data are available at six different dose levels for 11,562 genes. In total, six different antipsychotic compounds were used in the experiment: CompA, CompB, CompC, and three additional compounds: JnJa, JnJb, and JnJc. Table 14.1 presents the number of significant genes which were found to be significant using the likelihood ratio test for each compound. The BH-FDR adjusted p values are presented in Fig. 14.1. After adjustment for multiplicity, the number of significant genes ranges from 72 (for CompA) to 332 (for CompB). Figure 14.2 shows the expression levels for the gene FOS which was found to be significant for all compounds and will be used in the following sections for illustration.

For the remainder of this chapter, we use gene-specific parametric dose-response models, discussed in Chap. 4, of the form

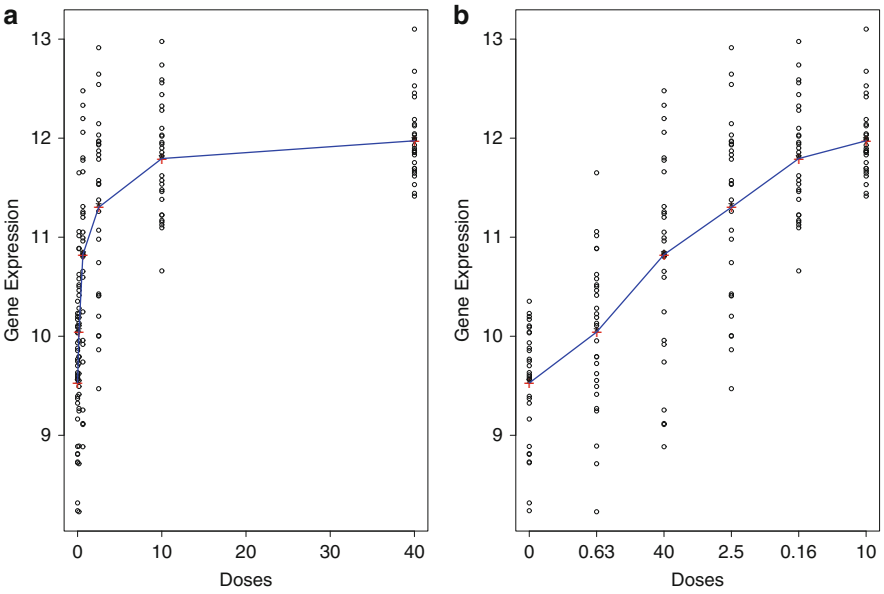

Fig. 14.2 Gene expression data and isotonic regression for the gene FOS. (**a**) Original scale; (**b**) logscale

$$Y_{ij} = f(\boldsymbol{\theta}, d_i) + \varepsilon_{ij}, \quad i = 1, \ldots, K, \; j = 1, \ldots, n_i. \tag{14.1}$$

Here, f is the unknown parametric function which describes the dose-response relationship, $\boldsymbol{\theta}$ is the unknown parameter vector to be estimated, Y_{ij} is the gene expression level of the jth sample at dose level i, and d_i is the dose. For illustration, we consider the following five dose-response nonlinear models:

$$f = \begin{cases} E_0 + \dfrac{E_{\max} d}{\mathrm{ED}_{50} + d} & \text{Emax model,} \\ E_0 + \dfrac{E_{\max} d^h}{\mathrm{ED}_{50}^h + d^h} & \text{sigmoid Emax model,} \\ E_0 + \dfrac{E_{\max}}{1 + \exp\left[\dfrac{(\mathrm{ED}_{50} - d)}{\delta}\right]} & \text{4PL model,} \\ E_0 + \delta d & \text{linear model,} \\ E_0 + \delta \log(d + \eta) & \text{log-linear model.} \end{cases} \tag{14.2}$$

For the log-linear model, η is a fixed offset. The above models will be used in this chapter for illustration but other parametric models, listed in the `DoseFinding` R package (Bornkamp, Pinheiro, and Bretz, 2009), can be used as well. This chapter is organized as follows. In Sects. 14.2 and 14.3, we focus on the analysis of a single gene and use the parametric models above in order to estimate the dose

for which the gene expression is halfway to the maximum level (ED_{50}), while in Sects. 14.4 and 14.5, we focus on the case in which several genes are ranked based on their ED_{50} parameter estimates.

14.2 Model Selection, Model Averaging, Akaike Weights, and Posterior Model Probabilities

In Chaps. 10 and 11, we used information criteria in order to cluster subsets of genes with similar dose-response curve shapes. In this chapter, we use the Akaike information criterion in order to obtain a model average estimate for both the mean gene expression and the ED_{50}. As discussed in Chap. 10, for a given set of candidate models, $f_1, \ldots, f_R$, the AIC values of the models are rescaled by

$$\Delta\text{AIC}_r = \text{AIC}_r - \text{AIC}_{\min}, \qquad r = 1, \ldots, R,$$

with $\text{AIC}_{\min} = \min(\text{AIC}(f_1), \ldots, \text{AIC}(f_R))$. The Akaike weight for the rth model is given by

$$P_{\text{A}}(f_r|D) = \frac{\exp\left(-\frac{1}{2}\Delta\text{AIC}_r\right) P(f_r)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta\text{AIC}_r\right) P(f_r)}.$$

As discussed in Chap. 10, the Akaike weight $P_{\text{A}}(f_r|D)$ can be interpreted as the weight of evidence that model f_r is the best KL model given a set of R models and given that one of the models in the set must be the best KL model. Akaike weights can be interpreted as the posterior probabilities of the models. For the case with non-informative prior probabilities $P(f_r) = 1/R$, the Akaike weights are identical to the weights discussed in Pinheiro et al. (2006b).

Let $\hat{f}_1, \hat{f}_2 \ldots, \hat{f}_R$ be the estimated mean gene expression obtained from a set of R candidate models and $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_R$ be the parameter estimates of primary interest (for example, the ED_{50}) obtained for each candidate model, respectively. A model selection procedure implies that one of the candidate models will be selected according to an information criterion, and the parameter of primary interest will be estimated from the selected model. Let us assume that the AIC is used for model selection. In this case, the post-selection estimate (Claeskens and Hjort 2008) for the mean gene expression and θ are given, respectively, by

$$\hat{f}_{\text{AIC}} = \sum_{r=1}^{R} I_r \hat{f}_r \quad \text{and} \quad \hat{\theta}_{\text{AIC}} = \sum_{r=1}^{R} I_r \hat{\theta}_r. \tag{14.3}$$

Here, I_r is an indicator variable such that

$$I_r = \begin{cases} 1 & \text{if } \text{AIC}(f_r) = \text{AIC}_{\min}, \\ 0 & \text{otherwise.} \end{cases} \quad \text{or} \quad \text{equivalently } I_r = \begin{cases} 1 & \Delta\text{AIC}_r = 0, \\ 0 & \Delta\text{AIC}_r > 0. \end{cases}$$

Note that $\sum_{r=1}^{R} I_r = 1$ and $I_r = 1$ for the model with the smallest value of AIC. The post-selection estimates in (14.3) do not take into account the uncertainty associated with the model selection procedure since they are based on a single model. Thus, the fact that R candidate models were fitted to the data from which the best model was selected is not reflected in the variability associated with the post-selection estimates. This can be done if we replace the indicator variable in (14.3) with a weight $w_r(R)$, $\sum_{r=1}^{R} w_r(R) = 1$. Hence, the estimates are averaged over all candidate models, that is

$$\hat{f}_{\text{MA}} = \sum_{i=1}^{R} w_i(R)\hat{f}_i \quad \text{and} \quad \hat{\theta}_{\text{MA}} = \sum_{i=1}^{R} w_i(R)\hat{\theta}_i.$$

$\hat{\theta}_{\text{MA}}$ is called the model average estimate for θ. Claeskens and Hjort (2008) referred to the case that Akaike weights are used to average the models, i.e., $w_r(R) = P_{\text{A}}(f_r|D)$ as the smooth AIC weights. In the context of dose-response modeling, model averaging methods are discussed by Pinheiro et al. (2006a,b), Bretz et al. (2005), Bornkamp et al. (2009) and Whitney and Ryan (2009).

14.2.1 *The Estimation of* $\mathbf{ED_{50}}$ *Within the Dose Range*

The ED_{50} parameter represents the dose at which the mean response is halfway to the maximum effect. As illustrated in Chap. 4, the R function `gnls()` allows one to estimate the ED_{50} directly as long as it can be specified as a parameter in the dose-response model (for example, θ_2 in the 4PL model). This implies that if the maximum effect, the minimum effect, or both are reached outside the range of the data, the ED_{50} can be estimated outside the range of the dose used in the experiment as well. The R function `MCPMod()` of the R package `DoseFinding` estimates the ED_{50} in a slightly different way. The ED_{50} estimate is estimated as the dose that corresponds to the halfway between the estimated mean at the lowest and the highest dose levels (hence, not as a parameter of the model). Figure 14.3 illustrates a scenario in which the maximum effect is reached at a dose level that is higher than the maximum dose administered in the experiment. The R function `MCPMod()` estimates the ED_{50} as the halfway point from the estimated mean response at the lowest dose to the estimated mean response at the highest dose. This is done by a grid search, in which the dose corresponding to the halfway point between the estimated mean response at the minimum and maximum doses is found. In that way, the ED_{50} represents the dose halfway to the maximum effect within the range of the data. In the context of dose-response microarray experiments, this approach has a major advantage. The difference between the (model-based) mean gene expression at the lowest dose and the highest dose is the fold change [or log(fold change)]. Hence, the ED_{50} can be interpreted as the dose that corresponds to the half-fold change.

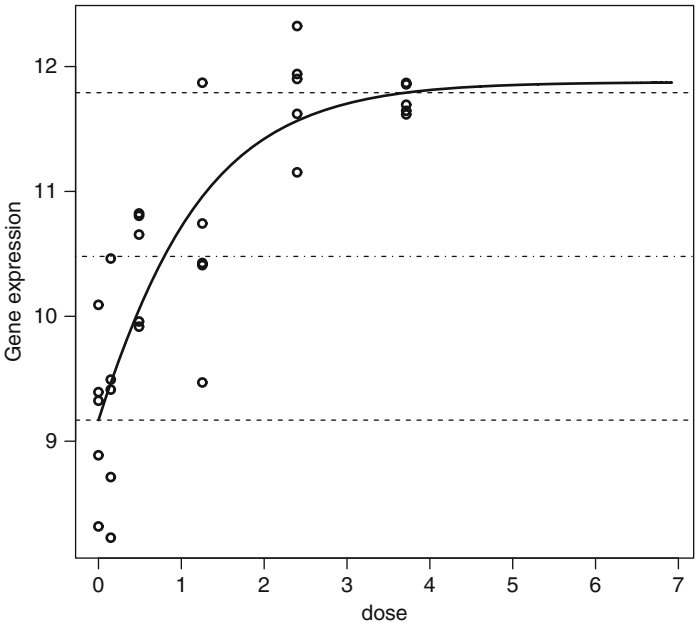

Fig. 14.3 Estimation of the ED_{50} by the R package `DoseFinding`. *Dashed lines* are the model-based minimum and maximum gene expression. The *dashed dotted line* represent the halfway from the minimum to the maximum effect

14.2.2 Fitting Parametric Models for Dose-Response Gene Expression Data Using the R Package `DoseFinding`

The methodology discussed above can be implemented in R using the package `DoseFinding`. For illustration, we use the dose-response gene expression data for the gene FOS with the compound CompA. A partial printout of the gene expression data is shown below.

```
> datafos1
  resp       dose  compound
9.393643    0.00   CompA
9.958406    0.63   CompA
11.869202  40.00   CompA
  .          .   .       .
  .          .   .       .
  .          .   .       .
10.409598   2.50   CompA
8.227862    0.16   CompA
8.712944    0.16   CompA
```

Within the `DoseFinding` R package, the function we use for modeling is the `MCPMod()`. A general call for the `MCPMod()` function has the form

```
MCPMod(response ~ dose, models to be fitted,  model selection criterion)
```

For our example, we use five different models in (14.2) specified in the R object `models1` as follows:

```
> models1 <- list(linear = NULL, linlog = NULL, emax = 1.288150,
sigEmax=c( 1.3762727, 0.8783642 ), logistic= c(-8.7298, 2.461841))
```

The R object `models1` is given as an input for the function `MCPMod()`. The option `selModel` = "`aveAIC`" implies that the AIC is used as the model selection criterion for the calculation of Akaike weights and the option `doseEst` = "`ED`" and `doseEstPar = 0.5` implies that the parameter for which model averaging will be performed is the ED_{50}. For our example, we use the following code:

```
> dfeA1 <- MCPMod(resp ~ dose,datafos1 , models1, alpha = 0.05,
                  doseEstPar = 0.5, pVal = TRUE,
                   selModel = "aveAIC", doseEst = "ED", off = 1,
                   critV = TRUE, optimizer = "bndnls", lenDose=1000)
```

Figure 14.4 shows the fitted models (panel a) and the model average estimate (panel b) calculated according to (14.3).

The object `dfeA1` contains the information about the prior probability for each model ($1/R = 0.2$, the default) in our case.

```
> summary(dfeA1)
MCPMod
Input parameters:
alpha = 0.05
alternative: one.sided, one sided
model selection: aveAIC
prior model weights:
sigEmax emax logistic linlog linear
0.2       0.2 0.2      0.2     0.2
dose estimator: ED (p = 0.5)
optimizer: bndnls
```

The next part of the output presents the AIC values for each model. The model with the best goodness of fit for the gene FOS is the Emax model with the AIC equal to 63.30 which corresponds to the highest Akaike weight $\hat{P}_A(f_r|D) = 0.64$. Note that Akaike weight of the linear model which fits the data poorly is equal to 0.

```
AIC criterion:
sigEmax emax logistic linlog linear
 65.15  63.30 67.53   69.51  86.37

Selected for dose estimation:
sigEmax emax logistic linlog linear

Model weights:
sigEmax emax logistic linlog linear
0.254  0.640  0.077    0.029 0.000
```

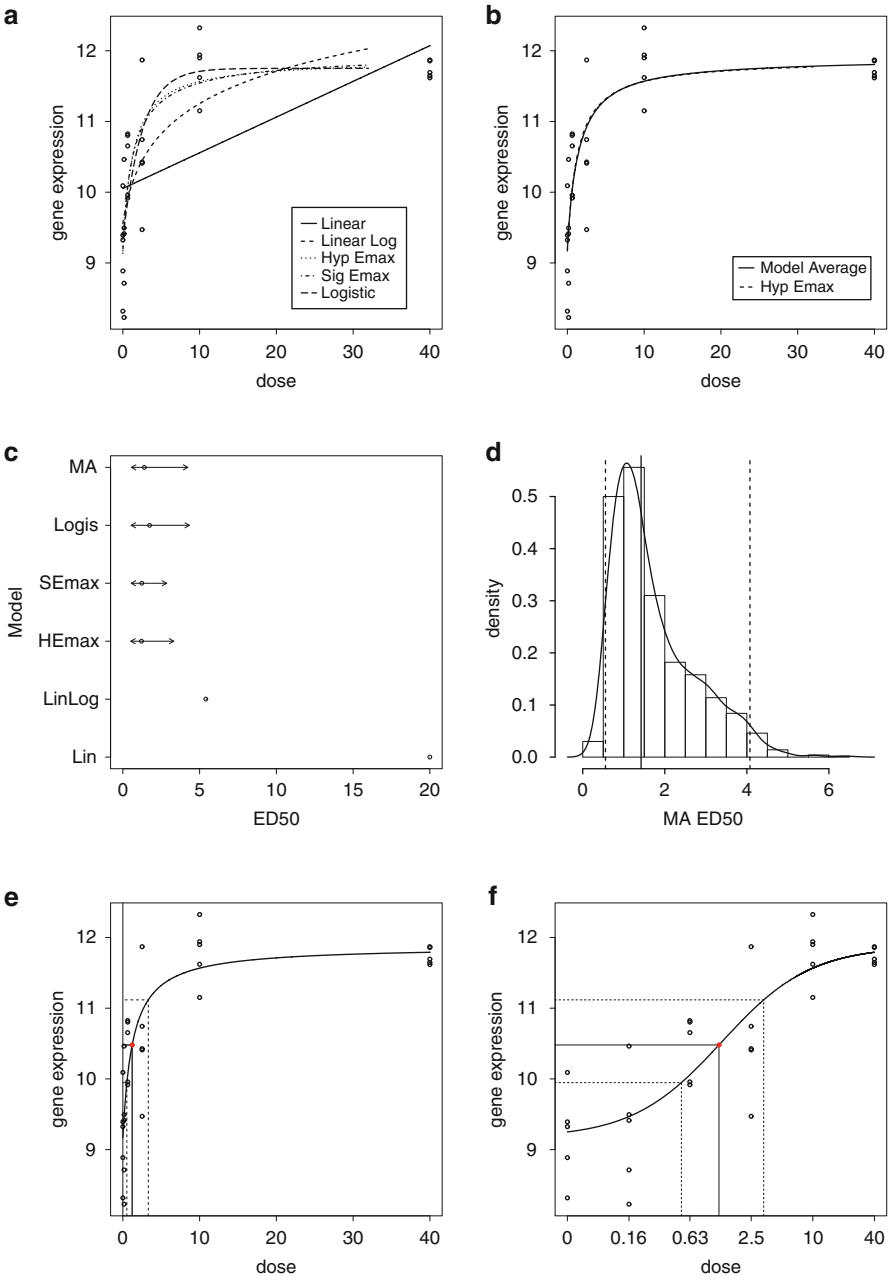

Fig. 14.4 Estimation of ED_{50} for the gene FOS. Panel (**a**): five parametric models. Panel (**b**): model average estimate for the mean gene expression and the best fitted model. Panel (**c**): model specific and model average estimates of the ED_{50}. Panel (**d**): histogram and density estimate for the bootstrap replicates for the model averaged estimates of the ED_{50}. The *vertical lines* represent the 95% CI. Panel (**e**): data and predicted model based on model averaging. Panel (**f**): data and predicted model presented on log scale

Parameter estimates for each model are shown in the panel below.

```
Parameter estimates:

sigEmax model:
 e0    eMax ed50   h
9.133 2.833 1.376 0.878

emax model:
e0     eMax ed50
9.169 2.706 1.288

logistic model:
 e0    eMax ed50 delta
7.102 4.653 0.040 1.815

linlog model:
e0    delta
9.558 0.707
linear model:
e0     delta
10.052 0.050
```

For the gene FOS, the ED_{50} calculated by `MCPMod()` for the different models is given below. The model average estimate for the ED_{50} calculated according to (14.3) is equal to 1.423. Note that, as discussed in Sect. 14.2.1, the ED_{50} is estimated by grid search within the dose range. For that reason, the parameter estimate for the ED_{50} obtained for the sigmoid E_{max} model (1.376) is not equal to the `Dose estimate` for this model (1.241).

```
Dose estimate
Estimates for models
      sigEmax emax  logistic linlog linear
ED50% 1.241   1.241  2.042   5.405   20.02

Model averaged dose estimate
ED50%
1.423
```

In the next stage, we obtain 90% bootstrap confidence intervals using the function `bootMCPMod()`. Figure 14.4c–f shows the parameter estimates for the ED_{50} obtained from the different models and their corresponding 90% bootstrap confidence intervals.

```
> resA1 <- bootMCPMod(dfeA1 , nSim = 1000, seed=123)
> resA1
MCPMod Bootstrap Dose Calculations
Number of Simulations : 1000
Perc. NA: 0
Bootstrap Mean: 1.725
Original Estimator: 1.423
Symmetric 0.9-Confidence Interval:
[ 0.621 , 3.758 ]
```

The model averaging technique allows us to compare between: (1) the expression levels of the same gene obtained for different compounds, (2) the expression levels of all significant genes for a specific compound, and (3) the expression levels of several genes obtained for different compounds. The three types of comparisons are discussed in Sects. 14.3–14.5, respectively.

14.3 Testing for Compound Effects Based on a Single Model

The first comparison of interest is the one between the expression levels of the same gene for different compounds. This allows us to evaluate the compound activity based on gene expression data. For the analysis presented in this section, six compounds are used: CompA, CompB, CompC, and three additional compounds: JnJa, JnJb, and JnJc. In the section, the comparison is based on the Emax model. In order to compare between the six compounds, we formulate the parameters in the model as a linear function of the treatment, that is,

$$\begin{pmatrix} E_{0_\ell} \\ \mathrm{ED}_{50_\ell} \\ E_{max_\ell} \end{pmatrix} = \begin{pmatrix} \theta_1 + \gamma_\ell \\ \theta_2 + \delta_\ell \\ \theta_3 + \eta_\ell \end{pmatrix}, \quad \ell = 2, \ldots, L. \tag{14.4}$$

Here, $\theta_1 + \gamma_\ell$, $\theta_2 + \delta_\ell$ and $\theta_3 + \eta_\ell$ are compound-specific E_0, ED_{50}, and $E_{\max}$, respectively. Formally, for each one of the parameters, we wish to test the null hypothesis of no compound effect. For example, for the ED_{50} parameter, we formulate the following hypotheses:

$$\begin{aligned} H_0 &: \delta_\ell = 0, \\ H_1 &: \delta_\ell \neq 0. \end{aligned} \tag{14.5}$$

The Emax model can be fitted in R using the function `gnls()` in the following way:

```
> fos.Com0 <- gnls(resp ~ E0 + ( (dose *Emax)/ (xmid+dose)),data = datafos,
+   start = c(Emax = dif,E0 = miny,xmid = 0.63))
```

Note that the model specified above does not include compound effect but assumes that there a single dose-response relationship for all compounds as can be seen in the panel below.

```
> summary(fos.Com0)
Generalized nonlinear least squares fit
  Model: resp ~ E0 + ((dose * Emax)/(xmid + dose))
  Data: datafos
       AIC      BIC    logLik
  384.7557 397.3457 -188.3779

Coefficients:
        Value Std.Error  t value p value
Emax 2.385194 0.1494264 15.96233   0e+00
E0   9.548798 0.1226443 77.85765   0e+00
xmid 0.630100 0.1624210  3.87942   1e-04
```

The next model we fit assumes that compounds are different in $E_{\max}$ and ED_{50}, but the baseline parameter E_0 is equal for all compounds. Note that since the first dose is equal to 0, E_0 can be used in order to estimate the baseline gene expression using data from all compounds. The change in parameterization can be implemented using the option `params`.

```
params= list(Emax+xmid~compound,E0~1)
```

In order to fit the model, we update the `gnls()` function as follows:

```
> ##  A model common E0  ##
> fos.Com1 <- gnls(resp ~ E0 + ((dose *Emax)/ (xmid+dose)),data = datafos ,
+   params= list(Emax+xmid~compound,E0~1),
+   start = c(Emax = dif,E0 = miny, rep(0,5), xmid = 0.63, rep(0,5)))
```

Parameter estimates are shown in the panel below.

```
> summary(fos.Com1)
Generalized nonlinear least squares fit
  Model: resp ~ E0 + ((dose * Emax)/(xmid + dose))
  Data: datafos
       AIC      BIC    logLik
  236.5949 280.6598 -104.2974

Coefficients:
                       Value Std.Error   t value p value
Emax.(Intercept)    2.527145 0.2105666  12.00164  0.0000
Emax.compoundJnJa   0.491722 0.4200471   1.17063  0.2435
Emax.compoundJnJb  -0.174543 0.2520476  -0.69250  0.4896
Emax.compoundJnJc  -0.222660 0.2544501  -0.87507  0.3829
Emax.compoundCompB -0.038892 0.2526987  -0.15391  0.8779
Emax.compoundCompC  0.720935 0.2410640   2.99064  0.0032
xmid.(Intercept)    1.869673 0.6478455   2.88599  0.0044
xmid.compoundJnJa   5.625482 2.9956672   1.87787  0.0622
xmid.compoundJnJb  -1.661514 0.6480431  -2.56389  0.0113
xmid.compoundJnJc  -1.535046 0.6517436  -2.35529  0.0197
xmid.compoundCompB -1.409720 0.6534863  -2.15723  0.0325
xmid.compoundCompC -1.725857 0.6472122  -2.66660  0.0085
E0                  9.424942 0.0725147 129.97280  0.0000
```

In the next step, two additional models are fitted. The first assumes a common $E_{\max}$ and a common E_0 for all compounds and the second assumes different parameters for each compound. The two models can be fitted using `gnls()` using the following two `params` statements, respectively:

```
> ##  For a model with the common E0 and Emax (fos.Com2)##
> params= list(Emax+E0~1,xmid~compound)
> ##  For a model with compound specific parameters (fos.com3) ##
> params=list(E0 + Emax+xmid~compound)
```

As can be seen in the panel below, the first model implies a compound-specific ED_{50}.

```
> summary(fos.Com2)
Generalized nonlinear least squares fit
  Model: resp ~ E0 + ((dose * Emax)/(xmid + dose))
  Data: datafos
       AIC      BIC    logLik
  261.1667 289.4941 -121.5833

Coefficients:
                        Value Std.Error   t value p value
Emax                 2.669646 0.1002376  26.63318  0.0000
E0                   9.441580 0.0774812 121.85639  0.0000
xmid.(Intercept)     2.285470 0.6196621   3.68825  0.0003
xmid.compoundJnJa    3.572604 1.6595000   2.15282  0.0328
xmid.compoundJnJb   -1.971663 0.6109348  -3.22729  0.0015
xmid.compoundJnJc   -1.744222 0.6112304  -2.85363  0.0049
xmid.compoundCompB -1.686293 0.6101441  -2.76376  0.0064
xmid.compoundCompC -2.215609 0.6161895  -3.59566  0.0004
```

According to the AIC criterion, the model with the best goodness of fit is the model with compound-specific parameter while the BIC criterion advocates the model with the common E_0. For the gene expression experiment, we prefer the later model since, as we argue above, E_0 is the baseline gene expression at dose zero and difference between compounds at this dose is only due to chance. The difference between the model with common E_0 (`fos.Com1`) and the model with common E_0 and $E_{\max}$ (`fos.Com2`) is shown in Fig. 14.5. Note that the model with common $E_{\max}$ does not imply that at the highest dose level the predicted values of all compounds are equal since $E_{\max}$ is the asymptote. This can be seen clearly in the upper panels of Fig. 14.5.

```
> anova(fos.Com0, fos.Com1 , fos.Com2, fos.Com3)

         Model df      AIC      BIC     logLik   Test   L.Ratio p value
fos.Com0     1  4 384.7557 397.3457 -188.37786
fos.Com1     2 14 236.5949 280.6598 -104.29743 1 vs 2 168.16088  <.0001
fos.Com2     3  9 261.1667 289.4941 -121.58335 2 vs 3  34.57184  <.0001
fos.Com3     4 19 232.4913 292.2937  -97.24564 3 vs 4  48.67540  <.0001
```

Let us focus on two compounds: CompA and JnJa. Figure 14.5 shows that the dose-response curve of CompA is above the dose-response curve of JnJa. This implies that gene FOS responds faster to CompA as compared with JnJa, and as a result, the ED_{50} of the gene for CompA is smaller than the ED_{50} of the gene for the compound JnJa as can be seen in Fig. 14.6.

14.4 Ranking Genes for a Single Compound Based on the ED_{50}

The model averaging methodology discussed above allows us to rank the response of genes to the increasing (decreasing) doses of a specific compound (CompA) based on model average estimate for the ED_{50}. Let $\hat{\theta}_{2_{g,r}}$ be the ED_{50} estimate for

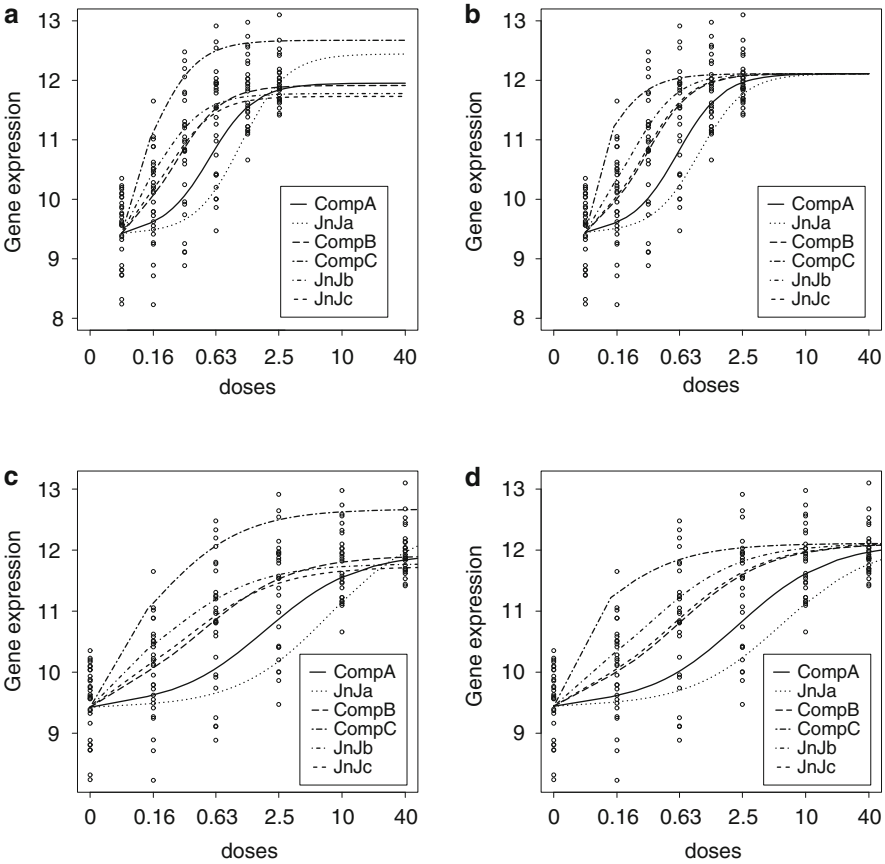

Fig. 14.5 Data and predicted models for the six compounds. *Upper panel*: models are presented outside the range of the dose. *Lower panel*: the models are presented in the range of the dose used in the experiment. (**a** and **c**) Common E_0; (**b** and **d**) common E_0 and $E_{\max}$

gene g and the rth model. As discussed in Sect. 13.2, the gene-specific model average estimate is given by

$$\hat{\theta}_{2_g} = \sum_{r=1}^{R} w_r(R)\hat{\theta}_{2_{g,r}},$$

where $w_r(R)$ is the Akaike weight (the posterior probability of the rth model) given in (10.7). Figure 14.7a, b shows the sorted values of the model average estimates for the ED_{50} for all significant genes for CompA with their corresponding 95% bootstrap confidence intervals.

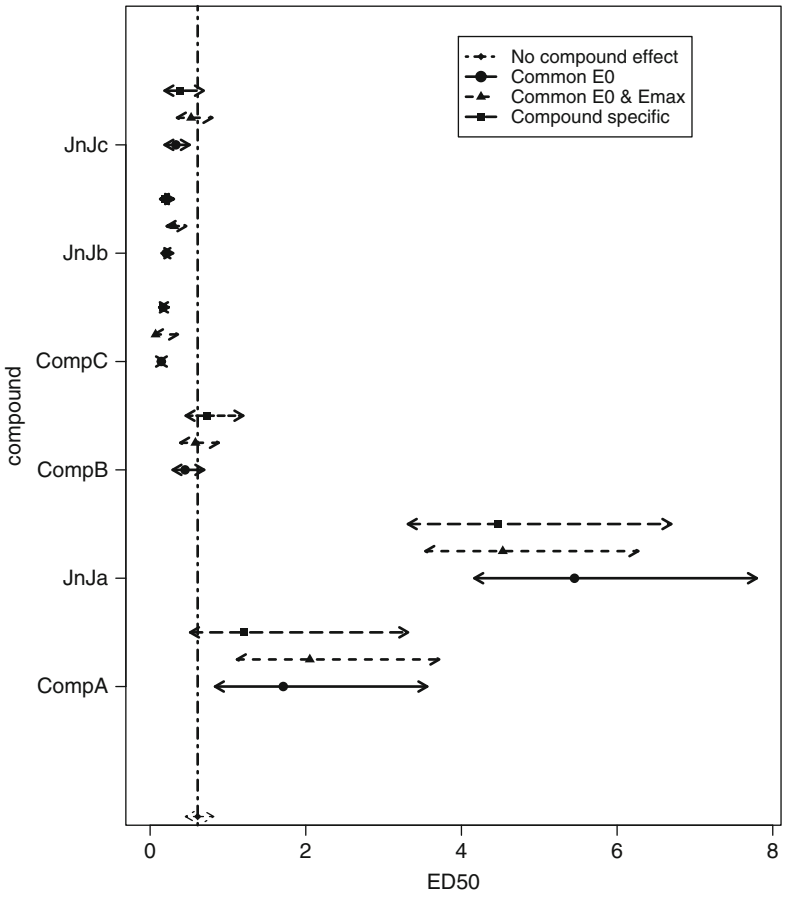

Fig. 14.6 Parameter estimates of the ED_{50} and 90% confidence intervals for each compound. *Vertical line*: $\widehat{ED}_{50}$ obtained from the model without compound effect

Figure 14.7c, d presents the data and the model average estimate for the dose-response curve for two genes (6792 and 83) with their ED_{50} equal to 1.143 and 4.96, respectively. Two main patterns can be observed in Fig. 14.7. First, the dose response curve for the gene 6792 indeed increases sharply compared to gene 83 and as a result the parameter estimate for the ED_{50} for gene 6792 is smaller as compared to gene 83. Second, the fold change of the gene 6792 is higher than the fold change of the gene 83. This implies that the gene 6792 responds to the increasing doses of CompA quicker (smaller value of ED_{50}) and with a higher increment of gene expression (a higher fold change).

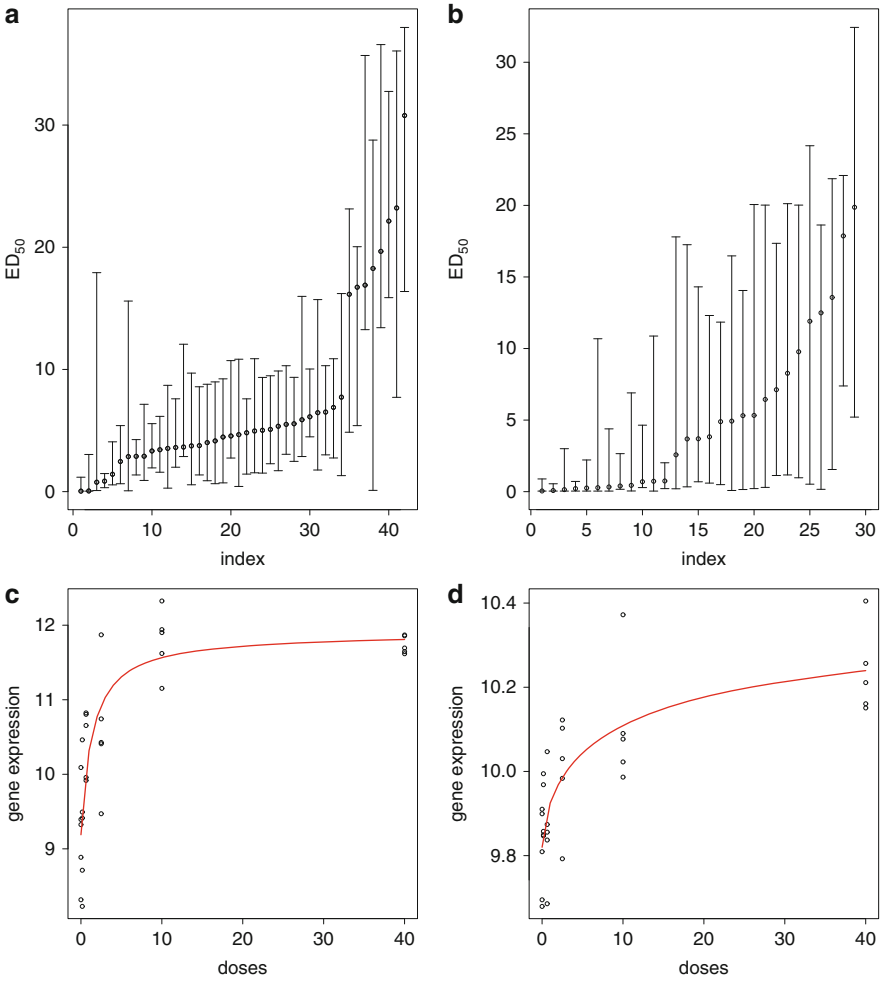

Fig. 14.7 Parameter estimates of the ED_{50} for CompA. *Upper panels*: model average estimates and 95% CI. *Lower panel*: data and model average of the dose-response curve for the genes 6792 and 83. (**a**) ED_{50} and 95% CI for genes with upward trend. (**b**) ED_{50} and 95% CI for genes with downward trend. (**c**) Gene 6792 ($ED_{50} = 1.143$). (**d**) Gene 83 ($ED_{50} = 4.96$)

14.5 Comparison Between Several Compounds Based on ED_{50} Ranking

In the previous section, we compare between the response of genes for a single compound dose-response experiment. In some cases, a multi-compound dose-response microarray experiment is conducted in order to evaluate a new compound against a known compound. In this setting, the comparison of primary interest is the

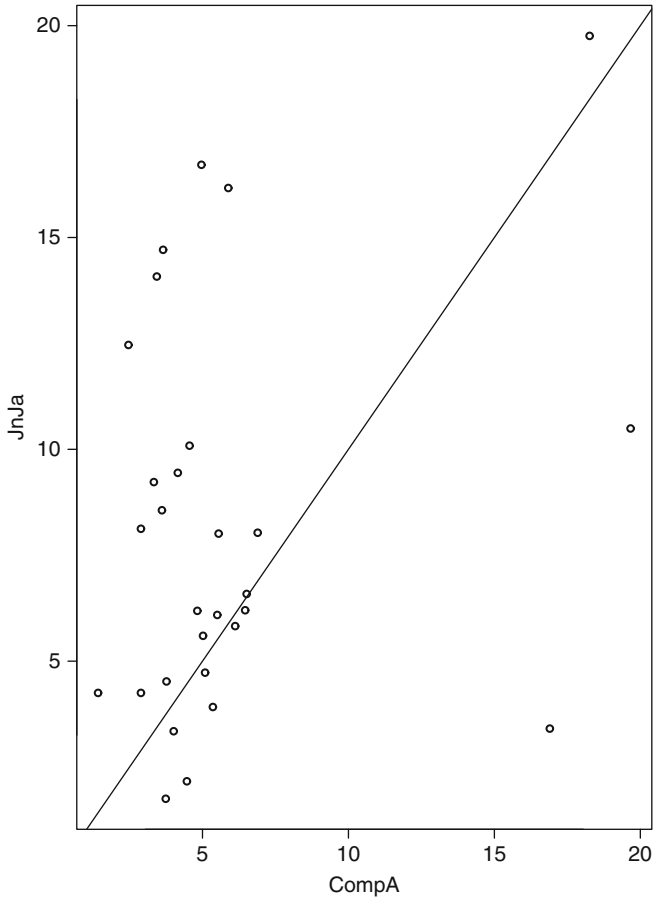

Fig. 14.8 Model average estimate of the ED_{50} for significant genes under two compounds: CompA and JnJa

change in gene expression across compounds for several genes. In this section, we illustrate the use of a model averaging technique in order to compare the change in gene expression between two compounds: CompA and JnJa. Figure 14.8 shows the model average estimates for the ED_{50} obtained for CompA and JnJa. We notice that the majority of the genes which were found to be significant for the two compounds respond faster to CompA, as compared to JnJa, i.e., their model average estimates for ED_{50} are higher under JnJa compared to the corresponding model average estimates for the ED_{50} under CompA. Figure 14.9 shows an example of two genes. The first gene (6792) responds quicker under CompA compared to the new compound (model average estimates of the ED_{50} are equal to 1.143 and 4.25 under CompA and JnJa, respectively). The same pattern can be observed for the second gene (83) with model average estimate of the ED_{50} equal to 4.96 and 16.92 under CompA and JnJa, respectively.

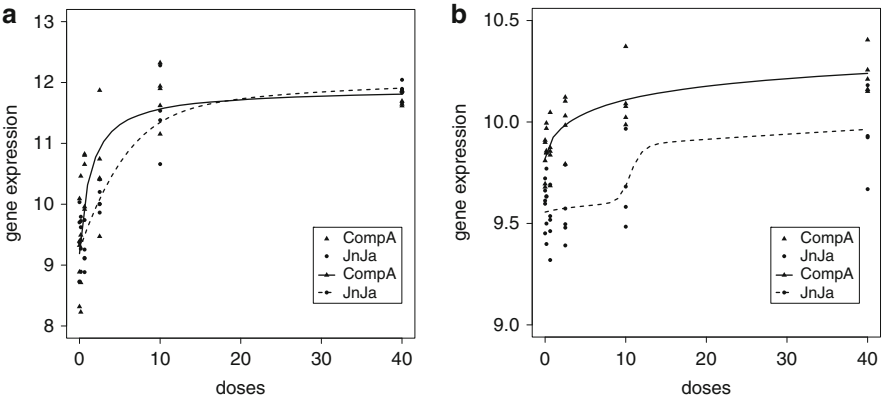

Fig. 14.9 Data and model average for the dose-response curves of two genes under CompA (*solid line*) and JnJa (*dashed line*). (**a**) Gene 6792 ($ED_{50} = 1.143$). (**b**) Gene 83 ($ED_{50} = 4.96$)

14.6 Discussion

The aim of the analysis presented in this chapter was not to detect differentially expressed genes with respect to an increasing dose, but to investigate in more details the response of genes which were found to have a significant dose-response relationship. In contrast with the previous chapter in which we assume that the dose-response relationship is a step function, in this chapter, we used parametric models in order to estimate the dose-response curve. This allows us to estimate the ED_{50} parameter for each gene and to compare the response of several genes for a single compound or several compounds. Such comparisons are of primary interest when the researcher would like to investigate the response of target genes to a certain compound to a certain compound or to compare the response of a gene list between compounds. In this chapter, we used five parametric dose-response models for illustration. The R package `DoseFinding` allows to use other parametric models according to the preference of the researcher.

References

Bornkamp, B., Pinheiro, J. C., & Bretz, F. (2009). MCPMod: an R package for the design and analysis of dose-finding studies. *Journal of Statistical Software, 29*(7).

Bornkamp, B., Pinheiro, J. C., & Bretz, F. (2012). Reference manual for the R Package `DoseFinding`. http://cran.r-project.org/web/packages/DoseFinding/index.html

Bretz, F., Pinheiro, J. C., & Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics, 61*, 738–748.

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging.* Cambridge Series in Statistical and Probabilistic Mathematics.

Pinheiro J.C., Bornkamp B., & Bretz F., (2006a). Design and Analysis of dose finding studies combining multiple comparisons and modeling procedures. *Journal of Biopharmaceutical Statistics, 6*, 639–656.

Pinheiro, J. C., Bretz, F., & Branson, M. (2006b). Analysis of dose-response studies—Modeling approaches. In N. Ting (Ed.), *Dose finding in drug development* (pp. 146–171). New York: Springer.

Whitney, M., & Ryan, L. (2009). Quantifying dose-response uncertainty using bayesian model averaging. In Cooke, R. M. (Ed.), *Uncertainty modeling in dose-response*. New York: Wiley.

Chapter 15
Multiple Contrast Tests for Testing Dose–Response Relationships Under Order-Restricted Alternatives

Dan Lin, Ludwig A. Hothorn, Gemechis D. Djira, and Frank Bretz

15.1 Introduction

In Chaps. 3, 7, and 8, we discussed five test statistics that can be used for testing the null hypothesis of homogeneity of means against order-restricted alternatives. A rejection of the null hypothesis implies a significant monotone trend of gene expression with respect to dose. In the case study of epidermal carcinoma cell line data in Sect. 10.1, we showed that 3,499 genes were found to be significant when the LRT (the $\bar{E}_{01}^2$ test statistic) was used to test the null hypothesis of no dose effect against the order-restricted alternatives and the BH-FDR procedure was used to control the FDR. Among the significant genes, 1,600 genes exhibited increasing trends and 1,899 genes showed decreasing trends. In this chapter, we employ an alternative method to find genes with monotonic trends, namely, the multiple contrast test (MCT).

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

L.A. Hothorn
Institute of Biostatistics, Leibniz University Hannover, Hannover, Germany
e-mail: hothorn@biostat.uni-hannover.de

G.D. Djira
Department of Mathematics and Statistics, South Dakota State University, Brookings, SD, USA
e-mail: Gemechis.Djira@sdstate.edu

F. Bretz
Integrated Information Sciences, Novartis Pharma AG, Novartis Campus, Basel, Switzerland
e-mail: frank.bretz@novartis.com

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_15,

The four t-type test statistics (Williams', Marcus', M, and the modified M), which were considered in Chaps. 3 and 7, can be used to test the mean expression levels between the highest dose and the control in a step-down fashion. As an analogue, Bretz (2006) proposed the use of MCTs for Williams' and Marcus' tests, when the question of interest in the dose-response study is to find genes with significant monotonic trends.

The content of the chapter is organized as follows. In Sect. 15.2, we summarize the pool-adjacent-violator algorithm (PAVA) to obtain the order-restricted mean responses using isotonic regression and introduce the MCTs. In particular, Williams- and Marcus-type MCTs are discussed. In Sect. 15.3, we show the use of the R package `multcomp` to illustrate the MCTs for these contrasts. In Sect. 15.4, we discuss the topic of partial order alternative and show that umbrella alternatives can be defined and derive suitable MCTs. The human epidermal carcinoma cell line data are used as case study in this chapter.

15.2 Computation of Isotonic Means and MCTs

In this section, we formulate the test for order-restricted inference and introduce the expression of Williams' and Marcus' contrasts in terms of the MCT proposed by Bretz (1999, 2006).

15.2.1 PAVA for the Computation of Isotonic Means

In order to obtain the isotonic means, the most widely used technique is the PAVA (Ayer et al. 1955; Barlow et al. 1972; Robertson et al. 1988). If sample means for neighboring doses are not in a monotonic restricted order, the method non-parametrically amalgamates the means, until the amalgamated means are completely ordered. The result of the algorithm can be linked to the following analytical expression using max–min formulas. Given $n_0, n_1, \ldots, n_K$ observations (arrays) and sample means $\bar{y}_0, \bar{y}_1, \ldots \bar{y}_K$, at doses $d_0, d_1, \ldots, d_K$, respectively, and assuming normally distributed data, the maximum likelihood estimates $\hat{\mu}_i^\star$, subject to the simple order restriction $\mu(d_0) \leq \mu(d_1) \leq \cdots \leq \mu(d_K)$, are given by

$$\hat{\mu}_i^\star = \max_{0 \leq u \leq i} \min_{i \leq v \leq K} \frac{\sum_{i=u}^{v} n_i \bar{y}_i}{\sum_{i=u}^{v} n_i}, \quad i = 0, 1, \ldots, K, \tag{15.1}$$

where $\bar{y}_i = \sum_{j=0}^{n_i} y_{ij} / n_i$ is the sample mean for dose $i = 0, 1, \ldots, K$.

15.2.2 Single and MCTs

The standardized statistic of a single contrast test (SCT) is defined as,

$$T^{\mathrm{SC}} = \frac{\sum_{i=0}^{K} c_i \bar{y}_i}{s\sqrt{\sum_{i=0}^{K} \frac{c_i^2}{n_i}}}, \tag{15.2}$$

where the contrast vector $\boldsymbol{c}$ with the weights c_i as the contrast coefficients fulfilling the condition $\sum_{i=0}^{K} c_i = 0$. T^{SC} is univariate central t-distributed with $\nu = \sum_{i=0}^{K} (n_i - 1)$ degrees of freedom under H_0.

MCTs were first described by Mukerjee et al. (1986, 1987) in a general and thorough manner. The test can be used to test for trend under order restrictions. The main reason for developing such a test was to obtain a test with a similar power behavior as the LRT, but easier to use.

MCTs seek to locate several contrast vectors, as good as possible in the alternative space (Bretz 1999). The aim of this approach is therefore to cover most parts of the alternative space by choosing some selected vectors within this space and conduct the MCT with respect to this vector. The resulting test, T^{MC}, is the maximum over r of such SCTs T^{SC} defined in (15.2):

$$\boldsymbol{T}^{\mathrm{MC}} = \max\{T_1^{\mathrm{SC}}, T_2^{\mathrm{SC}} \ldots T_r^{\mathrm{SC}}\}, \tag{15.3}$$

where r is the number of SCT.

The contrast matrix for $\boldsymbol{T}^{\mathrm{MC}}$ can be written as

$$\boldsymbol{C}^{\mathrm{MC}} = \begin{pmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_r \end{pmatrix} = \begin{pmatrix} c_{10} & c_{11} & \ldots & c_{1K} \\ c_{20} & c_{21} & \ldots & c_{2K} \\ \vdots & & \vdots & \\ c_{r0} & c_{r1} & \ldots & c_{rK} \end{pmatrix}. \tag{15.4}$$

Each row of the contrast matrix $\boldsymbol{C}^{\mathrm{MC}}$ corresponds to a contrast vector $\boldsymbol{c}$ of a SCT.

Multiple contrasts can be defined for many different test statistics. Somerville (1997, 1999) and Hothorn (2006) provided lists of several multiple comparison procedures (not necessarily designed for order-restricted testing), which can be formulated as MCTs, such as Dunnett (1955, 1964) many-to-one approach, or Tukey's (1953) all-pairwise comparisons. According to Robertson et al. (1988), the global LRT statistic can be expressed as the maximum of an infinite number of contrast statistics as well.

In what follows, we discuss Williams- and Marcus-type MCTs (Bretz 1999). They are two established multiple comparison procedures, which can be represented as above as MCTs. The advantage of Williams' test is that it is a simple order alternative with respect to a control, whereas Marcus' test is a competitor to a general simple order alternative.

15.2.2.1 MCT for an Order Restriction Alternative

Under an order-restricted alternative, one possibility of constructing the MCTs is to decompose the alternative $H_1^{\text{Up}} : \mu(d_0) \leq \mu(d_1) \leq, \ldots, \leq \mu(d_K)$ into $2^K - 1$ sub-hypotheses (Bretz 1999). These sub-hypotheses constitute all the possible trends under the simple order alternative. For example, for the case of $K = 3$, the alternative hypothesis can be decomposed as

$$H_1^{\text{Up}} = \bigcup_{i=1}^{7} H_{1(i)}^{\text{Up}},$$

where

$$\begin{aligned}
H_{1(1)}^{\text{Up}} &: \mu(d_0) = \mu(d_1) = \mu(d_2) < \mu(d_3),\\
H_{1(2)}^{\text{Up}} &: \mu(d_0) < \mu(d_1) = \mu(d_2) = \mu(d_3),\\
H_{1(3)}^{\text{Up}} &: \mu(d_0) = \mu(d_1) < \mu(d_2) = \mu(d_3),\\
H_{1(4)}^{\text{Up}} &: \mu(d_0) < \mu(d_1) = \mu(d_2) < \mu(d_3),\\
H_{1(5)}^{\text{Up}} &: \mu(d_0) < \mu(d_1) < \mu(d_2) < \mu(d_3),\\
H_{1(6)}^{\text{Up}} &: \mu(d_0) = \mu(d_1) < \mu(d_2) < \mu(d_3),\\
H_{1(7)}^{\text{Up}} &: \mu(d_0) < \mu(d_1) < \mu(d_2) = \mu(d_3).
\end{aligned}$$

Every true dose-response relationship will fall into exactly one of the subalternatives when H_1^{Up} is true. Each sub-hypothesis can be tested with a SCT. The advantages of using the MCTs for an order-restricted alternative are twofolds: (1) rejecting any of these sub-hypotheses indicates a significant dose-response relationship, and (2) the maximum of MCTs will determine the configuration of the isotonic means.

Note that these seven subalternatives are identical to the seven possible dose-response curve shapes given in Table 9.1. In Chap. 10, we classified genes into one of seven models (from g_1 to g_7) based on the posterior model probability using different information criteria. In this chapter, by means of defining all possible subalternatives as a set of SCTs, the maximum test statistic value of MCT will find the "best" dose-response relationship. Therefore, a MCT can be an alternative approach to identify the dose-response curve shape based on a parametric test.

15.2.2.2 A Williams-type MCT

In Williams' procedure (Williams 1971, 1972), the amalgamated mean for the highest dose $\mu(d_K)$ can be expressed as follows (Bretz 1999):

$$\begin{aligned}
\hat{\mu}_K^{\star} &= \max_{1\le u\le K} \sum_{i=u}^{K} n_i \bar{y}_i \Big/ \sum_{i=u}^{K} n_i \\
&= \max\left\{ \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + \ldots n_K\bar{y}_K}{n_1 + n_2 + \ldots n_K}, \ldots, \frac{n_{K-1}\bar{y}_{K-1} + n_K\bar{y}_K}{n_{K-1} + N_K}, \bar{y}_K \right\} \\
&= \max\left\{ \begin{pmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & \frac{n_{K-1}}{n_{K-1}+n_K} & \frac{n_K}{n_{K-1}+n_K} \\ \vdots & \ldots & \vdots & \vdots \\ \frac{n_1}{n_1+\ldots+n_K} & \cdots & \frac{n_{K-1}}{n_1+\ldots+n_K} & \frac{n_K}{n_1+\ldots+n_K} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_K \end{pmatrix} \right\} \\
&= \max \boldsymbol{C}\, \bar{\boldsymbol{y}}_{(-0)},
\end{aligned}$$

where $\bar{\boldsymbol{y}}_{(-0)} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_K)'$. Note that the arrays of the control group are not included in the amalgamation process. In Williams' test, we have

$$\begin{aligned}
\hat{\mu}_K^{\star} - \bar{y}_0 &= \max\left\{ \boldsymbol{C}\, \bar{\boldsymbol{y}}_{(-0)} - \bar{y}_0 \mathbf{1} \right\} \\
&= \max\left\{ [-\mathbf{1} + \boldsymbol{C}]\bar{\boldsymbol{y}} \right\} \\
&= \max \boldsymbol{C}^{\mathrm{Wil}} \bar{\boldsymbol{y}},
\end{aligned}$$

where $\bar{\boldsymbol{y}} = (\bar{y}_0, \bar{\boldsymbol{y}}_{(-0)})'$.

Hence, the Williams-type MCT matrix $\boldsymbol{C}^{\mathrm{Wil}}$ can be expressed as

$$\boldsymbol{C}^{\mathrm{Wil}} = \begin{pmatrix} -1 & 0 & \ldots & 0 & 1 \\ -1 & 0 & \ldots & \frac{n_{K-1}}{n_{K-1}+n_K} & \frac{n_K}{n_{K-1}+n_K} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ -1 & \frac{n_1}{n_1+\ldots+n_K} & \cdots & \frac{n_{K-1}}{n_1+\ldots+n_K} & \frac{n_K}{n_1+\ldots+n_K} \end{pmatrix}. \tag{15.5}$$

The maximum contrast (15.4) consists of comparisons of the control with the weighted average over the last i treatments ($i = 1, \ldots, K$). The contrast matrix $\boldsymbol{C}^{\mathrm{MC}}$ (15.4) for Williams' test is given by (15.5).

Williams' MCT takes the order restriction of the means into account through the contrast definition following the construction of the isotonic mean estimates

[in (15.1)]. As pointed out by Bretz (1999), Williams' t-type test statistic is not identical to Williams' MCT because they have different variance estimators. The variance in the MCT is the completely studentized statistic by making use of the mean square error, i.e.,

$$\sqrt{\sum_{i=0}^{K} 1/c_i \times \sum_{i=0}^{K}\sum_{j=0}^{n_i}(y_{ij}-\bar{y}_i)^2/\nu},$$

while Williams' and Marcus' tests used the mean square error for two sample t-test with

$$\sum_{i=0}^{K}\sum_{j=0}^{n_i}(y_{ij}-\bar{y}_i)^2/\nu(1/(n_0)+1/(n_K)).$$

15.2.2.3 The Marcus-type MCT

Marcus-type MCT can be derived in a similar way to the Williams-type MCT (Bretz 1999, 2006). The amalgamated mean for the control dose $\mu(d_0)$ is given by

$$\begin{aligned}\hat{\mu}_0^{\star} &= \min_{0\le v\le K}\sum_{i=0}^{v} n_i\bar{y}_i/\sum_{i=0}^{v} n_i \\ &= \min\left\{\bar{y}_0, \frac{n_0\bar{y}_0+n_1\bar{y}_1}{n_0+n_1},\ldots,\frac{n_0\bar{y}_0+n_1\bar{y}_1+\cdots+n_K\bar{y}_K}{n_0+n_1+\cdots+n_K}\right\}.\end{aligned}$$

Therefore,

$$\hat{\mu}_K^{\star}-\hat{\mu}_0^{\star} = \max\left\{0, \max_{0\le i,j\le K}\left\{\frac{n_j\bar{y}_j+\cdots+n_k\bar{y}_K}{n_j+\cdots+n_K}-\frac{n_0\bar{y}_0+\cdots+n_i\bar{y}_i}{n_0+\cdots+n_i}\right\}\right\}. \tag{15.6}$$

The difference $\hat{\mu}_K^{\star}-\hat{\mu}_0^{\star}$ can be represented as a simple maximum term in (15.6). A natural way of applying the MCT principle to Marcus' (1976) approach is to identify each element of (15.6) as a contrast. However, a closed form expression for the resulting contrast matrix $\boldsymbol{C}^{\mathrm{MC}}$ for Marcus' test depends on the number of dose levels.

15.2.3 Distribution of the MCT

Assuming the normality of the response, the joint distribution of $T_1,\ldots,T_r$ is a central r-variate t-distribution with ν degrees of freedom and correlation matrix $\boldsymbol{R}=\{\rho_{l,m}\}_{l,m}$, $l,m=1,\ldots,r$ (Bretz 2006). The entries of $\boldsymbol{R}$ consist of the

correlation between each two of the r estimated contrasts, say $\boldsymbol{c}_l$ and $\boldsymbol{c}_m$ under H_0

$$\rho_{lm} = \frac{\sum_{i=0}^{K} c_{li} c_{mi} / n_i}{\sqrt{(\sum_{i=0}^{K} c_{li}^2 / n_i)(\sum_{i=0}^{K} c_{mi}^2 / n_i)}},$$

assuming that the errors are normal distributed with homogeneous group-specific variances. Thus, the inference of MCTs can be based on the multivariate t-distribution (Genz and Bretz 2009).

15.3 Use of the `multcomp` Package to Case Study Data

In this section, we use the `multcomp` package (Bretz et al. 2010) to illustrate the analysis for the human epidermal carcinoma cell line data. To automatically obtain Williams' and Marcus' contrast coefficients from the package, the dose levels are required.

```
> library(multcomp)
> x.res=as.factor(c(rep(1,3),rep(2,3),rep(3,3),rep(4,3))) ##dose levels
> n <- table(x.res)
```

In the function `contrMat()`, the options to create the multiple comparisons can be seen below:

```
contrMat(n, type = c("Dunnett", "Tukey", "Sequen", "AVE",  "Changepoint",
"Williams", "Marcus", "McDermott", "UmbrellaWilliams", "GrandMean"),
base = 1)
```

where n is a vector containing the sample sizes under each dose group and `base` specifies the control group for Dunnett's test.

```
> conWilliams <- contrMat(n, type = "Williams")
> conMarcus <- contrMat(n, type = "Marcus")
> conWilliams
        Multiple Comparisons of Means: Williams Contrasts
     1      2      3      4
C 1 -1 0.0000 0.0000 1.0000
C 2 -1 0.0000 0.5000 0.5000
C 3 -1 0.3333 0.3333 0.3333
> conMarcus
        Multiple Comparisons of Means: Marcus Contrasts
          1       2       3      4
C 1 -1.0000  0.3333  0.3333 0.3333
C 2 -1.0000  0.0000  0.5000 0.5000
C 3 -0.5000 -0.5000  0.5000 0.5000
C 4 -1.0000  0.0000  0.0000 1.0000
C 5 -0.5000 -0.5000  0.0000 1.0000
C 6 -0.3333 -0.3333 -0.3333 1.0000
```

It is easy to note that Williams' contrasts are all contained in Marcus' contrasts (the fourth, second, and first rows). These are suitable for concave dose-response shapes, as the higher dose groups are being pooled and compared to the control. Two of Marcus' contrasts (given by the fifth and sixth rows) seem to be suitable to detect convex shapes, as they take the average over the lower doses.

For each gene, we first fit the ANOVA model with the dose as factor and then use the function `glht` with the contrasts obtained above to obtain the MCT. The general form of the function `glht()` is given by

```
glht(model, linfct,  alternative = c("two.sided", "less", "greater"), ...)
```

where `linfct` needs to be specified as the matrix of linear functions, which can be a matrix of contrast coefficients, a symbolic description of linear hypothesis, or multiple comparisons of means obtained from the `mcp()` function. The function `mcp()` can be used to create multiple comparisons defined in the types of the function `contrMat()`.

The example of using this function to obtain Williams-type MCT is shown below for the first gene:

```
> amod <- aov(gene1 ~ x.res)
> summary(glht(amod, linfct = mcp(x.res = "Williams")))
         Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Williams Contrasts
Fit: aov(formula = gene1 ~ x.res)
Linear Hypotheses:
         Estimate Std. Error t value p value
C 1 == 0  0.20915    0.20108   1.040   0.474
C 2 == 0  0.08395    0.17414   0.482   0.835
C 3 == 0 -0.01657    0.16418  -0.101   0.997
(Adjusted p values reported -- single-step method)
```

As expected, the MCTs with Williams' contrasts yield estimates of three contrasts with their standard error, test statistics, and p values. These p values are obtained based on a multivariate t-distribution. Therefore, the multiplicity adjustment is taken into account within the gene.

Similarly, Marcus' MCTs can be obtained by specifying the corresponding Marcus' contrasts. Since the correlation structure between the test statistics has changed from Williams' three contrasts to Marcus' six contrasts, the p values for the common three contrasts (C1, C2, and C4) have changed accordingly. For this gene, the null hypothesis of homogeneity of means in gene expression over doses is not rejected based on both Williams- and Marcus-type MCTs.

```
> summary(glht(amod, linfct = mcp(x.res = "Marcus")))
         Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Marcus Contrasts
Fit: aov(formula = gene1 ~ x.res)
Linear Hypotheses:
         Estimate Std. Error t value p value
C 1 == 0 -0.01657    0.16418  -0.101   0.999
C 2 == 0  0.08395    0.17414   0.482   0.930
C 3 == 0  0.19274    0.14218   1.356   0.446
C 4 == 0  0.20915    0.20108   1.040   0.626
C 5 == 0  0.31794    0.17414   1.826   0.245
C 6 == 0  0.29543    0.16418   1.799   0.254
(Adjusted p values reported -- single-step method)
```

For analysis of the whole case study data, the following code can be used to record all the p values of the MCT of all the genes:

```
> p.wil <-  matrix(NA, nrow(datax), nrow(conWilliams))
> p.mar  <- matrix(NA, nrow(datax), nrow(conMarcus))
> for (i in 1: nrow(data)) {
> genei <- as.numeric(data[i,])
> amod <- aov(genei ~ x.res)
> p.wil[i,] <- summary(glht(amod, linfct = mcp(x.res = "Williams")))
$test$pvalues
> p.mar[i,] <-summary(glht(amod, linfct = mcp(x.res = "Marcus")))
$test$pvalues
> cat(i)
> }
> minp.wil <- apply(p.wil, 1, min)
> minp.mar <- apply(p.mar, 1, min)
> adjp.wil <- p.adjust(minp.wil, method="BH")
> adjp.mar <- p.adjust(minp.mar, method="BH")
> sum(adjp.wil<=0.05)
> 2880
> sum(adjp.mar<=0.05)
> 3303
```

After adjusting for multiplicity using the BH-FDR procedure, the numbers of significant genes that result from using Williams- and Marcus-type MCTs are 2,880 and 3,303, respectively. These numbers are slightly smaller than the Williams' and Marcus' tests presented in Chap. 7. The plot of adjusted p values is given in Fig. 15.1.

15.4 Beyond the Simple Order Alternative

15.4.1 Dunnett's Test for a Simple Tree Alternative

The MCTs discussed in the previous sections assume a simple order alternative. In this section, we discuss two other ordered alternatives: the simple tree alternative and the unimodal partial order alternative (umbrella alternative). The latter was discussed in Chap. 11 where we discussed the ORIOGEN and ORICC algorithms for inference and clustering of order-restricted gene profiles. The simple tree alternative $\mu(d_0) \leq [\mu(d_1), \ldots, \mu(d_K)]$ can be used if the primary interest of the analysis is a comparison between dose zero and all other dose levels, $\mu(d_0) \leq \mu(d_i)$, $i = 1, \ldots, K$, but we do not specify any order restrictions among $\mu(d_1), \ldots, \mu(d_K)$. The contrast matrix for Dunnett's test can be constructed using the option `type = "Dunnett"`. We notice that, since there are no order restrictions among $\mu(d_1), \ldots, \mu(d_K)$, the contrast matrix for Dunnett's test implies that at each dose level the numerator of the test statistic is equal to $\bar{y}_i - \bar{y}_0$, while for Williams's test the numerator is equal to $\hat{\mu}_i^* - \bar{y}_0$.

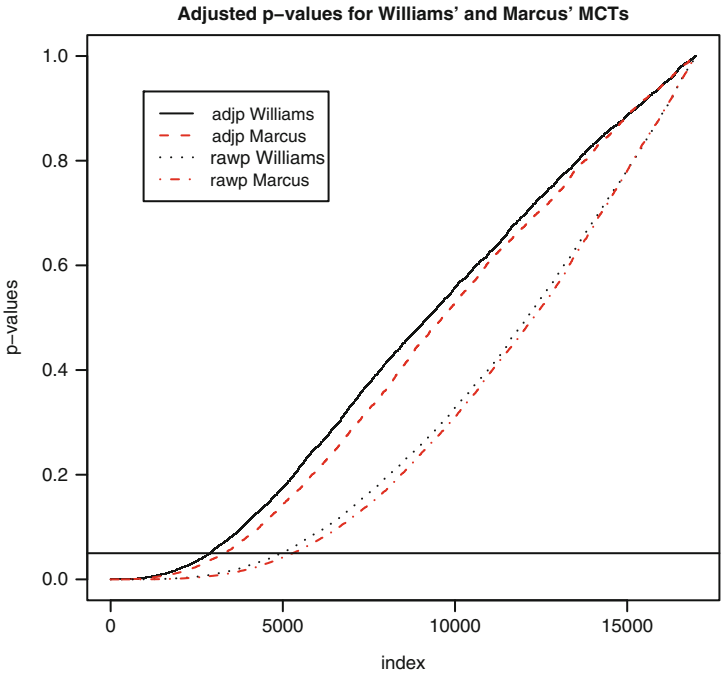

Fig. 15.1 Raw and adjusted p values for Williams- and Marcus-type MCTs

```
> conDunnett <- contrMat(n, type = "Dunnett")
> conDunnett

          Multiple Comparisons of Means: Dunnett Contrasts

       1 2 3 4
2 - 1 -1 1 0 0
3 - 1 -1 0 1 0
4 - 1 -1 0 0 1


> conWilliams
          Multiple Comparisons of Means: Williams Contrasts
     1      2      3      4
C 1 -1 0.0000 0.0000 1.0000
C 2 -1 0.0000 0.5000 0.5000
C 3 -1 0.3333 0.3333 0.3333
```

Dunnett's test can be performed in the `multcomp` package using the option `mcp(x.res "Dunnett")`. The panel below shows the test statistics for both Dunnett's and Williams' tests and, as expected, except for the last dose level, the test statistics are different:

```
>
> #Dunnett
> amod <- aov(gene2~x.res)
> summary(glht(amod, linfct = mcp(x.res = "Dunnett")))
Multiple Comparisons of Means: Dunnett Contrasts
Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0 -0.14830    0.08123  -1.826   0.2340
3 - 1 == 0  0.29035    0.08123   3.574   0.0181 *
4 - 1 == 0  2.31433    0.08123  28.490   <0.001 ***
---

> #Williams
> summary(glht(amod, linfct = mcp(x.res = "Williams")))
Multiple Comparisons of Means: Williams Contrasts
Linear Hypotheses:
         Estimate Std. Error t value Pr(>|t|)
C 1 == 0  2.31433    0.08123   28.49   <1e-06 ***
C 2 == 0  1.30234    0.07035   18.51   <1e-06 ***
C 3 == 0  0.81880    0.06633   12.35   <1e-06 ***
---
```

15.4.2 Umbrella Alternatives

As we showed in Chap. 11, an umbrella profile has the form

$$C_r = \left\{\mu \in R^{K+1} \ : \ \mu_0 \leq \mu_1 \leq \cdots \leq \mu_h \geq \cdots \geq \mu_K\right\}.$$

The umbrella point h is usually unknown. Bretz and Hothorn (2001, 2003) investigated MCTs for testing non-monotone order-restricted alternatives. The basic idea is that (for an up–down umbrella profile) the primary interest is not to investigate the (possibly monotone) decline of the response after dose h but to assess the trend up to dose h (Bretz and Hothorn 2003). Hence, for each possible value of h, Bretz and Hothorn (2003) define the following alternative:

$$H_1 : \mu_0 \leq \mu_1 \leq \cdots \leq \mu_h,$$

with at least $\mu_0 < \mu_h$. For the case study of human epidermal carcinoma data, there are four dose levels, and therefore, the umbrella point can be located at the first or the second dose levels (with the control as zero dose). The contrast matrix for a Williams-type umbrella alternative can be constructed using the option `type="UmbrellaWilliams"`. Note that the first three rows ($C_1 - C_3$) of the contrast matrix below are identical to the contrasts defined for Williams' MCT for simple order alternative. The three contrasts $C_4 - C_6$ correspond to an alternative with umbrella point at the second and third dose levels. For example, Fig. 15.2 illustrates a pattern in which there is a monotone trend up the third dose level. This pattern corresponds to contract C_5. Note that the mean gene expression in the last

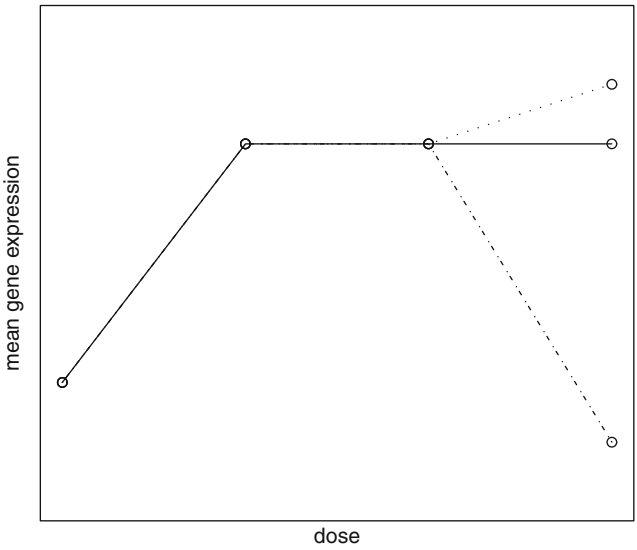

Fig. 15.2 Illustrative example of umbrella alternative. Possible mean patterns for a C_5 profile. *Solid line*: $\mu_0 < \mu_1 = \mu_2 = \mu_3$. *dotted line*: $\mu_0 < \mu_1 = \mu_2 < \mu_3$. *Dotted-dashed line* (an umbrella pattern with umbrella point at the third dose level): $\mu_0 < \mu_1 = \mu_2 > \mu_3$

dose level is not restricted, i.e., all relationships $\mu_2 = \mu_3$, $\mu_2 < \mu_3$, and $\mu_2 > \mu_3$ are possible.

```
> conUmbrellaWilliams <- contrMat(n, type = "UmbrellaWilliams")
> conUmbrellaWilliams

          Multiple Comparisons of Means: UmbrellaWilliams Contrasts

     1      2      3      4
C 1 -1 0.0000 0.0000 1.0000
C 2 -1 0.0000 0.5000 0.5000
C 3 -1 0.3333 0.3333 0.3333
C 4 -1 0.0000 1.0000 0.0000
C 5 -1 0.5000 0.5000 0.0000
C 6 -1 1.0000 0.0000 0.0000
```

Figure 15.3 shows an illustrative example for one gene. Note that the only difference (in terms of gene expression) between panel a and b is in the last dose level for which the gene expression in panel b shifted up.

For the expression levels presented in Fig. 15.3a, the test statistic for the maximum contrast is equal to 8.398 ($p < 0.0001$), indicating that the null hypothesis is rejected and that the most likely dose-response relationship is pattern C_5, $\mu_0 < \mu_1 = \mu_2$. For the expression levels presented in Fig. 15.3b, the test statistic for the maximum contrast test is equal to 9.041 ($p < 0.001$), indicating that the most likely profile is C_3, $\mu_0 < \mu_1 = \mu_2 = \mu_3$.

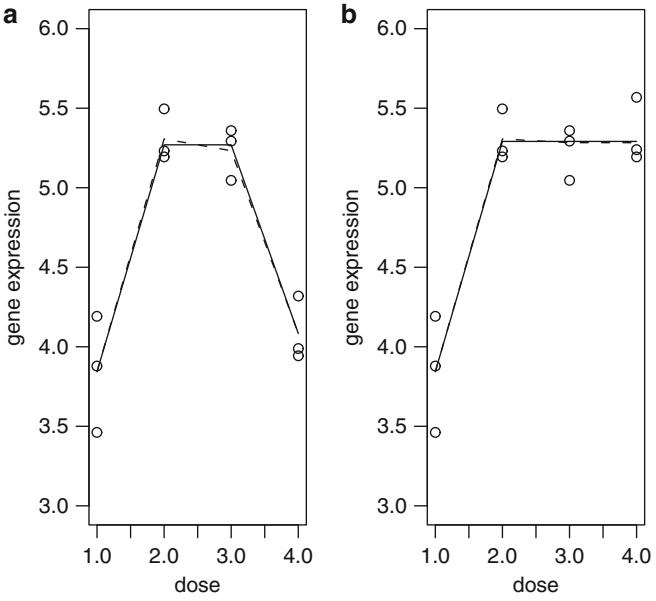

Fig. 15.3 Umbrella alternative. Illustrative example of one gene. *Dashed line*: ordered-restricted mean for up–down umbrella profiles with maximum at the second dose level. *Solid line*: ordered-restricted mean for up–down umbrella profiles with maximum at the third dose level

```
> ##### PANEL a  #####
> amod <- aov(gene1 ~ x.res)
> summary(glht(amod, linfct = mcp(x.res = "UmbrellaWilliams")))
Multiple Comparisons of Means: UmbrellaWilliams Contrasts

Linear Hypotheses:
         Estimate Std. Error t value Pr(>|t|)
C 1 == 0   0.2400     0.1961   1.224  0.51867
C 2 == 0   0.8144     0.1698   4.797  0.00395 **
C 3 == 0   1.0307     0.1601   6.438  < 0.001 ***
C 4 == 0   1.3889     0.1961   7.084  < 0.001 ***
C 5 == 0   1.4260     0.1698   8.398  < 0.001 ***
C 6 == 0   1.4632     0.1961   7.463  < 0.001 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
(Adjusted p values reported -- single-step method)

>#####  PANEL b  #####
> amod <- aov(gene1 ~ x.res)
> summary(glht(amod, linfct = mcp(x.res = "UmbrellaWilliams")))
Multiple Comparisons of Means: UmbrellaWilliams Contrasts

Linear Hypotheses:
         Estimate Std. Error t value Pr(>|t|)
C 1 == 0   1.4900     0.1961   7.599   <0.001 ***
C 2 == 0   1.4394     0.1698   8.477   <0.001 ***
C 3 == 0   1.4474     0.1601   9.041   <0.001 ***
C 4 == 0   1.3889     0.1961   7.084   <0.001 ***
C 5 == 0   1.4260     0.1698   8.398   <0.001 ***
C 6 == 0   1.4632     0.1961   7.463   <0.001 ***
---
```

15.5 Discussion

In this chapter, we have shown a representation of Marcus' test, i.e., the mean difference of gene expression level at the highest dose and the control under an order restriction by means of multiple contrasts, namely, the Marcus-type MCT. In order to answer the testing problem defined in Sect. 15.2.2.1, we transform the trend tests of Williams and Marcus into the corresponding Williams-type and Marcus-type multiple contrasts tests (Bretz 1999, 2006). This method efficiently utilizes the multivariate t-distribution as the basis for the inference on the MCTs. Therefore, the FWER is controlled within the gene in this case. At the same time, we have also applied the BH-FDR procedure across genes to control the overall type I error rate.

This approach of combining two types of multiple testing adjustments (i.e., the FWER/gene and the FDR/genes) is one option in controlling the type I error in this setting of ordered-restricted inference using the MCT. This two-stage approach has the advantage of easy implementation and a clear interpretation. We can also consider the approache of the FWER/gene and FWER/genes to compare their powers and controls of the error rate. However, this is beyond the scope of this book. We propose a simulation study to further investigate this topic.

We have compared this approach to the results of the likelihood ratio test, Williams' and Marcus' tests discussed in Chap. 7. Somehow, the number of significant findings is slightly smaller than that detected using Williams' and Marcus' tests. They differ in the standard error and estimation method using inference based on permutations or exact distribution. Inference and clustering for partial order alternative were discussed previously in Chap. 11. In this chapter, we address this topic within the MCT framework and we discussed MCTs for both simple tree alternative (using MCT for Dunnett's test) and umbrella Alternatives (using Williams-type umbrella MCT).

References

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics, 26*, 641–647.

Barlow, R. E., Bartholomew, D. J., Bremner, M. J., & Brunk, H. D. (1972). *Statistical inference under order restriction*. New York: Wiley.

Bretz, F. (1999). *Powerful modifications of Williams' test on trend*. Ph.D. dissertation. Vom Fachbereich Gartenbau der Universität Hannover, Hannover, http://www.biostat.uni-hannover.de/fileadmin/institut/pdf/thesis_bretz.pdf.

Bretz, F., & Hothorn, L. A. (2001). Testing dose-response relationship with a priori unknown, possibly non monotone shapes. *Journal of Biopharmaceutical Statistics, 11*, 193–207.

Bretz, F., & Hothorn, L. A. (2003). Statistical analysis of monotone and non-monotone dose-response data from in vitro toxicological assays. *Alternatives to Laboratory Animals (ALTA)* (Suppl. 1), *31*, 81–90.

Bretz, F. (2006). An extension of the Williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis, 50*, 1735–1748.

Bretz, F., Hothorn, T., & Westfall P. (2010). *Multiple comparisons using R*. Boca Raton: CRC press.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Society (JASA), 50*, 1096–1121.

Dunnett, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics, 20*, 482–491.

Genz, A., & Bretz, F. (2009) Computation of multivariate normal and t probabilities. *Lecture Notes in Statistics, 195* (Springer).

Hothorn, L. A. (2006). Multiple comparisons and multiple contrasts in randomized dose-response trials—Confidence interval orient approaches. *Journal of Biopharmaceutical Statistics, 16*, 711–731.

Marcus, R. (1976). The powers of some tests of the quality of normal means against an ordered alternative. *Biometrika, 63*, 177–83.

Mukerjee, H., Roberston, T., & Wright, F. T. (1986). Multiple contrast tests for testing against a simple tree ordering. In R. Dykstra (Ed.), *Advances in order restricted statistical inference* (pp. 203–230). Berlin: Springer.

Mukerjee, H., Roberston, T., & Wright, F. T. (1987). Comparison of several treatments with a control using multiple contrasts. *Journal of the American Statistical Association, 82*, 902–910.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Somerville, P. (1997). Multiple testing and simultaneous confidence intervals: calculation of constants. *Computational Statistics & Data Analysis, 25*, 217–233.

Somerville, P. (1999). Critical values for multiple testing and comparisons: One step and step down procedures. *Journal of Statistical Planning and Inference, 82*(1), 129–138(10).

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript. In *The collected works of John W. Tukey VIII. Multiple comparisons, 1948–1983* (pp. 1–300). New York: Chapman & Hall.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics, 27*, 103–117.

Williams, D. A. (1972). The comparision of several dose levels with a zero dose control. *Biometrics, 28*, 519–531.

Chapter 16
Simultaneous Inferences for Ratio Parameters Using Multiple Contrasts Test

Dan Lin, Gemechis D. Djira, Ziv Shkedy, Tomasz Burzykowski, and Ludwig A. Hothorn

16.1 Introduction

As we mentioned in Sect. 10.1, for the case study of epidermal carcinoma cell line data, 3,499 genes were found to be significant when the LRT (the E_{01}^2 test statistic) was used to test the null hypothesis (3.2) against the order-restricted alternative (3.3) or (3.4) and the BH-FDR procedure was used to control the FDR. However, the rejection of the null hypothesis (3.2) does not indicate the magnitude by which gene expression increases or decreases. In this chapter, we wish to search for genes for which the mean gene expression increases by $100 \times \delta\%$ from the control for at least one (perhaps the highest) dose, where δ is a relative biological ratio of interest.

Genes can be tested for a prespecified δ and can be ranked based on the significance of the ratio test. Genes found significant with a large δ are of higher interest to the biologists. In the case study of human epidermal carcinoma cell line data, the experiment includes four EGF doses (three doses and a control) for the control compound and three arrays per dose level. Figure 16.1 shows four genes, for which $\hat{\mu}_K^\star/\hat{\mu}_0^\star = 2.1$ (panels a and b) and $\hat{\mu}_K^\star/\hat{\mu}_0^\star = 0.48$ (panels c and d). The

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

Z. Shkedy · T. Burzykowski
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Ziv.shkedy@uhasselt.be; Tomasz.burzykowski@uhasselt.be

G.D. Djira
Department of Mathematics and Statistics, South Dakota State University, Brookings, SD, USA
e-mail: Gemechis.Djira@sdstate.edu

L.A. Hothorn
Institute of Biostatistics, Leibniz University Hannover, Hannover, Germany
e-mail: hothorn@biostat.uni-hannover.de

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_16,

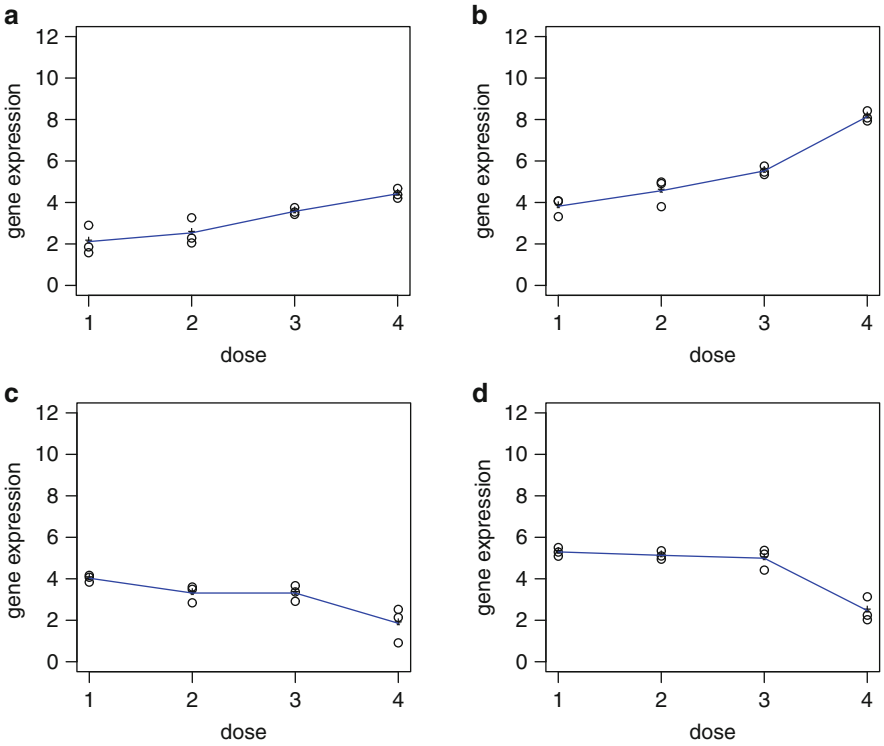

Fig. 16.1 Panels (**a**) and (**b**): example of two genes with an increasing trend where $\hat{\mu}_K^\star/\hat{\mu}_0^\star = 2.1$, but the estimated mean expression of the highest dose $\hat{\mu}_K^\star$ for gene in panel (**b**) is larger than that for gene in panel (**a**). Panels (**c**) and (**d**): example of two genes with a decreasing trend, $\hat{\mu}_K^\star/\hat{\mu}_0^\star = 0.48$, but the estimated mean expression at the highest dose $\hat{\mu}_K^\star$ for gene in panel (**c**) is smaller than that for gene in panel (**d**)

estimated mean expression for the control dose $\hat{\mu}_0^\star$ is different for genes a and b (genes c and d) and, clearly, the estimated mean expression for the highest dose $\hat{\mu}_K^\star$ is different for genes a and b (genes c and d). The aim of the study is to identify genes, for which the increase (decrease) of the mean gene expression from baseline is at least 110%, regardless of mean gene expression $\mu_0^\star$ at the zero dose. Note that for a log transformed data, the ratio test, discussed in this chapter, is less attractive since one can use the multiple contrasts tests (MCTs) discussed in Chap. 15 for the difference. However, we use the human epidermal carcinoma data to illustrate the potential use of the ratio test for the cases in which the ratio (the fold change) is of primary interest.

The four t-type test statistics (Williams', Marcus', M, and the modified M), which were considered in the previous chapters, can be used to test the mean expression levels between the highest dose and the control. Bretz (2006) proposed the use of MCTs for Williams' and Marcus' tests, when the question of interest in the dose-response study is formulated as multiple contrasts, i.e., as differences of treatment means. In this chapter, we extend this idea to test ratios of linear

combinations of treatment means. In particular, the interest is in the ratio parameter $\gamma = \mu(d_K)/\mu(d_0)$, which expresses the mean of the highest dose as a percentage of the mean at the lowest dose.

16.2 Multiple Tests for Ratios

We have discussed the link between Williams- and Marcus-type MCTs and the corresponding testing procedures in Chap. 15. The link between the ratio test and the MCT can be established in a similar way. In this section, we introduce a more general procedure used to perform multiple testing for several ratios of linear combination of treatment means.

A two-sided multiple contrast test for ratios of linear combinations of means $\mu(d_i)$ for doses d_i, $i = 0, \ldots, K$, is given by

$$H_{0\ell} : \frac{\boldsymbol{c}_\ell'\boldsymbol{\mu}}{\boldsymbol{d}_\ell'\boldsymbol{\mu}} = \psi \quad \text{against} \quad H_{1\ell} : \frac{\boldsymbol{c}_\ell'\boldsymbol{\mu}}{\boldsymbol{d}_\ell'\boldsymbol{\mu}} \neq \psi, \ \ell = 1, \ldots, r, \tag{16.1}$$

where $\boldsymbol{c}_\ell$ and $\boldsymbol{d}_\ell$ are vectors of known contrast coefficients associated with the numerator and denominator of the ℓth ratio, ψ is the relative threshold, and $\boldsymbol{\mu} = (\mu(d_0), \ldots, \mu(d_K))'$. Note that one can also test the r ratios against different thresholds ψ_ℓ, $\ell = 1, \ldots, r$.

The likelihood ratio statistics to test the set of hypotheses in (16.1) are given by

$$T_\ell(\psi) = \frac{\boldsymbol{c}_\ell'\widehat{\boldsymbol{\mu}} - \psi\boldsymbol{d}_\ell'\widehat{\boldsymbol{\mu}}}{\widehat{\sigma}[\psi^2\boldsymbol{d}_\ell'\boldsymbol{M}\boldsymbol{d}_\ell - 2\psi\boldsymbol{c}_\ell'\boldsymbol{M}\boldsymbol{d}_\ell + \boldsymbol{c}_\ell'\boldsymbol{M}\boldsymbol{c}_\ell]^{\frac{1}{2}}} \sim t(\nu), \quad \ell = 1, \ldots, r, \tag{16.2}$$

where $\widehat{\boldsymbol{\mu}}$ is the maximum likelihood estimator of $\boldsymbol{\mu}$, $\boldsymbol{M}$ is a diagonal matrix containing the reciprocals of the sample sizes n_k, and $\widehat{\sigma}^2$ is the usual variance estimator with $\nu = \sum_{k=0}^{K} n_k - (K+1)$ degrees of freedom. Under the normality assumption and the null hypotheses, defined in (16.1), the joint distribution of the test statistics in (16.2) is an r-variate t-distribution with ν degrees of freedom and correlation matrix $\boldsymbol{R_0}$, where the elements of the correlation matrix are given by

$$\rho_{ij} = \frac{(\boldsymbol{c}_i - \psi\boldsymbol{d}_i)'\boldsymbol{M}(\boldsymbol{c}_j - \psi\boldsymbol{d}_j)}{\sqrt{(\boldsymbol{c}_i - \psi\boldsymbol{d}_i)'\boldsymbol{M}(\boldsymbol{c}_i - \psi\boldsymbol{d}_i)}\sqrt{(\boldsymbol{c}_j - \psi\boldsymbol{d}_j)'\boldsymbol{M}(\boldsymbol{c}_j - \psi\boldsymbol{d}_j)}}, \tag{16.3}$$

$1 \leq i \neq j \leq r$. Note that all elements of the correlation matrix are known. Therefore, decisions about the tests in (16.1) can be based on the equicoordinate percentage point obtained from the aforementioned multivariate t-distribution (Dilba et al. 2004; Dilba et al. 2006a; Dilba et al. 2006b; Djira 2010). Given a family-wise error rate of α, a two-sided equicoordinate percentage point $c_{1-\alpha,\boldsymbol{R}_0}$ is determined such that

$$\text{Prob}\{|T_\ell(\psi)| \leq c_{1-\alpha,\boldsymbol{R}_0}, \ \ell = 1, 2, \ldots, r\} = 1 - \alpha.$$

For the hypotheses in (16.1), we conclude $H_{1\ell}$ if $|T_\ell(\psi)| > c_{1-\alpha,\boldsymbol{R}_0}$. The associated multiplicity-adjusted two-sided p values can be calculated as

$$\widetilde{p}_\ell = \text{Prob}\{|T_1(\psi)| > |t_\ell(\psi)|, \ldots, |T_r(\psi)| > |t_\ell(\psi)|\}, \quad \ell = 1, 2, \ldots, r$$

where $t_\ell(\psi), \ell = 1, 2, \ldots, r$ are the observed test statistics. Analogously, equicoordinate critical points and adjusted p values for one-sided tests can also be defined.

16.3 Ratio Test for the Highest Dose Versus the Control

As mentioned in Sect. 16.1, one of our primary interests is the inference about the ratio of the mean gene expression (under an order restriction) at the highest dose and the control in a dose-response microarray experiment. For this purpose, we formulate one-sided ratio hypotheses about an increasing and decreasing trend for each gene, respectively, as follows:

$$H_0^{*\text{U}} : \tfrac{\mu(d_K)}{\mu(d_0)} \leq (1+\delta) \text{ vs. } H_1^{*\text{U}} : \tfrac{\mu(d_K)}{\mu(d_0)} > (1+\delta) \text{ and } \mu(d_0) \leq, \ldots, \leq \mu(d_K),$$

$$H_0^{*\text{D}} : \tfrac{\mu(d_K)}{\mu(d_0)} \geq (1-\delta) \text{ vs. } H_1^{*\text{D}} : \tfrac{\mu(d_K)}{\mu(d_0)} < (1-\delta) \text{ and } \mu(d_0) \geq, \ldots, \geq \mu(d_K). \tag{16.4}$$

Note that under order restriction (increasing), μ_k/μ_0 is the largest ratio and therefore $\mu(d_K)/\mu(d_0) = \max(\mu(d_j)/\mu(d_i)), 0 \leq i < j \leq r$. But this is not necessarily true for $\max(\boldsymbol{c}'\boldsymbol{\mu}/\boldsymbol{d}'\boldsymbol{\mu})$, where $\boldsymbol{c}$ and $\boldsymbol{d}$ are arbitrary contrast coefficients. In other words, under order restriction, $\mu(d_K)/\mu(d_0)$ is the maximum of ratios of consecutive doses, which is true for Williams' and Marcus' tests. They are one-sided special cases of the general formulation given in the previous section. The relative thresholds are $\psi = 1+\delta$ and $\psi = 1-\delta$ for the increasing and decreasing trends, respectively. The relative mean difference δ quantifies the biological importance of the trend with respect to the increasing doses. For $\delta > 0$, rejecting the null hypothesis in (16.4) implies that the mean gene expression at the highest dose increases or decreases by $100 \times \delta\%$ compared to the mean expression for the control. Formulating the hypotheses as ratios implies that the relative threshold δ is chosen independently of the gene expression for the control. In general, for hypotheses stated as ratios, the thresholds do not depend on the dimension of the outcome variable.

For example, when $K = 3$, the numerator $\boldsymbol{C} = (\boldsymbol{c}_1', \ldots, \boldsymbol{c}_r')$ and denominator $\boldsymbol{D} = (\boldsymbol{d}_1', \ldots, \boldsymbol{d}_r')$ contrast coefficients defining Williams' trend test as ratios are respectively given by

$$\boldsymbol{C} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{n_2}{n_2+n_3} & \frac{n_3}{n_2+n_3} \\ 0 & \frac{n_1}{n_1+n_2+n_3} & \frac{n_2}{n_1+n_2+n_3} & \frac{n_3}{n_1+n_2+n_3} \end{pmatrix}$$

and

$$\boldsymbol{D} = \begin{pmatrix} 1\,0\,0\,0 \\ 1\,0\,0\,0 \\ 1\,0\,0\,0 \end{pmatrix}.$$

The $\boldsymbol{C}$ and $\boldsymbol{D}$ matrices define a ratio version of Williams' contrast matrix in (15.5). In general, standard multiple comparison procedures like Dunnett's (comparisons with a control, Dilba et al. 2004), Tukey's (all pairwise comparisons), Williams' (Hothorn and Djira 2010), and Marcus' tests for trend and many others can also be expressed as ratios using appropriate numerator and denominator contrast coefficients.

16.4 Use of the `mratios()` Package to the Case Study Data

In this section, we use the package `mratios` (Dilba et al. 2007) to illustrate the ratio test analysis for the human epidermal cell line data. Note that the ratio test is only applied to the 3,499 genes which are declared as significantly differentially expressed by using the likelihood ratio test in Chap. 7.

In order to test a specific δ, we need to specify the value for the margin in function `simtest.ratio()`. This function is to be applied to genes with increasing or decreasing monotonic trends, respectively, because we are interested in conducting a one-sided ratio test. The significance level for controlling the family-wise error rate for the r contrasts for a given contrast test needs to specified in the function as well.

The general form of the function `simtest.ratio()` is given by

```
library(mratios) simtest.ratio(gene~dose, data, type = c("Dunnett","Tukey",
"Sequen", "AVE", "GrandMean", "Changepoint", "Marcus","McDermott",
"Williams", "UmbrellaWilliams"), base = 1,alternative = c("two.sided",
"greater", "less"), Margin.vec = 1,FWER = 0.05,Num.Contrast = NULL,
Den.Contrast = NULL, names = TRUE).
```

There are several options of multiple contrasts possible in the mratios package. For example, "Dunnett option for": many-to-one comparisons with control in the denominator and "Tukey" for all-pair comparisons. For example, here for one gene with an increasing monotonic trend, we want to test whether the mean ratio of gene expression at the highest dose and control is larger than 1.1 (ψ=1.1, i.e., δ=0.1), with Marcus-type contrasts.

```
> data1 -> data.frame(gene1=gene1,x.res=x.res)
> ratio1 -> simtest.ratio(gene1~x.res, data=data1, type = "Marcus",
alternative = "greater", Margin.vec = 1.1, FWER = 0.05,  names = TRUE)
> ratio1
```

The output of the ratio test is given by

```
Tests for ratios of means assuming homogeneous variances
Alternative hypotheses: Ratios  greater  than margins
   margin estimate statistic p.value.raw p.value.adj
C1    1.1    1.029    -3.143    0.993126     0.99947
C2    1.1    1.049    -2.145    0.967886     0.99628
C3    1.1    1.055    -2.348    0.976579     0.99756
C4    1.1    1.147     1.731    0.060808     0.14119
C5    1.1    1.153     2.298    0.025307     0.06433
C6    1.1    1.170     3.198    0.006329     0.01749
```

The lowest p value can be observed for the sixth contrast, which can be used for further analysis when the p values are collected from all the genes to adjust for multiplicity.

To obtain the p values for all the 1,600 genes with increasing trends, we can use the `apply` function to efficiently gather the output into one vector. The following code can be used to test for genes with increasing monotonic trends:

```
> dat.mat.up <- data.sign[direction=="u",]  ## extract genes with
                                  ## significant increasing monotonic trends
> pval.up <- apply(dat.mat.up, 1, function(genei)
+            {min(simtest.ratio(genei~x.res, data=data.frame(genei, x.res),
+             type = "Marcus", alternative = "greater", Margin.vec = 1.1,
+             FWER = 0.05,  names = TRUE)$p.value.adj)} )
> pval.up
[1] 6.038935e-01 1.147182e-02 1.210143e-14  1.607646e-01 9.996908e-01
[6] 9.999983e-01 8.913892e-02 0.000000e+00 1.541190e-01 1.155104e-05
.
.
.
[1591] 0.975922437 0.733315102 0.173187492 0.705281319 0.935561286
[1600] 0.999718217 0.030762376 0.839369199 0.059804112 0.008509572
```

Similarly, for the 1,899 genes with decreasing trends, the following code can be used:

```
> dat.mat.dn <- data.sign[direction=="d",]  ## extract genes with
                                  ## significant decreasing monotonic trends
> pval.dn <- apply(dat.mat.dn, 1, function(genei) {min(
+                  simtest.ratio(genei~x.res, data=data.frame(genei, x.res),
+                  type = "Marcus", alternative = "less", Margin.vec = 0.9,
+                  FWER = 0.05,  names = TRUE)$p.value.adj)} )
> pval <- NULL
> pval[direction=="u"] <- pval.up
> pval[direction=="d"] <- pval.dn
> padj <- p.adjust(pval, method="BH")
> sum(padj[direction=="u"]<=0.05)
 429
> sum(padj[direction=="d"]<=0.05)
 330
```

While testing the null and alternative hypotheses, defined in (16.4), the value of δ can be chosen according to the biological interest. For genes with an increasing trend, the alternative will focus on the differences that are larger than $100 \times \delta\%$ of the control. For genes with a decreasing trend, differences smaller than $100 \times \delta\%$ of the control are of interest. When testing $\delta = 0$, in fact, a two-sided test is made. Nevertheless, in this case, we can categorize significant test results into increasing or decreasing trends.

Table 16.1 Number of genes with statistically significant test results for H_1^{*U} and H_1^{*D} with $\delta = 0, 0.05, 0.1, 0.15$, and 0.2

δ	0	0.05	0.1	0.15	0.2
H_1^{*U} Rejected	2,879	934	429	247	142
H_1^{*D} Rejected	3,387	968	330	133	66

For each gene in the case study dataset, a MCT with Marcus' contrasts is applied, and the BH-FDR procedure is used to control the FDR for all the genes. For testing $\delta = 0$, we obtain 2,024 genes with statistically significant increasing trends and 2,222 genes with decreasing trends. The number of significant tests equals 4,244 $(= 2,024 + 2,222)$ and is slightly different than the number of 3,499 obtained using the E_{01}^2 in Chap. 10. The difference is due to the fact that the inference for MCTs is based on the multivariate t-distribution, while the E_{01}^2 test is carried out based on the null distribution approximated by using permutations.

Moreover, different values of δ (0.05, 0.1, 0.15, and 0.2) are also used for testing H_0^{*U} (H_0^{*D}) against one-sided alternatives H_1^{*U} for genes with an increasing trend (H_1^{*D} for genes with a decreasing trend). For an increasing trend, tests for 934 genes are found significant for δ=0.05, 429 genes for δ=0.1, 247 genes for δ=0.15, and 142 genes for δ=0.2. As expected, as δ increases, the number of significant genes decreases. The same is true for genes with a decreasing trend. Note that genes with significant test results for a larger δ are always the subset of the genes significant test results for a smaller δ.

The R code given above can be modified to test for genes with different values of δ. The summary of results with numbers of significant genes is shown in Table 16.1.

Genes presented in Figs. 16.2 and 16.3 are found to be statistically significant when the $\bar{E}_{01}^2$ test was used. However, the $\bar{E}_{01}^2$ does not distinguish between genes a and b ($\delta = 0.2$) and genes c and d ($\delta = 0.1$). This distinction can be achieved using the ratio test. Moreover, using the t-type tests (Williams', Marcus', M, and the modified M tests) allows us to test the null hypothesis $H_0' : \mu(d_3) - \mu(d_0) = 0$ vs. $H_1^{'\text{Up}} : \mu(d_3) - \mu(d_0) > 0$. However, for a specific shift in mean gene expression between the largest dose and the control, one needs to test $H_0^{''} : \mu(d_3) - \mu(d_0) = \vartheta$ vs. $H_1^{''\text{Up}} : \mu(d_3) - \mu(d_0) > \vartheta$. The value of ϑ is gene specific and depends on the expression level. For example, the increase of 20% for gene a in Fig. 16.2 implies that the estimated difference of $\hat{\mu}_3^\star - \hat{\mu}_0^\star$ is equal to about 5.6, while the increase of 20% for gene b implies that the estimated difference of $\hat{\mu}_3^\star - \hat{\mu}_0^\star$ is equal to about 4.2, because the estimated mean for the control dose $\hat{\mu}_0^\star$ of gene b is larger than that of gene a. In contrast, using the ratio test allows us to test the null hypothesis $H_0^* : \mu(d_3)/\mu(d_0) = 1 + \delta$ vs. $H_1^{*\text{Up}} : \mu(d_3)/\mu(d_0) > 1 + \delta$ in (16.4) with same value of δ (=0.2) for all the genes with increasing trends. The same conclusion can be drawn for genes a, b, c, and d (shown in Fig. 16.3) with decreasing trends for testing $H_0^* : \mu(d_3)/\mu(d_0) = 1 - \delta$ vs. $H_1^{*\text{Down}} : \mu(d_3)/\mu(d_0) < 1 - \delta$, results for genes a and b are statistically significant for $\delta = 0.2$, and results for genes c and d are statistically significant for $\delta = 0.1$.

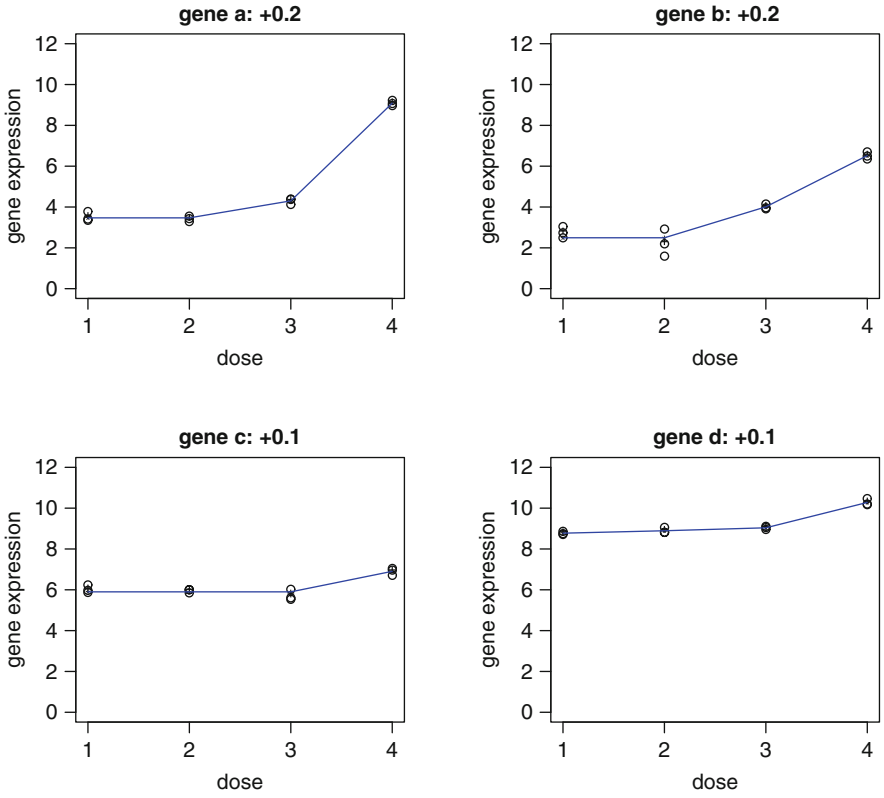

Fig. 16.2 Example of genes with statistically significant test results for the ratio test with $\delta = 0.2$ (*first row*) and $\delta = 0.1$ (*second row*)

16.5 Discussion

In this chapter, we focused on testing the ratio of the mean gene expression difference between the highest dose and the control using the MCT in the microarray setting. It is a special application of MCT for testing ratio parameters.

In the microarray setting, it is often expected that the increase of the dose implies a particular percentage increase/decrease of the mean expression level. Thus, this biological difference is predetermined and the ratio test becomes of special interest. It avoids testing for gene-specific difference in mean gene expression between two doses, which depends on the gene expression level. Based on the significance of ratios and gene expression of the control dose, it is also possible to rank genes, which are found significant using the five test statistics, i.e., the LRT, Williams', Marcus', *M*, and the modified *M* tests, which were discussed in Chap. 7.

In the previous chapter, we have shown a representation of the two mean gene expression level difference under an order restriction using multiple contrasts, i.e., the Marcus-type MCT. In order to answer the testing problem defined in (16.4),

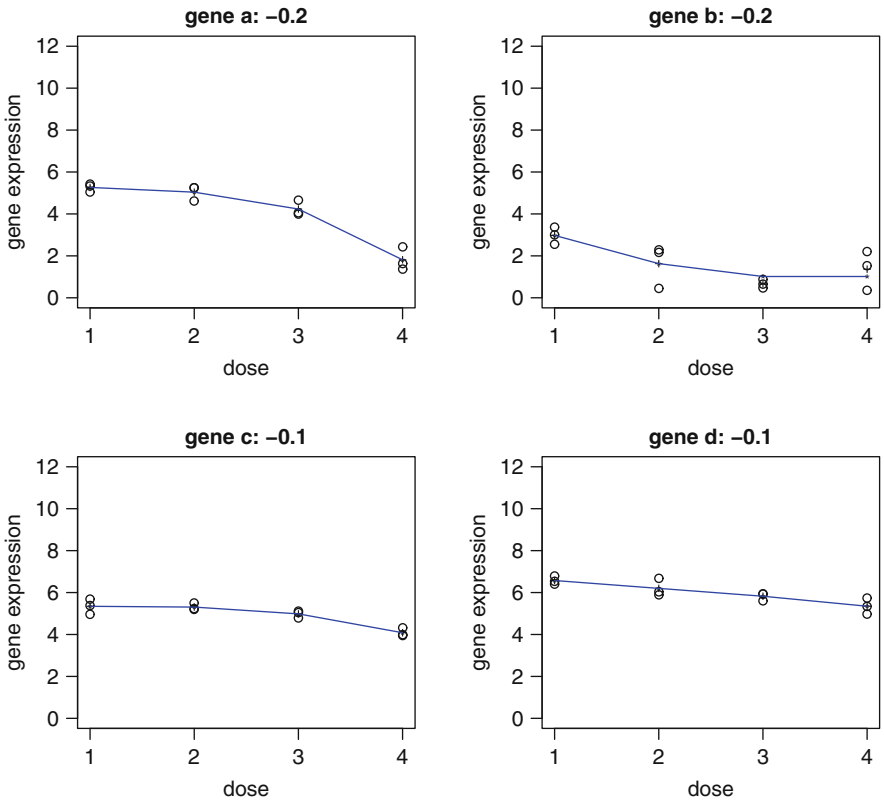

Fig. 16.3 Example of genes with statistically significant test results for the ratio test with $\delta = 0.2$ (*first row*) and $\delta = 0.1$ (*second row*)

the ratio test can be conducted using the MCTs with corresponding multiple ratio contrasts (Dilba et al. 2007). This method efficiently utilizes the multivariate t-distribution as the basis for the inference on the multiple ratio contrast tests.

By choosing a set of δ values in advance, we have performed the ratio tests for each δ. Increasing of δ values reduces the number of significant findings. Genes found with a larger value of δ and gene expression at control dose are of higher interest to the scientists for further investigation.

References

Bretz, F. (2006). An extension of the Williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis, 50*, 1735–1748.

Dilba, D., Bretz, F., Guiard, V., & Hothorn, L. A. (2004). Simultaneous confidence intervals for ratios with application to the comparison of several treatments with a control. *Method of Information in Medicine 43*, 465–469.

Dilba, G., Bretz, F., Hothorn, L. A., & Guiard, V. (2006a). Power and sample size computations in simultaneous tests for non-inferiority based on relative margins. *Statistics in Medicine, 25*, 1131–1147.

Dilba, G., Brets, F., & Guiard, V. (2006b). Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal of Statistical Planning and Inference, 136*, 2640–2658.

Djira, G. D. (2010). Relative potency estimation in parallel-line assys—Method comparison and some extensions. *Communications in Statistics—Theory and Methods, 39*, 1180–1189.

Dilba, G, Schaarschmidt, F., & Hothorn, L. A. (2007). Inferences for ratios of normal means. *RNews 7, 1*, 20–23.

Hothorn, L. A., & Djira, G. D. (2010). Ratio-to-control Williams-type test for trend. *Pharmaceutical Statistics*. doi:10.1002/pst.464.

Chapter 17
Multiple Confidence Intervals for Selected Ratio Parameters Adjusted for the False Coverage-Statement Rate

Dan Lin, Daniel Yekutieli, Gemechis D. Djira, and Ludwig A. Hothorn

17.1 Introduction

Benjamini and Yekutieli (2005) argued that two types of problems generally arise when providing statistical inference for multiple parameters: simultaneity refers to the need to provide inference that simultaneously applies to a subset of several parameters, while selective inference refers to the need to provide valid inferences for parameters that are selected after viewing the data. Since dose–response analysis is provided only for genes that are found to have either increasing or decreasing dose–response relationship and since the mean dose–response curve for each gene is a multivariate object, dose–response analysis of microarray experiments can be viewed as a selection-adjusted simultaneity problem.

In Chap. 16, we discussed using the ratio test in order to select a subset of genes with a significant increase (decrease) of $100 \times \delta\%$ in the mean gene expression. In this chapter, we discuss the construction of confidence intervals (CIs) offering simultaneous coverage for several contrasts of the mean dose–response curve for the subset of selected genes.

D. Lin (✉)
Veterinary Medicine Research and Development, Pfizer Animal Health, Zaventem, Belgium
e-mail: Dan.Lin2@pfizer.com

D. Yekutieli
Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel
e-mail: yekutiel@post.tau.ac.il

G.D. Djira
Department of Mathematics and Statistics, South Dakota State University, Brookings, SD, USA
e-mail: Gemechis.Djira@sdstate.edu

L.A. Hothorn
Institute of Biostatistics, Leibniz University Hannover, Hannover, Germany
e-mail: hothorn@biostat.uni-hannover.de

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_17,

The contents of this chapter are organized as follows. Section 17.2 introduces the methods for constructing the CI for a single ratio and multiple ratios, and we illustrate how to use the R package `mratios` to construct CIs for the ratios in Sect. 17.3 for a single prespecified gene. Section 17.4 discusses constructing CIs for the subset of selected parameters. Section 17.5 specifically discusses the construction of CIs for parameters selected by the BH-FDR procedure. An example of adjusting the false coverage-statement rate (FCR)-controlled CIs using R is given in Sect. 17.6.

17.2 CIs for Single and Multiple Ratios

Following the definitions in Sect. 16.2, for a single ratio, say $\gamma = \boldsymbol{c}'\boldsymbol{\mu}/\boldsymbol{d}'\boldsymbol{\mu}$, given in (17.3), a CI can be constructed using Fieller's theorem (Fieller 1954). A two-sided $(1-\alpha)\times 100\%$ CI for the parameter γ is the solution of the inequality

$$|t(\gamma)| = \frac{|(\gamma \boldsymbol{d} - \boldsymbol{c})'\bar{\boldsymbol{y}}|}{s[\gamma^2 \boldsymbol{d}'\boldsymbol{M}\boldsymbol{d} - 2\gamma \boldsymbol{c}'\boldsymbol{M}\boldsymbol{d} + \boldsymbol{c}'\boldsymbol{M}\boldsymbol{c}]^{1/2}} \leq t_{\alpha/2}(\nu), \tag{17.1}$$

where $t_{\alpha/2}(\nu)$ is the 1-$\alpha/2$ quantile of t-distribution with $\nu = \sum_{i=0}^{K}(n_i - 1)$ degrees of freedom. The limits of the confidence set in (17.1) can be expressed as a quadratic equation in γ:

$$A\gamma^2 + B\gamma + C \leq 0, \tag{17.2}$$

where

$$A = (\boldsymbol{d}'\bar{\boldsymbol{y}})^2 - \tilde{t}^2 s^2 \boldsymbol{d}'\boldsymbol{M}\boldsymbol{d},$$
$$B = -2[(\boldsymbol{c}'\bar{\boldsymbol{y}})(\boldsymbol{d}'\bar{\boldsymbol{y}}) - \tilde{t}^2 s^2 \boldsymbol{c}'\boldsymbol{M}\boldsymbol{d}],$$
$$C = (\boldsymbol{c}'\bar{\boldsymbol{y}})^2 - \tilde{t}^2 s^2 \boldsymbol{c}'\boldsymbol{M}\boldsymbol{c} \quad \text{and} \quad \tilde{t} = t_{\alpha/2}(\nu).$$

Depending on the value of the leading coefficient A, and the discriminant $B^2 - 4AC$, there are three possible solutions to the equation in (17.2) (Kendall 1999). If $A > 0$, then it can be shown that also $B^2 - 4AC > 0$, and there are two solutions to Eq. (17.2). Consequently, the CI is a finite interval lying between the two roots of (17.2). The other two cases, when $A \leq 0$ and $B^2 - 4AC \leq 0$, result in either a region containing all values lying outside the finite interval defined by the two roots of (17.2), or containing the entire γ-axis, which is commonly referred as Fieller's problem. If $\boldsymbol{d}'\bar{\boldsymbol{\mu}}$, the denominator of the ratio, is significantly different from 0, the last two cases occur only with a small probability.

Following the multiple testing problem for ratios in Sect. 16.2, now let us define the unknown ratios of linear combinations of the treatment means by

$$\gamma_\ell = \frac{\boldsymbol{c}_\ell'\boldsymbol{\mu}}{\boldsymbol{d}_\ell'\boldsymbol{\mu}}, \quad \ell = 1, \ldots, r, \tag{17.3}$$

where $\boldsymbol{c}_\ell$ and $\boldsymbol{d}_\ell$ are vectors of known contrast coefficients associated with the numerator and denominator of the ℓth ratio, and $\boldsymbol{\mu} = (\mu(d_0), \ldots, \mu(d_K))'$. To derive simultaneous CIs for $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)'$, we apply the following linear form (Fieller 1954) originally used for single ratios. Let

$$L_\ell = (\boldsymbol{c}_\ell - \gamma_\ell \boldsymbol{d}_\ell)' \hat{\boldsymbol{\mu}}, \quad \ell = 1, \ldots, r.$$

Under the normality assumption for the response variable, L_ℓ is distributed as $N(0, \sigma^2_{L_\ell})$, where

$$\sigma^2_{L_\ell} = \mathrm{Var}(L_\ell) = \sigma^2 (\boldsymbol{c}_\ell - \gamma_\ell \boldsymbol{d}_\ell)' \boldsymbol{M} (\boldsymbol{c}_\ell - \gamma_\ell \boldsymbol{d}_\ell),$$

and $\mathrm{Var}(\hat{\boldsymbol{\mu}}) = \sigma^2 \boldsymbol{M}$, with $\boldsymbol{M}$ being a diagonal matrix containing the reciprocals of the sample sizes n_i. Let S^2 be the unbiased pooled variance estimator of the common variance σ^2 based on $\nu = \sum_{i=0}^{K} (n_i - 1)$ degrees of freedom. Therefore, an unbiased estimator $S^2_{L_\ell}$ of $\sigma^2_{L_\ell}$ is given by

$$S^2_{L_\ell} = S^2 (\boldsymbol{c}_\ell - \gamma_\ell \boldsymbol{d}_\ell)' \boldsymbol{M} (\boldsymbol{c}_\ell - \gamma_\ell \boldsymbol{d}_\ell).$$

Since S_{L_ℓ} is distributed as $(\sigma^2_{L_\ell} \nu^{-1} \chi^2(\nu))^{1/2}$ and is independent of L_ℓ, the test statistic

$$T_\ell(\gamma_\ell) = L_\ell / s_{L_\ell}$$

follows a t-distribution with ν degrees of freedom. Jointly for multiple ratios $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)$, the vector of test statistic $\boldsymbol{T} = (T_1, \ldots, T_r)'$ follows a central r-variate t-distribution with ν degrees of freedom and correlation matrix $\boldsymbol{R} = [\rho_{ij}]$, where

$$\rho_{ij} = \frac{(\boldsymbol{c}_i - \gamma_i \boldsymbol{d}_i)' \boldsymbol{M} (\boldsymbol{c}_j - \gamma_j \boldsymbol{d}_j)}{\sqrt{(\boldsymbol{c}_i - \gamma_i \boldsymbol{d}_i)' \boldsymbol{M} (\boldsymbol{c}_i - \gamma_i \boldsymbol{d}_i)} \sqrt{(\boldsymbol{c}_j - \gamma_j \boldsymbol{d}_j)' \boldsymbol{M} (\boldsymbol{c}_j - \gamma_j \boldsymbol{d}_j)}}$$

$1 \le i \ne j \le r$. Unlike the correlation matrix $\boldsymbol{R}_0$ associated with the multiple test in the previous chapter, the correlation matrix associated with the estimation problem is a function of the unknown ratio parameters $\gamma_\ell, \ell = 1, \ldots, r$. This means that the equicoordinate critical point of multivariate t-distribution which we use for simultaneous CI calculations will depend on the unknown $\boldsymbol{\gamma}$ (see Dilba et al. 2004, 2006 for detailed discussions of this problem). In general, a two-sided $(1-\alpha)\times 100\%$ CI for γ_ℓ can be obtained by solving the inequality

$$|T_\ell(\boldsymbol{\gamma})| = \frac{|(\gamma_\ell \boldsymbol{d}_\ell - \boldsymbol{c}_\ell)' \hat{\boldsymbol{\mu}}|}{S[\gamma^2_\ell \boldsymbol{d}'_\ell \boldsymbol{M} \boldsymbol{d}_\ell - 2\gamma^2_\ell \boldsymbol{c}'_\ell \boldsymbol{M} \boldsymbol{d}_\ell + \boldsymbol{c}'_\ell \boldsymbol{M} \boldsymbol{c}_\ell]^{1/2}} \le Q, \quad \ell = 1, \ldots, r \tag{17.4}$$

where Q is a quantile determined from certain distribution. For $Q = t_{\alpha/2}(\nu)$, one obtains marginal Fieller CIs for γ_ℓ, and if Q is multiplicity-adjusted critical point, the intervals will be simultaneous CIs. For example, Bonferroni-adjusted simultaneous CIs are obtained by taking $Q = t_{\frac{\alpha}{2r}}(\nu)$. A less conservative approach is to use Šidák (1967) inequality and take Q to be an equicoordinate critical point of a multivariate t-distribution with ν degrees of freedom and correlation matrix the identity matrix ($\boldsymbol{I}_r$). Yet, another method is the plug-in approach by which we replace the unknown γ_ℓs in the correlation matrix $\boldsymbol{R}$ by their corresponding maximum likelihood estimates ($\widehat{\gamma}_\ell = \boldsymbol{c}'_\ell\widehat{\boldsymbol{\mu}}/\boldsymbol{d}'_\ell\widehat{\boldsymbol{\mu}}$) and then take Q as an equicoordinate critical point of an r-variate t-distribution with ν degrees of freedom and estimated correlation matrix $\boldsymbol{R}(\widehat{\boldsymbol{\gamma}})$. The latter approach is observed to possess better simultaneous coverage (Dilba et al. 2006; Djira 2010).

17.3 Use of the `mratios` Package to Construct CIs for Ratios

In this section, we illustrate the use of the function `sci.ratio()` in the `mratios` package to construct CIs for the multiple ratios of one gene. The general form of the function is written as,

```
sci.ratio(y~x, data, type = c("Dunnett","Turkey","Marcus",\ldots), base = 1,
 method = c("Plug","MtI", "Bonf", "unadj"), Num.Contrast = NULL,
 Den.Contrast = NULL,  alternative = c("two.sided", "greater","less"),
 conf.level = 0.95, names=TRUE)
```

in which the type of contrasts can be specified in the same way as in the function `contrMat` or user can define numerator and denominator contrast matrices. The methods to estimate the CI are "Plug" of the maximum likelihood estimators of the ratio parameters into the correlation matrix as discussed above, "MtI" for Šidák (1967) or Slepian adjustment, "Bonf" for Bonferroni adjustment, and "unadj" for unadjusted CIs.

We use two genes with significant ratio of 1.2 as an example to illustrate use of the `mratios` for the construction of CIs for multiple ratios in the MCTs.

```
> gene1 <- data.sign[3, ]
> data1 <- data.frame(gene1=gene1,x.res=x.res)
> CI.wil1=sci.ratio(gene1~x.res,data=data1,conf.level=0.95,method = "Plug",
type="Williams",alternative="greater")
> CI.mar1=sci.ratio(gene1~x.res,data=data1,conf.level=0.95,method = "Plug",
type="Marcus",alternative="greater")
> gene2 <- data.sign[11, ]
> data2 <- data.frame(gene2=gene2,x.res=x.res)
> CI.wil2=sci.ratio(gene2~x.res,data=data2,conf.level=0.95,method = "Plug",
type="Williams",alternative="greater")
> CI.mar2=sci.ratio(gene2~x.res,data=data2,conf.level=0.95,method = "Plug",
type="Marcus",alternative="greater")
```

```
> CI.wil1
Simultaneous 95-% confidence intervals
   estimate  lower
C1   1.3465 1.3162
C2   1.1950 1.1700
C3   1.1226 1.0998

> CI.mar1
Simultaneous 95-% confidence intervals
   estimate  lower
C1   1.1226 1.0963
C2   1.1950 1.1661
C3   1.2084 1.1849
C4   1.3465 1.3116
C5   1.3617 1.3322
C6   1.3371 1.3107

> CI.wil2
Simultaneous 95-% confidence intervals
   estimate  lower
C1   2.5878 2.4409
C2   1.9063 1.7987
C3   1.5961 1.5063

> CI.mar2
Simultaneous 95-% confidence intervals
   estimate  lower
C1   1.5961 1.4899
C2   1.9063 1.7790
C3   1.9298 1.8306
C4   2.5878 2.4141
C5   2.6197 2.4837
C6   2.4257 2.3215
```

As we can see, the three contrasts of Williams-type MCT (C1–C3) are included in the Marcus-type MCT (C4, C2, and C1) as discussed in Sect. 14.3. Since we adjust for six and three CIs for Marcus' type ratio test and Williams' type ratio test, respectively, the CIs for Marcus' ratio test are slightly wider as compared to that of William' CIs.

17.4 False Coverage-Statement Rate

Benjamini and Yekutieli (2005) assumed that there are m parameters $\theta_1 \cdots \theta_m$, with corresponding estimators $T_1, \ldots, T_m$, and the goal is to construct valid CIs for the parameters selected by a given selection criterion $\tilde{S}(T_1 \cdots T_m) \subseteq \{1 \cdots m\}$. They showed that CIs constructed for selected parameters no longer ensure nominal coverage probability and suggested the FCR, a generalization of the FDR, as the appropriate criterion to capture the error for CIs constructed for selected parameters. The FCR is also defined as $E\{V/\max(R, 1)\}$, where R is the number of CIs constructed and V is the number of non-covering CIs.

Benjamini and Yekutieli (2005) introduced a general method of ensuring FCR $\leq q$ for independently distributed test statistics and any selection criterion: construct marginal $1 - R \cdot q/m$ CIs for each of the R selected parameters. In dose–response

microarray experiments, R is the number of selected genes, and the CIs constructed for each gene actually consist of the set of marginal CIs constructed for the contrasts of the mean dose–response curve for this gene. Thus, V is the number of selected genes for which at least one mean dose–response contrast is not covered by the corresponding marginal CI, and the level 0.05 FCR control implies that for approximately 95% of the selected genes all the CIs for the mean dose–response contrasts cover the corresponding parameter.

17.5 FCR-Adjusted BH-Selected CIs

The construction of the FCR-adjusted BH-selected CIs, proposed by Benjamini and Yekutieli (2005), consists of four steps. In the first three steps, we adjust the multiplicity for the ratio tests in a similar way as what was done in Chap. 15 for the MCTs:

1. Sort the p values used for testing the m hypotheses regarding the ratio parameters, $P_{(1)} \leq \cdots \leq P_{(m)}$. Note that P_i is the minimum p value from the MCT ratio test for each gene.
2. Calculate $R_{\text{CI}} = \max\{i : P_{(i)} \leq i \cdot q/m\}$, where q is the FDR level to be controlled.
3. Select the R_{CI} parameters, for which $P_{(i)} \leq R_{\text{CI}} \cdot q/m$, corresponding to the rejected hypotheses.
4. Construct a $1 - R_{\text{CI}} \cdot q/m$ CI for each parameter selected.

The last step is related to the construction of the FCR-adjusted CIs. The length of the constructed CIs increases as the number of parameters considered (m) increases, but decreases as the number of selected parameters (R_{CI}) increases (Benjamini and Yekutieli 2005).

Constructing the FCR-adjusted CIs is directly linked to the BH-FDR procedure that controls the FDR. The FCR-adjusted BH-selected procedure ensures the control of the FCR as well as the control of the FDR. If the null hypotheses are rejected in the testing step, the FCR-adjusted CIs will not cover their null parameter values. This property only holds if the CIs correspond to the p values, meaning one(two)-sided p values should correspond to the one(two)-sided CIs for selected parameters.

17.6 Construction of FCR-Adjusted CIs Using the `mratios` Package

In this section, we construct the FCR-adjusted BH-selected CIs for the ratio parameters of genes found significant by the ratio test, with $\delta = 0.2$ in Sect. 16.4 at the significance level of 0.05 with BH-FDR adjustment. The number of rejected hypotheses for $\delta = 0.2$ is 142 for the increasing trend and 66 for the decreasing

trend. The FCR-adjusted $1 - R_{\mathrm{CI}} \cdot q/m$CI for the selected ratios is constructed, where $R_{\mathrm{CI}} = 208$ and $m = 3,499$.

By adjusting the significance level in the function `sci.ratio` at $1 - R_{\mathrm{CI}} \cdot q/m$, i.e., 0.997 (=$1 - 0.05 * (208/3499)$) for each gene, the simultaneous CIs for ratios of genes with increasing trends can be constructed using the following code:

```
> # sign.1.2 is the list of genes with ratio of mean expression
between the highest dose and the control larger than 1.2
> data.1.2ratio <- data.sign[sign.1.2,]
> sign.level <- round(1-0.05*(208/3499),4)
> CIs <- apply(data.1.2ratio, 1, function(genei) sci.ratio(genei~x.res,
data=data.frame(genei, x.res),conf.level=sign.level,method = "Plug",
type="Marcus",alternative="greater") )
> CIs[[1]]
Simultaneous 99.7-% confidence intervals
   estimate  lower
C1   1.1226 1.0757
C2   1.1950 1.1434
C3   1.2084 1.1663
C4   1.3465 1.2841
C5   1.3617 1.3088
C6   1.3371 1.2895

> CIs[[2]]
Simultaneous 99.7-% confidence intervals
   estimate  lower
C1   1.5961 1.4115
C2   1.9063 1.6851
C3   1.9298 1.7548
C4   2.5878 2.2857
C5   2.6197 2.3798
C6   2.4257 2.2401
```

Comparing to the adjusted CIs constructed for these two genes in Sect. 17.3 by using Marcus' contrast matrix, it is easy to observe that the simultaneous CIs are wider. We also compare these simultaneous CIs adjusted by the FCR with the Bonferroni-adjusted CIs, which are expected to widen the CIs even more. Figure 17.1 shows the CIs for ten selected genes found significant by the ratio test with $\delta = 0.2$ using the Bonferroni (with the confidence level of 0.9999857 (=$1 - 0.05/3,499$), the FCR-adjusted, and without any multiplicity adjustment (namely, unadjusted).

Figure 17.2 shows the width of the CIs for ratios of mean gene expressions between the highest dose and the control for the 208 significant genes. The unadjusted CIs (dotted line with pluses shown in Fig. 17.2) are always the shortest, while the Bonferroni CIs are the widest (solid line).

17.7 Discussion

In the microarray setting, the control of the FDR is well addressed when testing thousands of genes simultaneously (Ge et al. 2003). The multiplicity issue in the construction of simultaneous CIs for the selected genes needs to be addressed in

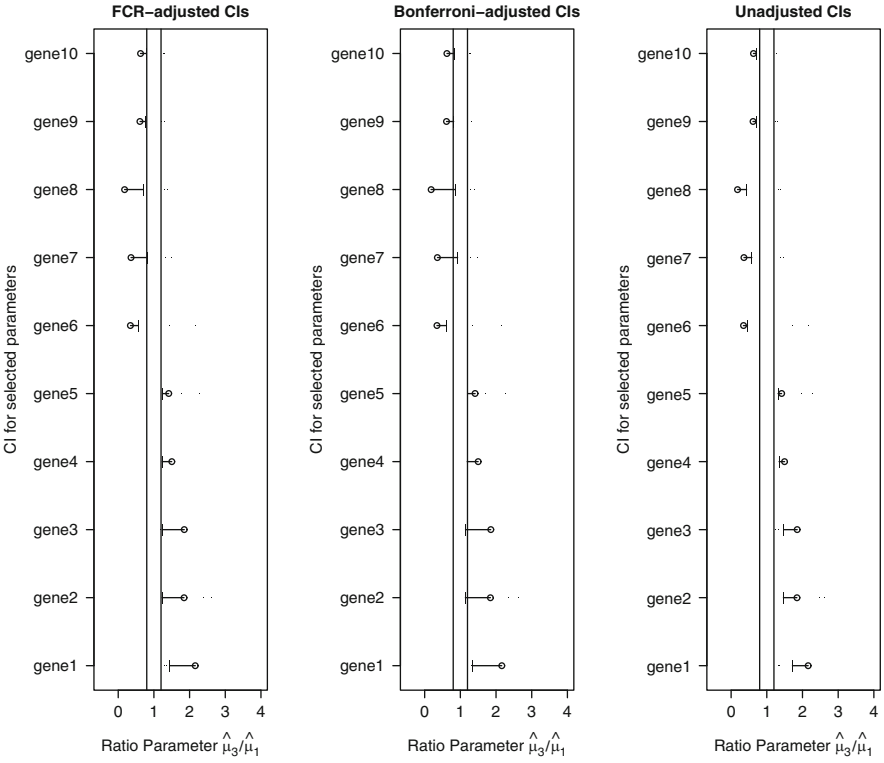

Fig. 17.1 Example of CIs constructed by the Bonferroni, FCR-adjusted, and the adjusted method for selected genes found significant by the ratio test with $\delta = 0.2$

this context as well. In this chapter, our aim was to construct simultaneous CIs for the ratio parameters for genes found significant by the ratio tests in Sect. 16.4.

We applied the procedure proposed by Benjamini and Yekutieli (2005) to adjust the simultaneous confidence levels to ensure the FCR. The FCR-adjusted BH-selected procedure links the control of FCR in constructing CIs with control of the FDR in testing the hypotheses. We compared the length of CIs obtained using the FCR adjustment, no adjustment, and Bonferroni adjustment. The constructed FCR-adjusted BH-selected CIs did not cover the null value of parameter tested, while the Bonferroni-adjusted CIs were too wide. The use of the FCR-adjusted CIs needs to be highlighted as an analogue to the control of the FDR in addressing for multiple testing problem in the microarray setting.

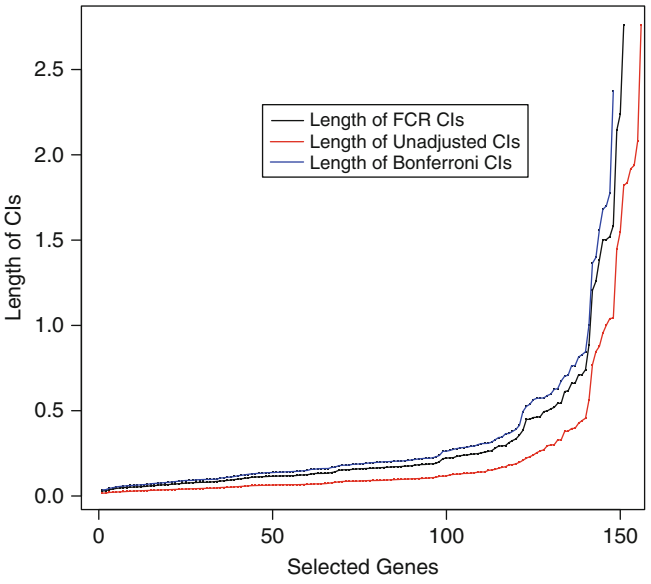

Fig. 17.2 Length of CIs for 208 genes found significant by the ratio test with $\delta = 0.2$ for comparing the mean gene expression of the highest dose versus the control, using the unadjusted, Bonferroni and FCR-adjusted approaches

References

Benjamini, Y., & Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association, 100*, 71–81.

Dilba, G., Bretz, F., Hothorn, L. A., & Guiard, V. (2006). Power and sample size computations in simultaneous tests for non-inferiority based on relative margins. *Statistics in Medicine, 25*, 1131–1147.

Dilba, G., Bretz, F., Guiard, V., & Hothorn, L. A. (2004). Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods of Information in Medicine, 43*(5), 465–469.

Djira, G. D. (2010). Relative potency estimation in parallel-line assys—Method comparison and some extensions. *Communications in Statistics—Theory and Methods, 39*, 1180–1189.

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of Royal Statistical Society B, 16*, 175–185.

Ge, Y., Dudoit, S., & Speed, P. T. (2003). *Resampling based multiple testing for microarray data analysis* (Technical report, 633). University of Berkeley.

Kendall, M. G. (1999). *The Advanced Thoery of Statistics*, Volume 2A: Classical Inference and the Linear Model, M. Kendall, A. Stuart, J.K. Ord, S. Arnold, Oxford University Press, 1999.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*, 626–633.

Chapter 18
Interfaces for Analyzing Dose–Response Studies in Microarray Experiments: `IsoGeneGUI` and `ORIOGEN`

Setia Pramana, Philippe Haldermans, and Tobias Verbeke

18.1 Introduction

According to Ernst and Bar-Joseph (2006), in 39.1% of the 786 datasets in the Gene Expression Omnibus of 2005 are studies with ordered-restricted design variable such as age, time, temperature, and dose. Table 18.1 presents a list of free software developed for the analysis of gene expression experiments with order-restricted design variable. There is a substantial amount of overlapping between the different packages presented in Table 18.1 and the same or similar analysis can be conducted using more than one package.

In this chapter, we present two interfaces for the analysis of dose–response microarray data. The `IsoGeneGUI` is a bioconductor menu-based R package. The package has the same graphical support as the `IsoGene` package, discussed previously in the book, and therefore all the output provided by the `IsoGene` library can be produced with the `IsoGeneGUI` as well. The `ORIOGEN` package Peddada et al. (2003) is a java-based interface which can be used to produce the order-restricted analysis presented in Chap. 11. For illustration of the capacity

S. Pramana (✉)
Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden
e-mail: setia.pramana@ki.se

P. Haldermans
Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium
e-mail: Philippe.haldermans@uhasselt.be

T. Verbeke
OpenAnalytics BVBA, Heist-op-den-Berg, Belgium
e-mail: tobias.verbeke@openanalytics.be

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2_18,

Table 18.1 Free software available for order-restricted analysis of ordered gene expression data

Package	Type	Location	Reference
orQA	R	CRAN	Klinglmueller et al. (2011)
IsoGene	R	CRAN	Pramana et al. (2010a)
IsoGeneGUI	R	Bioconductor	Pramana et al. (2010b)
ORIOGEN	Java	http://dir.niehs.gov/dirbb/oriogen/index.cfm	Peddada et al. (2003)
ORIClust	R	CRAN	Liu et al. (2009)
STEM	Java	http://www.cs.cmu.edu/~jernst/stem	Ernst and Bar-Joseph (2006)
ORCME	R	CRAN	Otava et al. (2011)

of the packages, the antipsychotic compound study analyzed in Chap. 14 will be used.

18.2 The `IsoGeneGUI` Package

18.2.1 Reading the Data

In the first step of the analysis, the data are loaded into R using the `File` menu. The package can read data in an R workspace file (*.RData), excel (*.xls), or text (*.txt) format. In order to load the R workspace, we choose the following sequence:

```
File > Open dataset > R workspace
```

Text and Excel files can be loaded as well in a similar way.

18.2.2 Data Analysis

The `IsoGeneGUI` package provides three options for analysis, which we discussed earlier in Chaps. 3, 7, and 8: (1) analysis with the likelihood ratio test based on the exact distribution of the test statistic (Chap. 3) (2) resampling-based analysis (Chap. 7) and (3) significance analysis of microarrays (Chap. 8).

18.2.2.1 Likelihood Ratio Test Using Exact *p* Values

To perform the analysis with the LRT using the exact distribution of $\bar{E}^2_{01}$, discussed in Chap. 3, we choose the following sequence in the `Analysis` menu:

```
Analysis > Likelihood Ratio Test (E2)
```

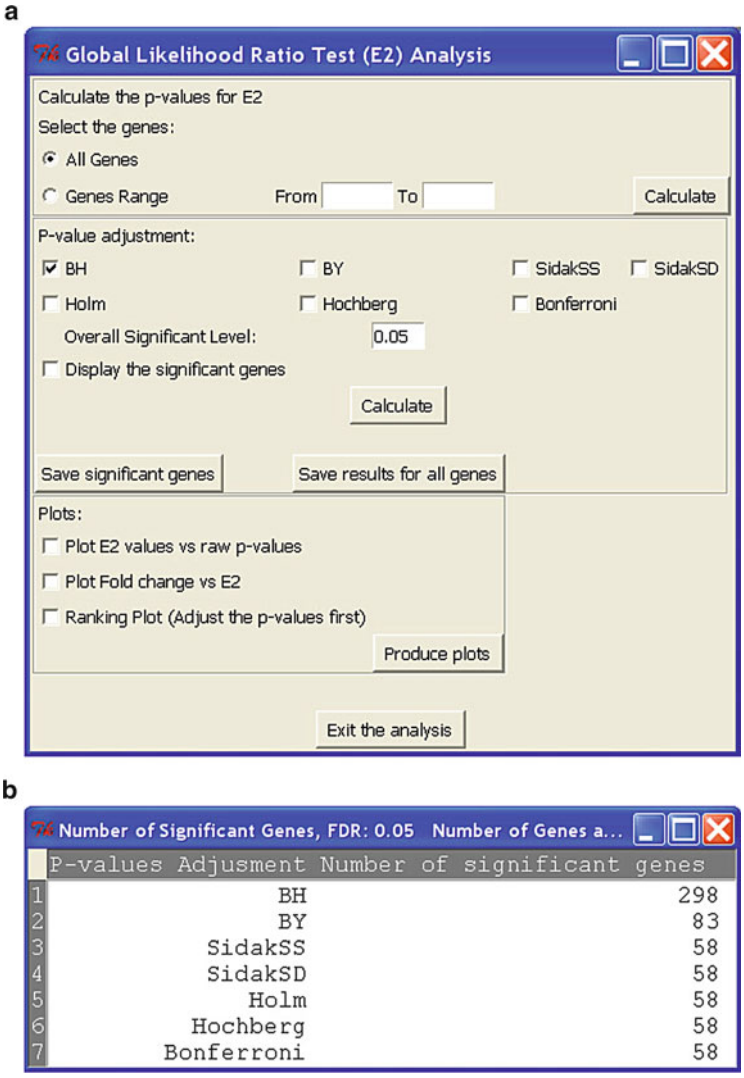

Fig. 18.1 The main windows for the likelihood ratio test. (**a**) The dialog box for the LRT statistic E^2_{01}. (**b**) Numerical output for the LRT statistic E^2_{01}

The main dialog box for `Analysis` based on the LRT is shown in Fig. 18.1a. Note that, we can choose to perform the analysis for all the genes in the array or on a predefined subset of genes. In addition, in order to adjust for multiplicity, one can select from the menu the multiplicity adjustment method to use and the type I error rate to control.

For the analysis of the antipsychotic data, all the genes are analyzed and the exact p values were adjusted using all the methods available in the package. The

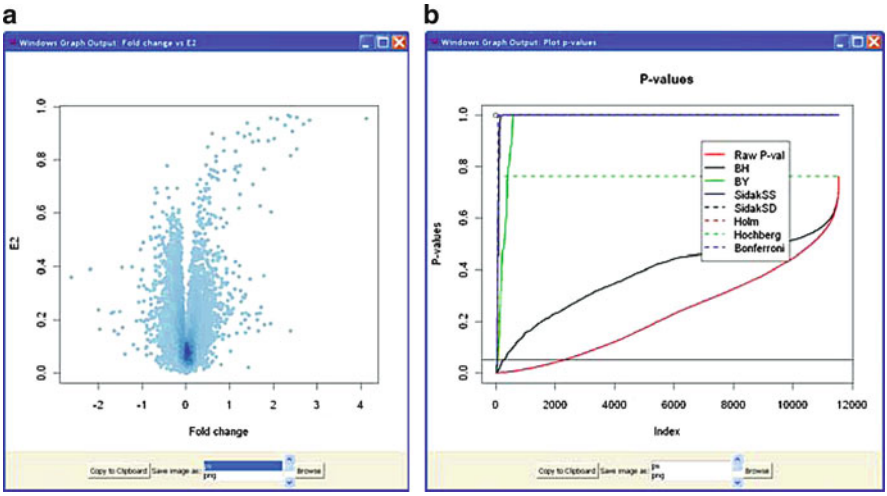

Fig. 18.2 Graphical outputs of the likelihood ratio test. Panel (**a**): volcano plot of $\bar{E}_0^2$. Panel (**b**): raw and adjusted p values for the LRT statistic

output window in Fig. 18.1b shows numbers of significant genes for each selected adjustment method. In this study, 298 genes were found to be significant using the FDR-BH method. Default graphs are provided by the package as shown in Fig. 18.2. All graphical displays can be copied into clipboard and saved in a separate file or alternatively can be saved into several image formats (*.ps, *,png, *.jpeg, *.bmp, and *.tiff).

18.2.3 Resampling-Based Methods

Figure 18.3a shows the main menu for the resampling-based inference discussed in Chap. 7. We need to specify number of permutations, test statistic(s) to be used, and multiplicity adjustment methods. Numerical and graphical output can be produced in a same way discussed in the previous section. Figure 18.3 shows the main dialog box for the permutation tests.

18.2.3.1 Significance Analysis of Microarrays

Significance analysis of dose–response microarray data, discussed in Chap. 8, can be performed by using the last sub-option of the `Analysis` menu shown in Fig. 18.3b:

```
Analysis > Significance Analysis of Microarrays > SAM permutation
```

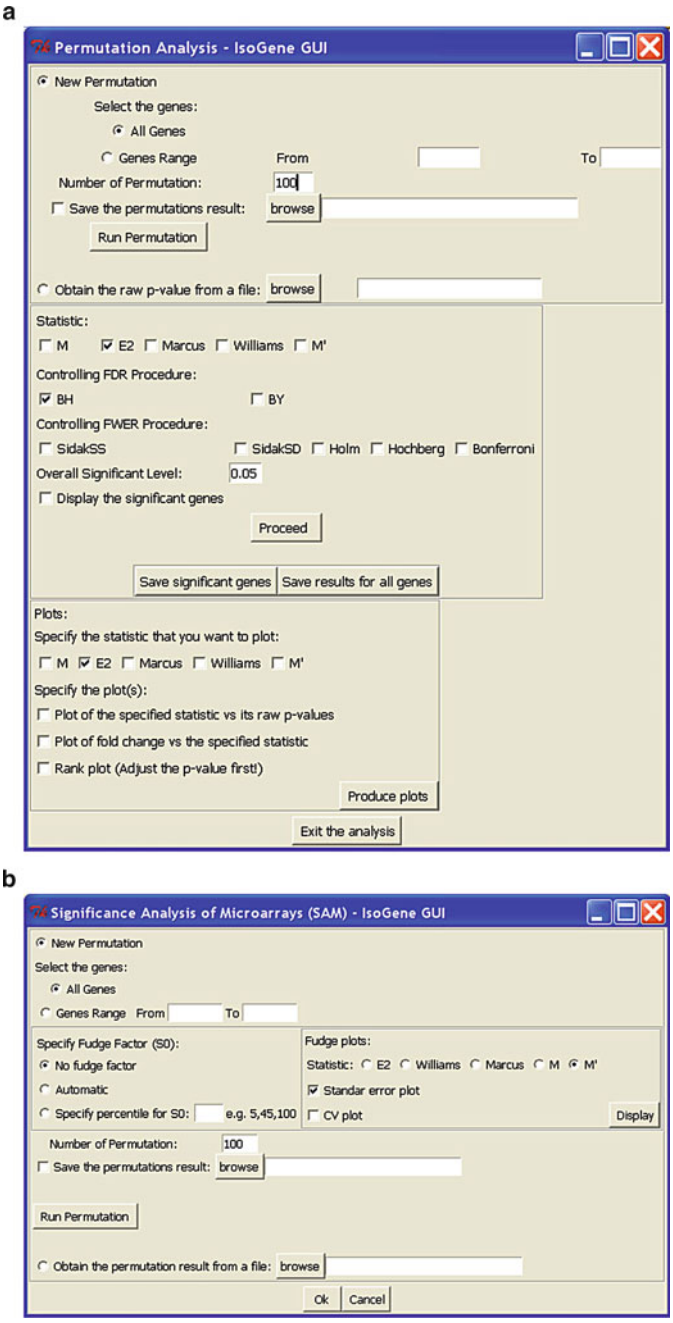

Fig. 18.3 Resampling-based inference in the `IsoGeneGUI`. (**a**) Permutation based inference; (**b**) the SAM window

Figure 18.3b presents the window for calculating the SAM regularized test statistics using permutations. The analysis can be carried out with automatic selection or a predefined percentile or without fudge factor. In the case of analysis with the automatic choice of fudge factor by the SAM, the package obtains the fudge factor by minimizing the coefficient of variations (CV) of median absolute deviations (MAD) of the selected test statistic.

Once the gene specific test statistics are computed or loaded from an external (saved) file, we can now perform the SAM by following the sequence:

```
Analysis > Significance Analysis of Microarrays > SAM Analysis
```

For the antipsychotic study, using 100 permutations, we found that the number of genes with a significant monotonic trend is equal to 170 (FDR $=$ 0.05 and delta $=$ 0.67, see Fig. 18.4a), in which 93 and 77 are under increasing and decreasing trends, respectively. Figure 18.4b shows the SAM plot of the expected versus the observed test statistics. Other exploratory plots, summary statistics, and user-defined plots can be produced easily using the `Plots` menu.

18.3 The `ORIOGEN` Package

The `ORIOGEN` package (Peddada et al. 2005) is a java-based interface which can be used to perform an order-restricted inference for ordered gene expression data. In comparison with the `IsoGeneGUI`, the `ORIOGEN` package allows to test for partial order alternatives assuming heteroscedasticity. The `ORIOGEN` is a resampling-based inference method and uses the SAM methodology to carry out the analysis. As explained in Chap. 11, the `ORIOGEN` clusters the genes found to be significant into subgroups with the same dose–response profile.

Figure 18.5 shows the main dialog box in which we specify: (1) the input data (a text file with the expression matrix), (2) the output file, (4) number of dose levels, (5) number of replicates at each dose level, (6) number of bootstrap samples, (7) FDR level, and (8) the quantile for the fudge factor. A complete description of each item in the input window is given in the help file of `ORIOGEN`.

In the next step, we need to specify in advance the mean profiles of primary interest for the analysis. Note that if we specify only decreasing or increasing profiles, the `ORIOGEN` and the SAM in the `IsoGeneGUI` perform similar analysis, although the analysis using `IsoGeneGUI` is done under the assumption of homoscedasticity and automatic selection of the fudge factor to minimize the CV of the MAD of the test statistic, while the analysis in the `ORIOGEN` is done under the assumption of heteroscedasticity and uses a fixed quantile for the fudge factor.

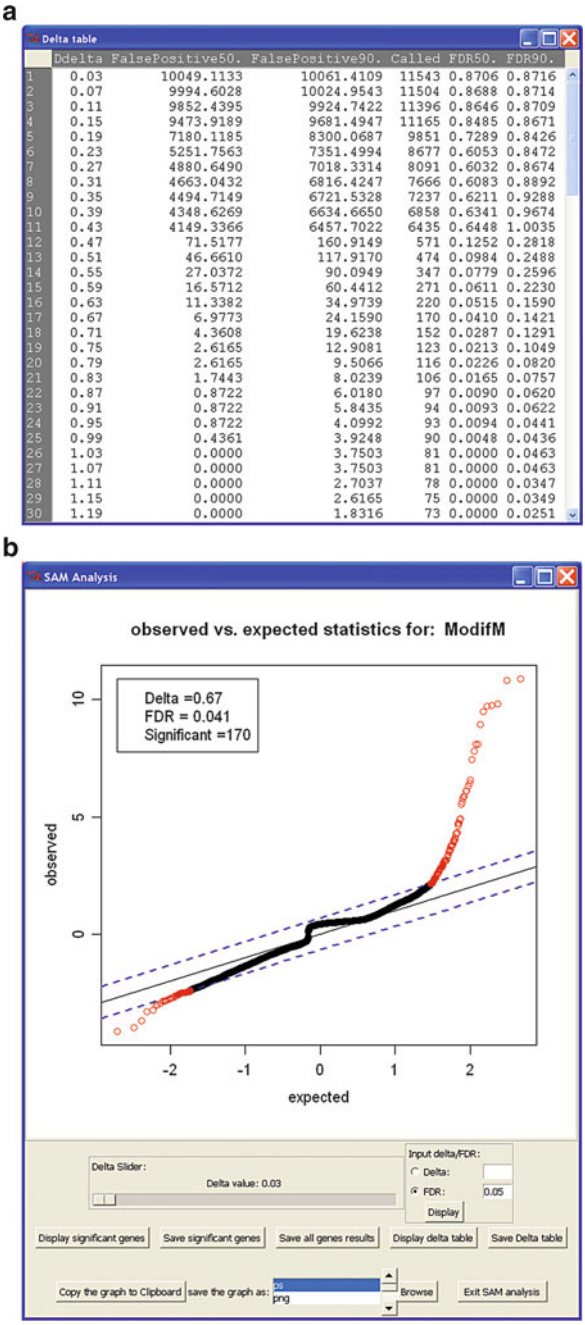

Fig. 18.4 Graphical and numerical output of the SAM. (**a**) The delta table; (**b**) expected versus observed test statistics

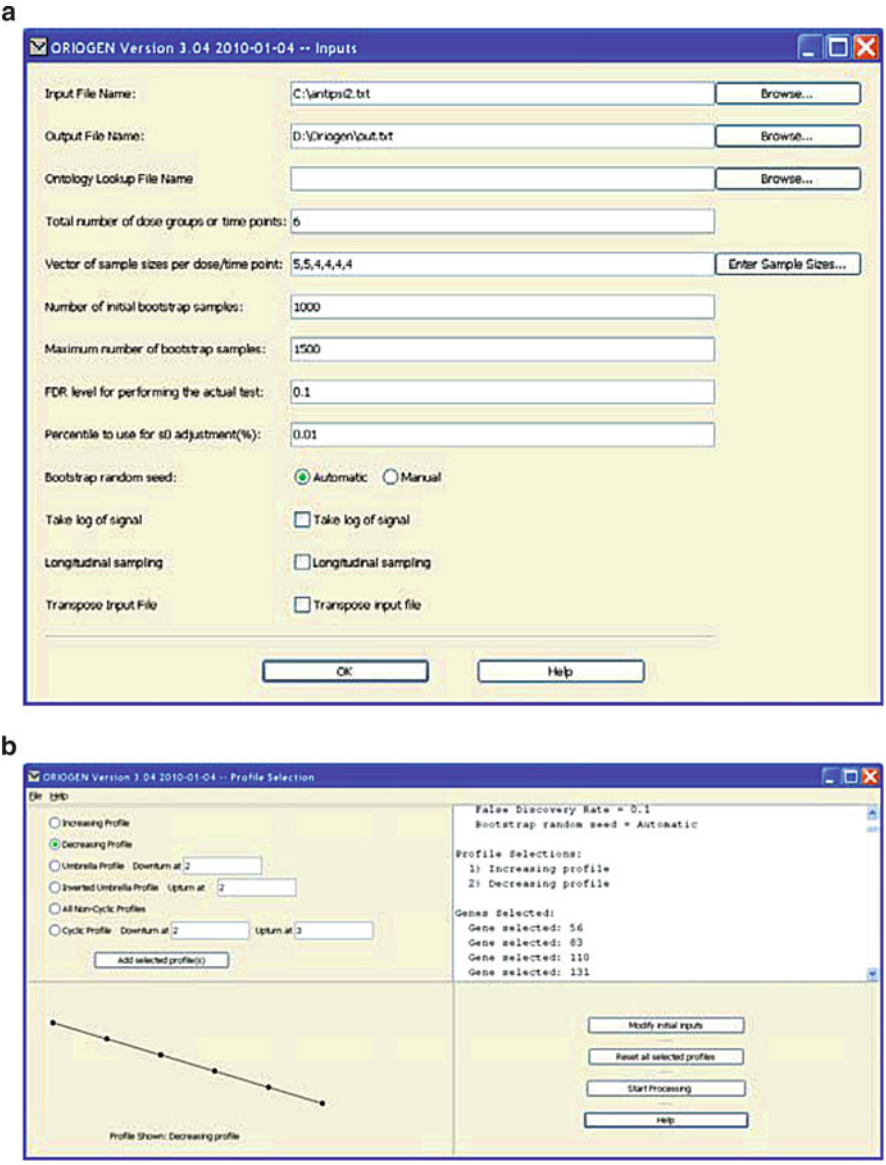

Fig. 18.5 The main input screens of the ORIOGEN package. (**a**) Specification of initial values for in ORIOGEN. (**b**) Specification of mean profiles of interest

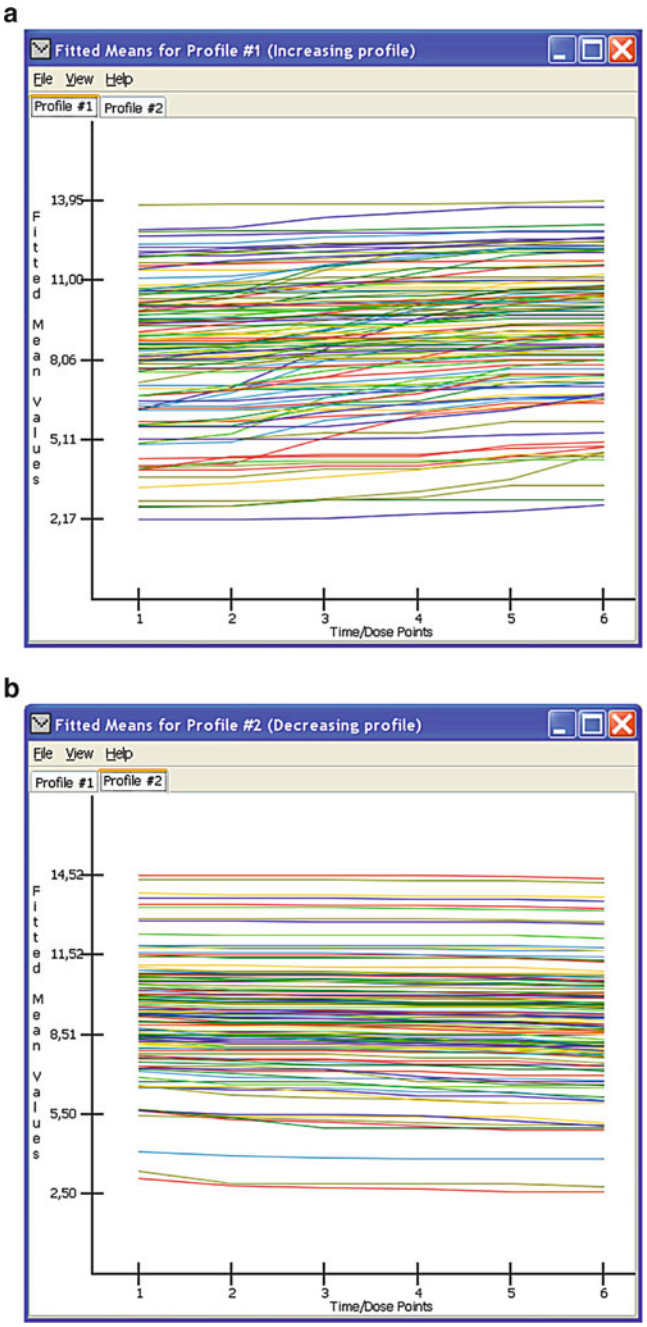

Fig. 18.6 Genes with significant increasing/decreasing profiles. (**a**) Increasing profiles; (**b**) decreasing profiles

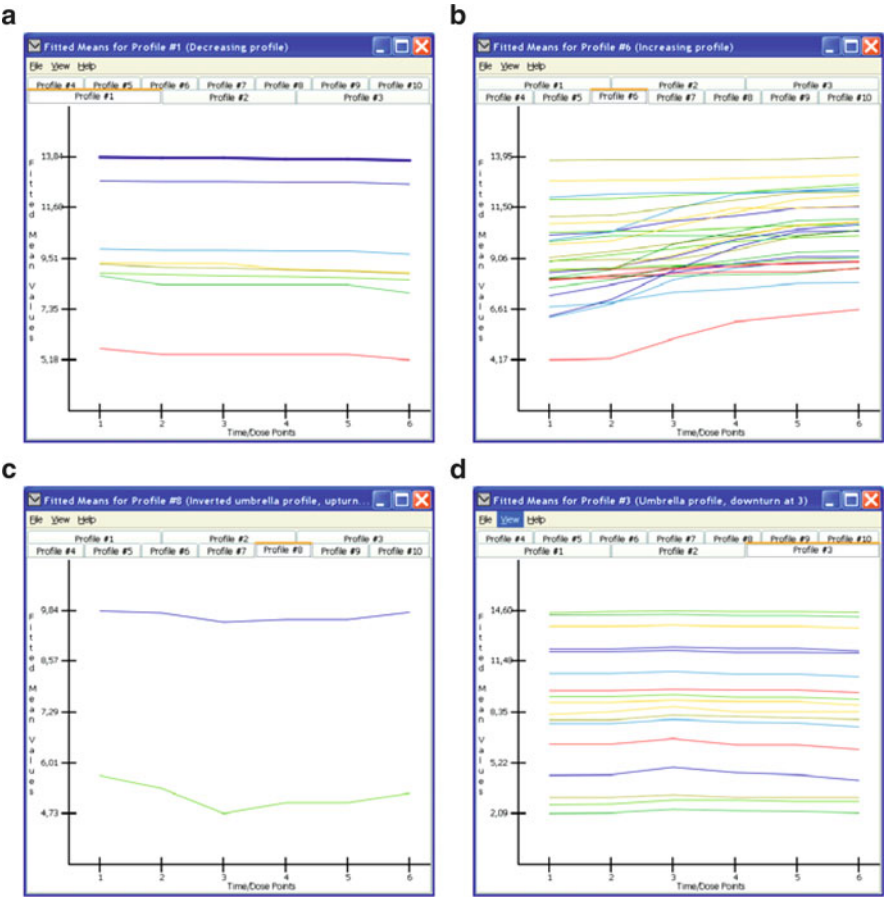

Fig. 18.7 Genes with significant profiles for the analysis of noncyclical profiles. (**a**) Increasing profiles; (**b**) decreasing profiles; (**c**) inverted umbrella profiles (upturn at 3); (**d**) umbrella profiles (downturn at 3)

18.3.1 Analysis of the Antipsychotic Experiment with the `ORIOGEN` *Package*

18.3.1.1 Analysis of Simple Order Alternatives

In Sect. 18.2.3.1, we discussed the SAM analysis for the antipsychotic experiment. In order to perform a similar analysis in the `ORIOGEN`, we need to specify that the profiles of primary interest are either increasing or decreasing. In that way, the `ORIOGEN` performs an analysis for simple order alternatives. Using the FDR level of 10% and $s_0 = s_{(5\%)}$, 146 significant genes were found to be significant (compared with 123 genes which were found in Sect. 18.2.3.1 for the $\text{FDR}_{90} = 0.1$ and $\Delta = 0.75$, see Fig. 18.3b). The significant genes for each direction are shown in Fig. 18.6.

18.3.1.2 Analysis of Partial Order Alternatives

In this section, we specify all noncyclical profiles as alternative profiles. Since there are six dose levels, there are ten possible noncyclical profiles.

```
Profile Selections:
1   Decreasing profile
2   Umbrella profile, downturn at 2
3   Umbrella profile, downturn at 3
4   Umbrella profile, downturn at 4
5   Umbrella profile, downturn at 5
6   Increasing profile
7   Inverted umbrella profile, upturn at 2
8   Inverted umbrella profile, upturn at 3
9   Inverted umbrella profile, upturn at 4
10  Inverted umbrella profile, upturn at 5
```

Using the FDR level of 10% and $s_0 = s_{(5\%)}$, 146 significant genes were found to be significant. Figure 18.7 shows examples for increasing/decreasing and umbrella profiles, with downturn and upturn at the third dose level.

References

Ernst, J., & Bar-Joseph, Z. (2006). STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinformatics, 7*, 191.

Klinglmueller, F., Tuechler, T., & Posch, M. (2011). Cross-platform comparison of microarray data using order restricted inference. *Bioinformatics, 27*(7), 953–960.

Liu, T., Lin, N., Shi, N., & Zhang, B. (2009). *Order-restricted information criterion-based clustering algorithm. Reference manual*. http://cran.r-project.org/web/packages/ORIClust/.

Otava, M., Kasim, A., & Verbeke, T. (2011). *Order restricted clustering for microarray experiments. Reference manual*. CRAN. http://cran.r-project.org/web/packages/ORCME/ORCME.pdf.

Pramana, S. Lin, D., Haldermans, P., Shkedy, Z., Verbeke, T., Göhlmann, H., et al. (2010a). IsoGene: an R package for analyzing dose-response studies in microarray experiments. *The R Journal, 2*(1), 5–12.

Pramana, S., Lin, D., & Shkedy, Z. (2010b). IsoGeneGUI package vignette. *Bioconductor*. http://www.bioconductor.org.

Peddada, S., Lobenhofer, E. K., Li, L., Afshari, C. A., Weinberg, C. R., & Umbach, D. M. (2003). Gene selection and clustering for time-course and dose-response microarray experimants using order-restricted inference. *Bioinformatics, 19*(7), 834–841.

Peddada, S., Harris, S., & Harvey E. (2005). ORIOGEN: order restricted inference for ordered gene expression data. *Bioinformatics, 21*(20), 3933–3934.

Index

D. Lin et al. (eds.), *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, Use R!, DOI 10.1007/978-3-642-24007-2,