Alberto Bernardini
Fulvio Tonon

# Bounding Uncertainty in Civil Engineering

## THEORETICAL BACKGROUND

Springer

Bounding Uncertainty in Civil Engineering

Alberto Bernardini and Fulvio Tonon

# Bounding Uncertainty in Civil Engineering

## Theoretical Background

Springer

Prof. Dr. Alberto Bernardini
Dipartimento di Costruzioni e Trasporti (DCT)
Università degli Studi di Padova
Via Marzolo 9
Padova, 35131
Italy
E-mail: alberto.bernardini@unipd.it

Prof. Dr. Fulvio Tonon
The University of Texas at Austin
Department of Civil Engineering
1 University Station C1792
Austin, TX 78712-0280
USA
E-mail: tonon@mail.utexas.edu

ISBN 978-3-642-11189-1 e-ISBN 978-3-642-11190-7

DOI 10.1007/978-3-642-11190-7

Library of Congress Control Number: 2009942254



*Typesetting:* Scientific Publishing Services Pvt. Ltd., Chennai, India

*Cover Design:* WMX Design, Heidelberg, Germany

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

# Preface

The theories described in the first part of this book summarize the research work that in past 30-40 years, from different roots and with different aims, has tried to overcome the boundaries of the classical theory of probability, both in its objectivist interpretation (relative frequencies of expected events) and in its subjective, Bayesian or behavioral view. Many compelling and competitive mathematical objects have been proposed in different areas (robust statistical methods, mathematical logic, artificial intelligence, generalized information theory). For example, fuzzy sets, bodies of evidence, Choquet capacities, imprecise previsions, possibility distributions, and sets of desirable gambles.

Many of these new ideas have been tentatively applied in different disciplines to model the inherent uncertainty in predicting a system's behavior or in back analyzing or identifying a system's behavior in order to obtain parameters of interest (econometric measures, medical diagnosis, …). In the early to mid-1990s, the authors turned to random sets as a way to formalize uncertainty in civil engineering.

It is far from the intended mission of this book to be an all comprehensive presentation of the subject. For an updated and clear synthesis, the interested reader could for example refer to (Klir 2005). The particular point of view of the authors is centered on the applications to civil engineering problems and essentially on the mathematical theories that can be referred to the general idea of a convex set of probability distributions describing the input data and/or the final response of systems. In this respect, the theory of random sets has been adopted as the most appropriate and relatively simple model in many typical problems. However, the authors have tried to elucidate its connections to the more general theory of imprecise probabilities. If choosing the theory of random sets may lead to some loss of generality, it will, on the other hand, allow for a self-contained selection of the arguments and a more unified presentation of the theoretical contents and algorithms.

Finally, it will be shown that in some (or all) cases the final engineering decisions should be guided by some subjective judgment in order to obtain a reasonable compromise between different contrasting objectives (for example safety and economy) or to take into account qualitative factors. Therefore, some formal rules of approximate reasoning or multi-valued logic will be described and implemented in the applications. These rules cannot be confined within the boundaries of a probabilistic theory, albeit extended as indicated above.

**Subjects Covered:** Within the context of civil engineering, the first chapter provides motivation for the introduction of more general theories of uncertainty than the classical theory of probability, whose basic definitions and concepts (à la Kolmogorov) are recalled in the second chapter that also establishes the nomenclature and notation for the remainder of the book. Chapter 3 is the main point of departure for this book, and presents the theory of random sets for one uncertain variable together with its links to the theory of fuzzy sets, evidence theory, theory of capacities, and imprecise probabilities. Chapter 4 expands the treatment to two or more variables (random relations), whereas the inclusion between random sets (or relations) is covered in Chapter 5 together with mappings of random sets and monotonicity of operations on random sets. The book concludes with Chapter 6, which deals with approximate reasoning techniques. Chapters 3 through 5 should be read sequentially. Chapter 6 may be read after reading Chapter 3.

**Level and Background:** The book is written at the beginning graduate level with the engineering student and practitioner in mind. As a consequence, each definition, concept or algorithm is followed by examples solved in detail, and cross-references have been introduced to link different sections of the book. Mathematicians will find excellent presentations in the books by Molchanov (2005), and Nguyen (2006) where links to the initial stochastic geometry pathway of Matheron (1975) is recalled and random sets are studied as stochastic models.

The authors have equally contributed to the book.

# Contents

# Chapter 1
# Motivation

Before embarking on studying the following chapters, motivations are provided as to why random sets are useful to formalize uncertainty in civil engineering. Pros and cons in using the theory of random sets are contrasted to more familiar theories such as, for example, the theory of random variables.

## 1.1 Why Use Random Sets?

### *1.1.1 Histograms*

Consider the case where statistical information on a quantity of interest is presented in histogram form. For example, Figure 1.1 shows the annual rainfall intensity at a certain location. It tells us that the frequency that an annual rainfall intensity be in the range between 38 and 42 inches is about 10%. One can also calculate the frequency that an annual rainfall intensity be in the range between 38 and 46 inches: this is done by summing up the frequencies relevant to the [38, 42] in. ($m_1$) and [42, 46] in. ($m_2$) intervals, i.e. $m_1 + m_2 = 10 + 24 = 34\%$. But, what if one wants to know the frequency in the 40 to 48 in. range?

A histogram gives the frequency that an event be *anywhere* in a chosen bin, even if one does not know exactly *where* in that bin. Call $m_3$ the frequency in [46, 50] in.. Given the available information, one may just consider two extreme cases. In the first extreme case, one might think that events were actually recorded only in the [38, 40] in. range for the first bin, and in the [48, 50] in. range for the third bin. As for the second bin, one does not care where the events were recorded because the [42, 46] in. range falls entirely within the [40, 48] range. In this case, the frequency in the [40, 48] in. range is equal to $m_2$, i.e. 24%.

In the second extreme case, one might think that events were actually recorded only in the [40, 42] in. range for the first bin, and in [46, 48] in. for the third bin. The frequency in the [40, 48] in. range is thus equal to $m_1 + m_2 + m_3 = 10 + 24 + 18 = 52$ %. As a result, one can only say that the frequency of the [40, 48] in. range is between 24% and 52%.

The reader has just encountered the first example of a random set, i.e. a collection of intervals (histogram bins) with weights (frequencies) attached to them. The reader has also performed the first example of calculation of upper and lower bounds on the frequency of an event of interest.

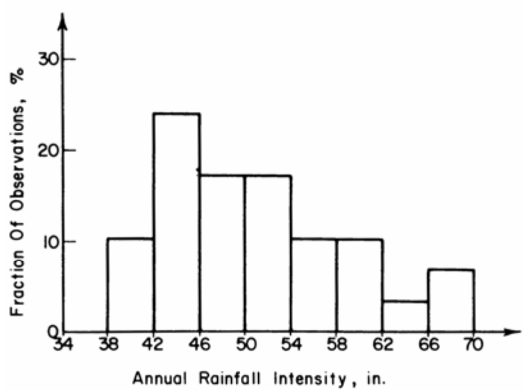



**Fig. 1.1** Histogram of rainfall intensity (Esopus Creek Watershed, NY, 1918-1946), (modified after Ang and Tang (1975), with permission)

## *1.1.2 Empirical Limitations in Data Gathering*

### 1.1.2.1 Measurements

Limitations in time and funds available for data gathering may lead to imprecise or incomplete measurements. Consider, for example, the measurement of the uniaxial compressive strength (UCS) of an intact rock specimen using the Schmidt hammer. The two quantities are correlated by the chart in Figure 1.2 presented in the Suggested Methods of the International Society of Rock Mechanics (ISRM 1978) and by Harrison and Hudson (1997). Since a single hammer reading yields an interval of UCS values, a set of readings yields a set of intervals, each with its own frequency. This set of intervals, each with its own frequency, is a random set.

With a large enough budget and timeframe, laboratory tests may be carried out that do not exhibit this imprecision. However, the low cost and short duration of Schmidt hammer measurements allow one to take many more readings than lab tests and thus obtain a more representative sample.

Additionally, in the presence of inhomogeneous intact rock, repeated Schmidt hammer readings are invaluable to determine the extents of a homogeneous zone. Finally, regardless of the available budget and timeframe, the Schmidt hammer is the only piece of equipment that allows one to measure the joint compression strength (JCS) in discontinuities, especially if weathered. The JCS is then used to evaluate the shear strength of rock discontinuities (Barton 1976).

Examples of correlations are replete in geotechnical engineering practice, especially when using the results of *in situ* tests. Figure 1.3 and Figure 1.4 show two examples: one for deformation parameters to be used in consolidation settlement calculations, and one for friction angle to be used in stability calculations, respectively. Even in this case, laboratory tests may yield more precise results, but one needs to account for disturbance of lab specimens. Additionally, as occurred in rock, the number of lab tests is always small when compared to the large number of data points obtainable using correlations.

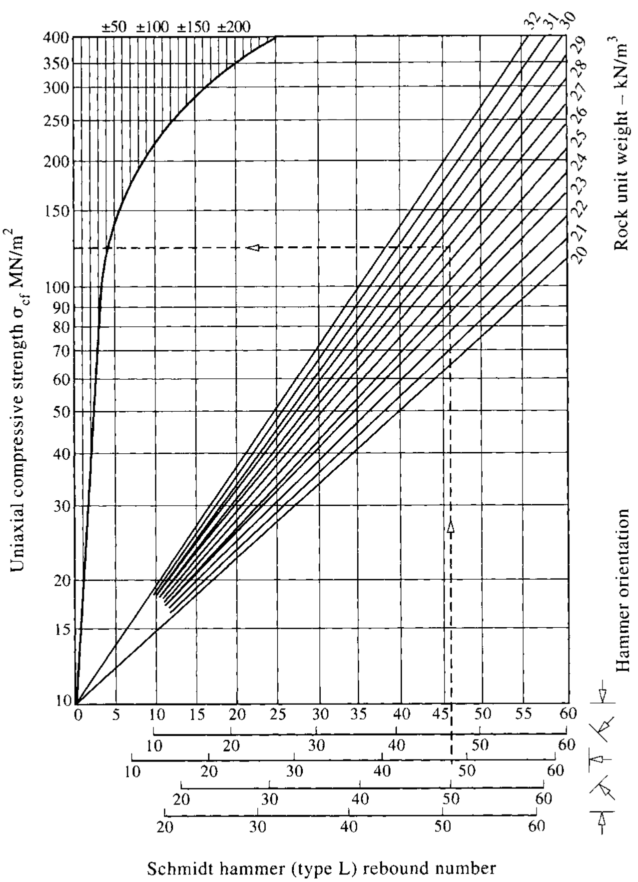



**Fig. 1.2.** Correlation between Schmidt hammer rebound number (r) and uniaxial compressive strength for different rock densities, (after Hudson and Harrison (1997), with permission)

| Compression index, $C_c$ | Comments | Source/Reference |
|---|---|---|
| $C_c = 0.009(w_L - 10)$ ($\pm 30\%$ error) | Clays of moderate $S_t$ | Terzaghi and Peck (1967) |
| $C_c = 0.37(e_o + 0.003w_L + 0.0004w_N - 0.34)$ | 678 data points | Azzouz et al. (1976) |
| $C_c = 0.141G_s\left(\frac{\gamma_{sat}}{\gamma_{dry}}\right)^{2.4}$ | All clays | Rendon-Herrero (1983) |
| $C_c = 0.0093w_N$ | 109 data points | Koppula (1981) |
| $C_c = -0.0997 + 0.009w_L + 0.0014I_P + 0.0036w_N + 0.1165e_o + 0.0025C_P$ | 109 data points | Koppula (1981) |
| $C_c = 0.329[w_N G_s - 0.027w_P + 0.0133I_P(1.192 + C_P/I_P)$ | All inorganic clays | Carrier (1985) |
| $C_c = 0.046 + 0.0104I_P$ | Best for $I_P < 50\%$ | Nakase et al. (1988) |
| $C_c = 0.00234w_L G_s$ | All inorganic clays | Nagaraj and Srinivasa Murthy (1985, 1986) |
| $C_c = 1.15(e_o - 0.35)$ | All clays | Nishida (1956) |
| $C_c = 0.009w_N + 0.005w_L$ | All clays | Koppula (1986) |
| $C_c = -0.156 + 0.411e_o + 0.00058w_L$ | 72 data points | Al-Khafaji and Andersland (1992) |
| **Recompression index, $C_r$** | | |
| $C_r = 0.000463w_L G_s$ | | Nagaraj and Srinivasa Murthy (1985) |
| $C_r = 0.00194(I_P - 4.6)$ | Best for $I_P < 50\%$ | Nakase et al. (1988) |
| $= 0.05$ to $0.1C_c$ | In desperation | |
| **Secondary compression index, $C_\alpha$** | | |
| $C_\alpha = 0.00168 + 0.00033I_P$ | | Nakase et al. (1988) |
| $= 0.0001w_N$ | | NAFAC DM7.1 p. 7.1–237 |
| $C_\alpha = 0.032C_c$ | $0.025 < C_\alpha < 0.1$ | Mesri and Godlewski (1977) |
| $= 0.06$ to $0.07C_c$ | Peats and organic soil | Mesri (1986) |
| $= 0.015$ to $0.03C_c$ | Sandy clays | Mesri et al. (1990) |

*Notes:* 1. Use $w_L, w_P, w_N, I_P$ as percent, not decimal.
2. One may compute the in situ void ratio as $e_o = w_N G_s$ if $S \to 100$ percent.
3. $C_p$ = percent clay (usually material finer than 0.002 mm).
4. Equations that use $e_o$, $w_N$, and $w_L$ are for both normally and overconsolidated soils.

**Fig. 1.3** Correlation equations for the compression and recompression index of soils, (after Bowles (1996), with permission)

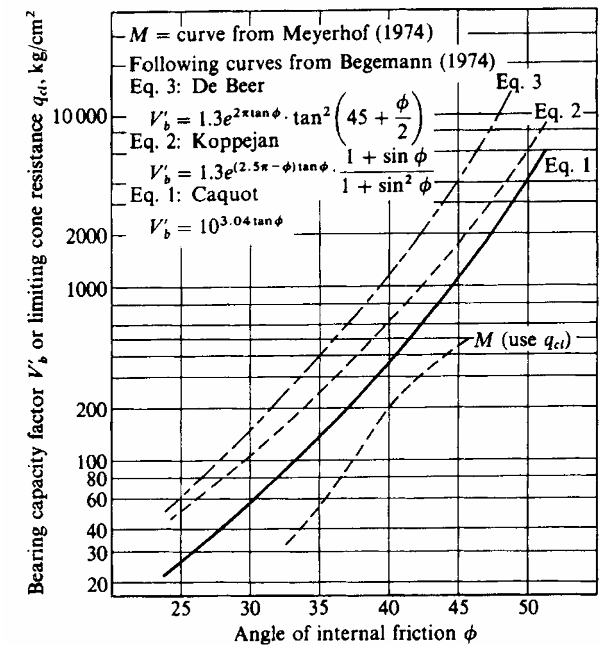



**Fig. 1.4** Correlations between cone penetrometer data and friction angle of soils. $V'_b = q'_c/p'_0$, where $q'_c$ = (cone resistance – pore water pressure); $p'_0$ = initial vertical effective stress, (after Bowles (1996), with permission)

### 1.1.2.2 Experts

Another empirical limitation occurs when eliciting information from experts. In typical risk assessment procedures (e.g., those adopted by the US Bureau of Reclamation and by the International Tunneling Association), experts convey their information on an event of interest (e.g., failure of a dam component) through linguistic terms, which are then converted into numerical probability intervals as per Figure 1.5. Notice, however, the very large discrepancy between the values in the two tables in Figure 1.5; this discrepancy may be explained by considering that the values in Figure 5a refer to the construction period, whereas the values in Figure 5b are not referred to a time interval. By polling a group of experts, a set of probability intervals will be collected. This information can be converted into a random set.

| Expression | Single-number probability equivalent, % (median of responses) | Specified range, % (median upper and lower bounds) |
|---|---|---|
| Almost impossible | 2 | 0 to 5 |
| Very improbable | 5 | 1 to 15 |
| Very unlikely | 10 | 2 to 15 |
| Very low chance | 10 | 5 to 15 |
| Improbable | 15 | 5 to 20 |
| Unlikely | 15 | 10 to 25 |
| Low chance | 20 | 10 to 20 |
| Possible | 40 | 40 to 70 |
| Medium chance | 50 | 40 to 60 |
| Even chance | 50 | 45 to 55 |
| Probable | 70 | 60 to 75 |
| Likely | 70 | 65 to 85 |
| Very possible | 80 | 70 to 87.5 |
| Very probable | 80 | 75 to 92 |
| High chance | 80 | 80 to 92 |
| Very likely | 85 | 75 to 90 |
| Very high chance | 90 | 85 to 99 |
| Almost certain | 90 | 90 to 99.5 |

a)

| Frequency class | Interval | Central value | Descriptive frequency class |
|---|---|---|---|
| 5 | >0.3 | 1 | Very likely |
| 4 | 0.03 to 0.3 | 0.1 | Likely |
| 3 | 0.003 to 0.03 | 0.01 | Occasional |
| 2 | 0.0003 to 0.003 | 0.001 | Unlikely |
| 1 | <0.0003 | 0.0001 | Very unlikely |

The central value represents the logarithmic mean value of the given interval.

b)

**Fig. 1.5** a) Numerical responses and ranges for 18 probability expressions (after Vick (1999), and Reagan et al. (1989)); b) frequency of occurrence during a tunnel's construction period, (after Eskesen et al. (2004), with permission)

## *1.1.3 Modeling*

### 1.1.3.1 Different Competing Models

In order to gain confidence in their predictive ability, engineers instinctively use two or more models of the same engineering system. In the simplest case, these models may simply be two different analytical formulations, but in the more complex cases they can be completely independent studies.

As a first example, consider the calculation of the bearing capacity for a footing. Several bearing capacity models have been proposed in the literature, and Figure 1—6 shows the comparison between the set of values calculated using a set of five different models and relevant test results. When a set of models are used, a set of results (bearing capacity values) is obtained for any vector of input values (e.g., $q_{ult} \in \{9.4, 8.2, 7.2, 8.1, 14.0\}$ kg/cm$^2$ for Test 1 in Figure 1.6). If the vector of input values, v*, is not deterministic, but has a probability of occurrence equal to, say, 30%, then the set of bearing capacity values obtained using v* has probability equal to 30%. Proceeding in this fashion for all possible input vectors, one obtains sets of bearing capacity values with a probability mass attached to each set of bearing capacity values, i.e., a random set.

| Bearing-capacity method | Test 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| | $D$ = 0.0 m | 0.5 | 0.5 | 0.5 | 0.4 | 0.5 | 0.0 | 0.3 |
| | $B$ = 0.5 m | 0.5 | 0.5 | 1.0 | 0.71 | 0.71 | 0.71 | 0.71 |
| | $L$ = 2.0 m | 2.0 | 2.0 | 1.0 | 0.71 | 0.71 | 0.71 | 0.71 |
| | $\gamma$ = 15.69 kN/m$^3$ | 16.38 | 17.06 | 17.06 | 17.65 | 17.65 | 17.06 | 17.06 |
| | $\phi$ = 37°(38.5°) | 35.5(36.25) | 38.5(40.75) | 38.5 | 22 | 25 | 20 | 20 |
| | $c$ = 6.37 kPa | 3.92 | 7.8 | 7.8 | 12.75 | 14.7 | 9.8 | 9.8 |
| Milović (tests) | | | | $q_{ult}$, kg/cm$^2$ = | 4.1 | 5.5 | 2.2 | 2.6 |
| Muhs (tests) | $q_{ult}$ = 10.8 kg/cm$^2$ | 12.2 | 24.2 | 33.0 | | | | |
| Terzaghi | $q_{ult}$ = 9.4* | 9.2 | 22.9 | 19.7 | 4.3* | 6.5* | 2.5 | 2.9* |
| Meyerhof | 8.2* | 10.3 | 26.4 | 28.4 | 4.8 | 7.6 | 2.3 | 3.0 |
| Hansen | 7.2 | 9.8 | 23.7* | 23.4 | 5.0 | 8.0 | 2.2* | 3.1 |
| Vesić | 8.1 | 10.4* | 25.1 | 24.7 | 5.1 | 8.2 | 2.3 | 3.2 |
| Balla | 14.0 | 15.3 | 35.8 | 33.0* | 6.0 | 9.2 | 2.6 | 3.8 |

*After Milovic (1965) but all methods recomputed by author and Vesić added.

*Notes:*

1. $\phi$ = triaxial value ( ) = value adjusted as $\phi_{ps} = 1.5\phi_{tr} - 17$ (Eq. 2-57).
2. Values to nearest 0.1.
3. $\gamma$, $c$ converted from given units to above values.
4. All values computed using computer program B-31 with subroutines for each method. Values all use $\phi_{ps}$ for $L/B > 1$.
5. * = best → Terzaghi = 4; Hansen = 2; Vesić and Balla = 1 each.

**Fig. 1.6** Comparison of bearing capacities computed using different methods with experimental values, (after Bowles (1996), with permission)

#### 1.1.3.2 Upper and Lower Bounds in Plastic Limit Analysis

For elasto-perfectly plastic solids with no dilatancy, limit analysis yields static (lower) and kinematic (upper) load multipliers. Greenberg-Prager theorem then assures us that the load multiplier that causes failure is the largest static multiplier and the smallest kinematic multiplier. Oftentimes, it is not possible to calculate the largest static multiplier and the smallest kinematic multiplier, and thus the engineer is left with upper and lower bounds on the load multiplier. Consider, for example, the pressure $q$ that must be exerted on a tunnel's face to ensure its stability. In an elasto-perfectly plastic ground with Mohr-Coulomb failure criterion (cohesion = $c$, and friction angle = $\varphi$), one has:

$$q = Q_\gamma \cdot \gamma \cdot a + Q_s \cdot q_s + (Q_s - 1) \cdot c \cdot ctg(\varphi) \tag{1.1}$$

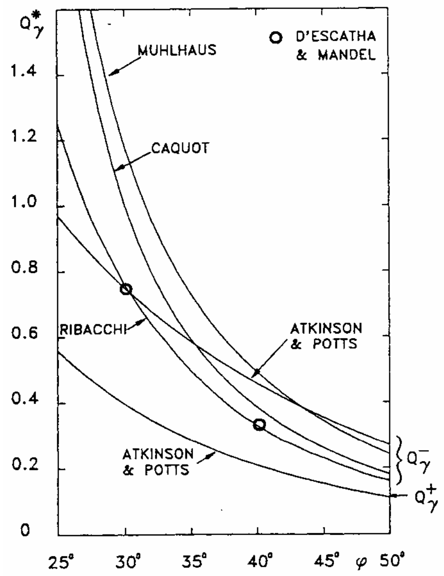



**Fig. 1.7** Coefficients $Q_\gamma$ obtained using limit analysis, (after Ribacchi (1993), with permission)

where: $Q_\gamma$ = coefficient from limit analysis, $\gamma$ = unit weight of the ground, $a$ = tunnel radius, $Q_s = \left(a/H\right)^{N-1}$, $N = \left(1+\sin\varphi\right)/(1-\sin\varphi)$, $H$ = tunnel cover, $q_s$ = pressure on ground surface.

Figure 1.7 shows coefficients $Q_\gamma$ obtained using kinematic analysis ($Q_\gamma^+$), and coefficients $Q_\gamma$ obtained using static analysis ($Q_\gamma^-$) versus the friction angle. The different solutions for $Q_\gamma^-$ originate from different assumptions on the equilibrated stress distribution at failure.

If the friction angle is not known deterministically, but one knows that the probability of $\varphi^*$ is, say, 60%, then one can calculate upper and lower bounds (i.e., an interval) on the face pressure $q^*$. This pressure will have probability equal to 60%. By calculating the face pressure intervals for all possible values of the friction angle, one obtains a collection of intervals, each one with its own probability, i.e. a random set.

#### 1.1.3.3 Discretization Errors

One of the first uses of digital computers was to approximately simulate physical systems by numerically solving differential equations. This approach leads to numerical computation that is at least three levels removed from the physical world represented by those differential equations:

1) One models a physical phenomenon using a differential equation (or a system of differential equations) or a variational principle.
2) Then, one obtains the algebraic forms of the differential equation(s) or variational principle by forcing them into the mold of discrete time and space; and
3) Finally, in order to commit those algebraic forms to algorithms, one projects real-valued variables onto finite computer words, thus introducing round-off during computation and truncation.

Errors included in Steps 1 through 3 are to be addressed during verification and validation of numerical models (Oberkampf *et al*., 2003). A large body of literature has been devoted to estimating the discretization errors introduced in Step 2. For example, Dow (1998), Babuska and Strouboulis (2001), Oden *et al*. (2005), and an issue of the journal *Computer Methods in Applied Mechanics and Engineering* (2006) give an overview of results in the finite element discretization method. Peraire and coworkers have developed algorithms for calculating guaranteed bounds on these errors (Sauer-Budge *et al*., 2004; Xuan *et al*., 2006); however, their calculations are performed in floating-point arithmetic. Figure 1.8 illustrates the discretization error bounds for the Laplace equation in an L-shaped domain: the finite element solution is comprised in the error interval, whose width decreases quadratically with the mesh size.

Figure 1.9 shows bounds on displacements and tractions for a notched specimen: although convergence is not quadratic, it is still superlinear.

Consider the displacement in Figure 1.9c and fix the mesh size, *h*: if the vector of input values, v*, is not deterministic, but has a probability of occurrence equal to, say, 70%, then the interval of displacement values obtained using v* has probability equal to 70%. Proceeding in this fashion for all possible input vectors, one obtains a collection of displacement intervals with a probability mass attached to each displacement interval, i.e., a random set.

Errors involved in Step 3 have been vigorously attacked by the "reliable computing" community using interval analysis started by Warmus (1956) and Moore (1966); the reader is referred to the journal *Reliable Computing* (formerly *Interval Computations*) and to the web site (www.cs.utep.edu/interval-comp/main.html) for up-to-date information. One can repeat the same reasoning above to obtain a random set for any quantity of interest.

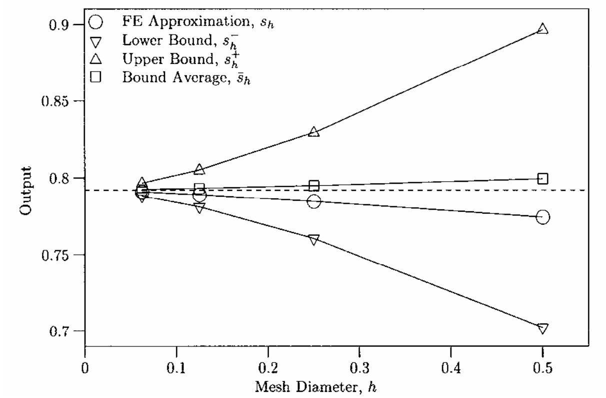



**Fig. 1.8** Error bounds on the discretized solution of the Laplace equation, (after Sauer-Budge et al. (2004)). Copyright ©2004 Society for Industrial and Applied Mathematics. Reprinted with permission. All rights reserved

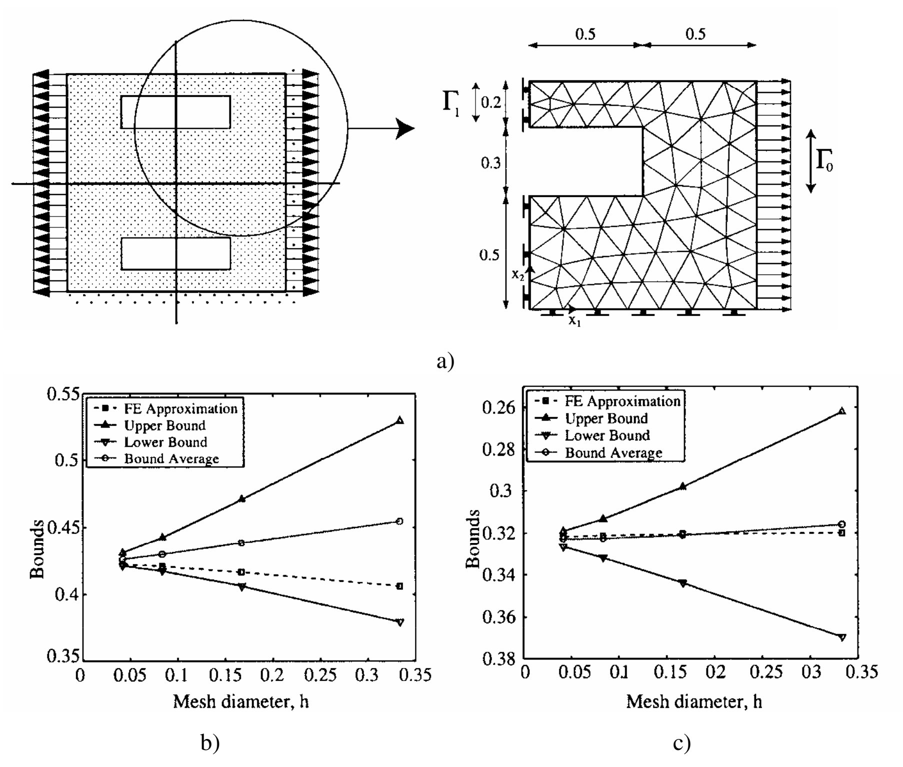



**Fig. 1.9** a) Model problem and initial mesh; b) average normal displacement over the boundary $\Gamma_0$; c) integrated normal component of the traction in $\Gamma_1$, (after Parès et al. (2006), with permission). Copyright ©2004 Society for Industrial and Applied Mathematics. Reprinted with permission. All rights reserved

## 1.2 Imprecise Information Cannot Give Precise Conclusions

The most attractive advantage in using the theories described in this book is the possibility of taking into account the available information about the engineering systems to be evaluated, without any other unjustified hypothesis.

For example, if some data obtained through imprecise instruments are given (and in fact really every measurement has a bounded precision), it is not reasonable to force the interval of confidence to a single central value; or in the case of a sample of measurements, it is not reasonable to force the statistics of intervals to a conventional histogram or finally to a precise probability distribution.

In other cases, the available information could consist of a very poor estimation of some parameters of the unknown probabilistic distribution: for example the mean value or an interval containing the mean value. Sometimes this information derives from subjective judgment or from opinions of experts, and is therefore characterized by the unavoidable uncertainty inherent in every human assessment.

Forcing these opinions to a particular probabilistic distribution (for example, a lognormal distribution) with precise parameters seems to be unjustified; but, on the contrary, it is unreasonable to disregard all sources of information that cannot be forced to a precise probabilistic distribution in the analysis or in decision-making.

Even if one assumes that precise distributions can be attached to each random variable in the probabilistic approach to engineering problems, frequently very little evidence is available about the correlation between these random variables. Without any well-grounded motivation, independence is oftentimes assumed in order to calculate the joint distribution. But in many cases this hypothesis seems to be unrealistic, or at least not justified. This assumption, however, in many cases strongly influences the final conclusions of the analysis, and sometimes it is not on the safe side. For example, consider the load, $L$, on a ground-floor column of a multistory building (Ang and Tang 1975, page 195). The load contribution from each floor to $L$ is an increasing function of the correlation among floor loads; therefore, the assumption of statistical independence would yield results on the unsafe side with respect to any other hypothesis of positive correlation.

The unrealistic character of many assumptions supporting most applications of the classical probabilistic methods to civil engineering systems is particularly evident when one then considers the computational effort required to evaluate the performance or the safety of these systems in complex real-world applications. Closed-form solutions for propagating the probabilistic information from the input random variables to the system response are rarely available. Only numerical solutions (e.g., Monte Carlo simulations of large-scale finite element models) can then be used: the computational time and effort necessary to obtain such an approximation could be dramatically large, but at the end the conclusion may be of questionable validity because of the initial (unwittinly added) assumptions on the probabilistic information.

A further limitation of the probabilistic approach sometimes appears when model uncertainties are combined with a precise joint distribution for the random variables of the considered engineering system. Recall, for example, the bounding intervals in the evaluation of collapse loading of elastic-perfectly plastic structures using limit analysis (Figure 1.7), or the unavoidable errors when a continuous model is forced to a discrete one in finite element procedures (Figure 1.8).

These problems appear when the deterministic modeling of a system's behavior yields a multi-valued mapping from the space of the input variables to the space of the response output variables. Validation of the obtained results and calibration of a reasonable compromise between competitive models of different complexity cannot be performed without taking into account all the available information and the actual evidence required to support design choices or decision-making in the management of civil infrastructures.

## 1.3 Describing Void Information

The power of the approach considered in this book is also apparent when considering cases of total lack of information. In this context, the probabilistic approach seems to require or suggest the selection of a particular precise probabilistic distribution, for example based on the so called "Principle of Indifference" or "Maximum Entropy".

The literature on the paradoxical conclusions that can derive from this choice is very rich. Here, we discuss a simple way to gain money using the "Principle of Indifference" (Ben Haim 2004).

Two envelopes containing a positive amount of money are offered for your choice and you know only that one envelope contains twice as much money as the other envelope. You choose one envelope and find \$ 100 inside. Now, you are given the option to exchange the envelope for the other, which could contain either \$ 50 or \$ 200. On the basis of the "Principle of Indifference", you could assign equal probabilities (1/2) to both possible results and try to make the best decision by evaluating the expected reward:

$$E(\text{Reward}) = 50 \cdot \frac{1}{2} + 200 \cdot \frac{1}{2} = 125\$ \tag{1.2}$$

The expected percentage increase of the reward (25%) does not depend on the value (\$ 100) that you have found in the first envelope: therefore, without opening the first envelope, you could decide to exchange it for the second, and so augment on average the reward by 25%. You can also try to gain more money exchanging the envelopes again, again and again…

Considering the same story within the optics of random sets, you can only admit that, on the basis of the available information, the overall probability of the two alternative rewards (\$ 200 or \$ 50) is exactly equal to 1. Therefore, you have a histogram with one bin covering the two rewards, and the probability of each reward is between 0 and 1. The lower and upper bounds of the expected reward are:

$$\underline{E}(\text{Reward}) = 50 \cdot 1 + 200 \cdot 0 = 50\$ \tag{1.3}$$

$$\overline{E}(\text{Reward}) = 50 \cdot 0 + 200 \cdot 1 = 200\$ \tag{1.4}$$

The bounds contain the previous result, but now the situation is clear: the choice is in your hands!

## 1.4 Bounding Uncertainty

Recalling Hamlet's words, a wise engineer, and perhaps any reasonable person, should be suspicious of a perfectly precise proposition about future events:

"*There are more things in heaven and hearth, Horatio, than are dreamt of in your philosophy*"

The authors do not think that random sets or imprecise probabilities could help in solving this dramatic philosophical question. However, they suggest that the *true* solution does not exist, or, if it does, it can only be bounded by incomplete or imprecise information through uncertain mathematical and physical models.

Additionally, by knowing these bounds, the engineer may ascertain if what he/she knows about the expected behavior of the system is enough to make final decisions about the design, safety assessment or management of the system. When the reply is affirmative, any further investigation is not justified, or is only motivated by personal curiosity or higher engineering fees!

On the contrary, when the reply is negative, new or more precise information is necessary, or more sophisticated models should be employed to narrow the bounds of the final evaluations.

# Chapter 2
# Review of Theory of Probability and Notation

The basic definitions of a probability space are briefly reviewed, thus introducing the notation useful for the theoretical developments presented in the book. Particular attention is given to continuous and discrete random variables and to the concept of expectation of a random variable, defined through both Lebesque and Stieltjes integrals. The theory is extended to joint probability spaces and random vectors.

## 2.1 Probability Measures

The following is mainly taken from (Burrill 1972, Cariolaro and Pierobon 1992, Fetz and Oberguggenberger 2004, Papoulis and Pilai 2002); for additional details, the reader is referred to (Halmos 1950, Kolmogorov 1956, Loève 1977 and 1994,). Let $S$ be any set, and let $A^C$ indicate the complement of set $A$. A *σ-algebra* $\mathcal{S}$ on $S$ is a nonempty collection of subsets of $X$ such that the following conditions hold:

1. $S \in \mathcal{S}$.
2. $A \in \mathcal{S} \Rightarrow A^C \in \mathcal{S}$.
3. If $\{A^i\}$ is a sequence of elements of $\mathcal{S}$, then $\bigcup_i A^i \in \mathcal{S}$.

If $\mathcal{C}$ is any collection of subsets of $S$, then one can always find a σ-algebra containing $\mathcal{C}$, namely the power set (set of all subsets) of $S$. By taking the intersection of all σ-algebras containing $\mathcal{C}$, we obtain the smallest such σ-algebra. We call the smallest σ-algebra containing $\mathcal{C}$ the *σ-algebra generated by* $\mathcal{C}$. On the set of real numbers, $\mathbb{R}$, the σ-algebra generated by $\mathcal{C} = \{(-\infty, a]: a \in \mathbb{R}\}$ is called the *Borel σ-algebra*, $\mathcal{B}$, and contains all intervals of $\mathbb{R}$. If $S$ is finite and $|S|$ is the cardinality, the σ-algebra generated by $S$ is the power set of $S$, with cardinality $2^{|S|}$.

A *measurable space* is a pair $(S, \mathcal{S})$. Given a *measurable space* $(S, \mathcal{S})$, a *probability measure*, $P$, on $\mathcal{S}$ is a mapping $\mathcal{S} \to [0, 1]$ such that:

$$P(\varnothing)=0,\ \ P(S)=1,\ \ P\left(\bigcup_i A^i\right)=\sum_i P\left(A^i\right) \tag{2.1}$$

whenever subsets $A^i \in \mathcal{S}$ are disjoint.

A *probability space* is a triple $(S, \mathcal{S}, P)$. If $S = \{s^1,\ldots, s^n\}$ is finite, or more generally $\{s^1,\ldots, s^n\}$ is a finite partition of $S$ through the "*singletons*" or "*elementary events*" $s^i$ ($s^i \cap s^j = \varnothing$ and $\cup_i s^i = S$), $P$ on the σ-algebra generated by $\mathcal{C} = \{s^1,\ldots, s^n\}$ can be assigned by using the *probability of elementary events*, $\{s^i\}$, $P(s^i):=P(\{s^i\})$, which has to satisfy the two conditions:

$$P(s^i)\geq 0,\ \ \sum_{i=1}^n P\left(s^i\right)=1 \tag{2.2}$$

Since elementary events are disjoint, the probability of $T \subseteq S$ is calculated using Eq. (2.1):

$$P(T)=P\left(\bigcup_{s^i\in T}\{s^i\}\right)=\sum_{s^i\in T} P\left(\{s^i\}\right)=\sum_{s^i\in T} P\left(s^i\right) \tag{2.3}$$

A *Borel measure* is a probability measure on $\mathbb{R}$ such that its σ-algebra contains the Borel σ-algebra, $\mathcal{B}$. A *point mass* or *Dirac measure* at $s^0 \in S$ is the measure, $\delta_0$, concentrated at $s^0$, i.e. such that $\delta_0(A) = 1$ if $s^0 \in A$ and $\delta_0(A) = 0$ if $s^0 \notin A$, $A\in \mathcal{S}$.

Let $T^2 \in \mathcal{S}$, and $P(T^2)>0$. The *conditional probability* of $T^1\in\mathcal{S}$ conditioned on $T^2$ is defined as

$$P(T^1 | T^2) := P(T^1 \cap T^2)/P(T^2) \tag{2.4}$$

Let $P(T^2)>0$ and $P(T^1)>0$. From Eq. (2.4): $P(T^1 \cap T^2) = P(T^1|T^2)\, P(T^2) = P(T^2|T^1)\, P(T^1)$.

Thus (*Bayes' Theorem*):

$$P(T^1 | T^2) = P(T^2 | T^1)\, P(T^1)/ P(T^2) \tag{2.5}$$

If the occurrence of $T^1$ does not affect the probability of occurrence of $T^2$, the two sets (events) are said *statistically independent*.

Therefore: $P(T^1|T^2) = P(T^1)$ and $P(T^2|T^1) = P(T^2)$; moreover from eq. (2.4):

$$P(T^1 \cap T^2)= P(T^1)\, P(T^2) \tag{2.6}$$

Note that alternatively Eq. (2.6) could be assumed as defining statistical independence, from which identities of conditional to unconconditional probabilities follow.

If subsets $\{T^i\}$ are a partition of *S*, then for any subset *T*, $T=\cup_i T\cap T^i$, and Eqs. (2.1) and (2.4) give (*Total Probability Theorem*):

$$P(T) = \sum_i P(T\cap T^i) = \sum_i P(T \mid T^i)\, P(T^i) \tag{2.7}$$

Eqs. (2.5) and (2.7) define the Bayes' rule for updating a probability space (for example the probability of any singleton $\{s^i\}$) observing the occurrence of an event *B*, when the conditional probability $P(B|\{s^i\})$ are known

$$P_{POSTERIOR}\left(s^i\right) = P\left(\{s^i\} \mid B\right) = \frac{P\left(B \mid \left\{s^i\right\}\right)}{\sum_j P\left(B \mid \left\{s^j\right\}\right) P_{PRIOR}\left(s^j\right)} P_{PRIOR}\left(s^i\right) \tag{2.8}$$

More generally the posterior updated probabilities can be calculated when a *likelihood function* $L(s^i)$ proportional to $P(B|\{s^i\})$ is known for the observed event B or also for the observation $x$ on a *sampling space X* where *likelihood values* proportional to conditional probabilities $P(x|s^i)$ are known:

$$P_{POSTERIOR}\left(s^i\right) = P\left(\{s^i\} \mid L\right) = \frac{L\left(s^i\right) P_{PRIOR}\left(s^i\right)}{\sum_j L\left(s^j\right) P_{PRIOR}\left(s^j\right)} \tag{2.9}$$

## 2.2 Random Variable

Given two measurable spaces $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$, a function $g : S_1 \rightarrow S_2$ is *measurable* if, for every $T \in \mathcal{S}_2$, $A = g^{-1}(T) \in \mathcal{S}_1$. The particular case $(S_2, \mathcal{S}_2)$, $= (\mathbb{R}, \mathcal{B})$ is of great relevance. Let $(S, \mathcal{S}, P)$ be a probability space; a real function $x : S \rightarrow \mathbb{R}$, defined on $S$ is a *random variable* on $(S, \mathcal{S}, P)$ if $x(s)$ is *Borel-measurable*, i.e. if, for every $a \in \mathbb{R}$, $\{s\colon x(s) \leq a\} \in \mathcal{S}$.

The (*cumulative*) *distribution* (CDF) of a random variable on $(S, \mathcal{S}, P)$ is the function $F_x : \mathbb{R} \rightarrow [0, 1]$, $a \mapsto P\left(\left\{s : x(s) \leq a\right\}\right)$; the CDF allows one to calculate the dependent probability $P_x$ that $x$ be in any Borel set. A random variable, $x$, is *continuous* if $F_x$ is continuous; its *probability density* (pdf) of $x$ is

$$f_x = \frac{dF_x}{da} \tag{2.10}$$

Otherwise, let $B$ be the set of discontinuity points of $x$ (they are either finite or infinitely numerable) and let $p_x(a) := P(\{s\colon x(s) = a\}) = F_x(a) - F_x(a^-) > 0$ be the discontinuity jump at $a \in B$; if

$$\sum_{a \in B} p_x(a) = 1 \tag{2.11}$$

then $x$ is a discrete random variable and $p_x$ is called the *mass distribution*; $p_x$ is not a probability measure, in fact it is not even defined on a σ-algebra. $x$ is finite if $B$ is finite: in this case, $p_x$ allows one to calculate the probability of any subset of $B$ using Eq. (2.1) in a way similar to the probability of elementary events (Eq. (2.2)).

In many numerical engineering applications the space $S$ could be a subset or a partition of the real numbers $\mathbb{R}$, and the probability $P$ is defined through the *probability of elementary events* (or *singletons*, $P(s^i) := P(\{s^i\})$; hence, for the discrete random variable defined by the identity $x(s) = s$, the mass distribution equals the probabilities of the ordered elementary events. In many examples presented in the book this hypothesis is implicitly assumed.

In order to understand the concept of expectation $E[x]$ of a random variable $x$, one needs to introduce some more notions. For a continuous random variable the expectation is defined by means a (Riemann) integral, supposed absolutely convergent, of $x$ multiplied by the density function $f_x$:

$$E[x] = \int_{-\infty}^{+\infty} a\, f_x(a)\, da \tag{2.12}$$

This definition can be extended to discrete random variable by summation, supposed absolutely convergent if $|B| = \infty$, of $x$ multiplied the mass distribution $p_x$:

$$E[x] = \sum_{a \in B} a\, p_x(a) \tag{2.13}$$

In more general terms, the definition of expectation should be given through the Lebesque integral on the original probability space ($S$, $\mathcal{S}$, $P$) or, alternatively, by the Stieltjes integral on the dependent probability space ($\mathbb{R}$, $\mathcal{B}$, $P_x$).

Let $A \in \mathcal{S}$. The *characteristic function* (or *"indicator"* $I_A$) of the set $A$, $\chi_A(s)$: $S \to \{0, 1\}$ is defined as $\chi_A(s) = 1$ if $s \in A$, $\chi_A(s) = 0$ if $s \notin A$. Observe that $\chi_A$ is a discrete random variable with $B = \{0, 1\}$ and Eq. (2.13) demonstrates that $E[\chi_A] = P(A)$.

Let $\mathcal{C} = \{A^1,\ldots, A^n\}$ be a finite partition of $S$: a *simple function* is a finite linear combination of characteristic functions of the form $x_j(s) = \sum_i a_j{}^i \chi_{A^i}$ where $a^i \in \mathbb{R}$, $A^i \in \mathcal{S}$. A simple function is a discrete random variable with finite set $B = \{a^1,\ldots, a^n\}$; the expectation is given by:

$$E[x] = \sum_{i=1}^{n} a^i\ P\left(A^i\right) \tag{2.14}$$

Given a probability space ($S$, $\mathcal{S}$, $P$), a function $f$: $S \to \mathbb{R}$ is said to be $P$-measurable (or $\mathcal{S}$-measurable) if $f$ is pointwise the limit of a monotonic not decreasing sequence of simple functions $x_j$. It is possible to demonstrate that any non negative measurable function $x$ is pointwise the limit of a monotonic not decreasing sequence of simple functions $x_j$, i.e. it is $P$-measurable (e.g., (Hunter and Bruno 2001), page 343, Theorem 12.26). Hence $E[x_j]$ is a monotonic not decreasing sequence of real numbers converging to the *Lebesque integral* of $x$ with respect to $P$ defined as

$$\int_S x(s)\,dP(s) := \sup_j \left\{\sum_i a_j{}^i P\left(A^i\right)\right\} \tag{2.15}$$

For a general $x$, the positive and negative parts are considered separately: the Lebesque integral equals the difference between the two Lebesque integrals of the positive and negative parts, supposing that they are not both converging to $+\infty$ (otherwise the function does not admit Lebesque integral).

On the other hand, in the dependent probability space ($\mathbb{R}$, $\mathcal{B}$, $F_x$), it is possible to demonstrate that the expectation can be evaluated through the the Stieltjes integral of $x$ with weight function $F_x$ on the interval $[a_0, a_n]$:

$$\int_{a_0}^{a_n} a\ dF_x(a) = \lim_{\varepsilon \to 0} \sum_{i=0}^{n-1} a^{i'} \left(F_x(a^{i+1}) - F_x(a^i)\right) \tag{2.16}$$

where $a^0 < a^1 < \ldots < a^i < \ldots < a^n$ defines a partition of $[a_0, a_n]$, $a^{i'} \in (a^i, a^{i+1}]$ and $\varepsilon$ is the maximum amplitude of the partition. The integral can be extended to the entire $\mathbb{R}$ by considering the limits $a_0 \to -\infty$, $a_n \to +\infty$. When $F_x$ *is* continuous and hence the probability density function is defined by Eq. (2.10), the Stiltjes integral is equivalent to the Rieman integral.

The result can be extended to a *function g of the random variable x*. Let $x$ a random variable on $(S, \mathcal{S}, P)$ and $g: \mathbb{R} \to \mathbb{R}$ a real measurable function (generally a Borel measurable function). Then $y = g(x(s))$ is a random variable and its CDF $F_y$ can be alternatively calculated by using:

- The original space : $F_y(b) = P(x^{-1}(g^{-1}(y \le b)))$
- The dependent space $(\mathbb{R}, \mathcal{B}, P_x)$: $F_y(b) = P_x(g^{-1}(y \le b))$.

Additionally, if $x$ is a continuous random variable:

$$F_y(b) = \int_{g^{-1}(y \le b)} f_x(a)\, da \tag{2.17}$$

The expectation of $y$ can for example be evaluated by the Stieltjes integral:

$$E[y = g(x)] = \int_{-\infty}^{+\infty} g(a)\, dF_x(a) \tag{2.18}$$

When the probability density of $x$ exists, the expectation can be more directly given, according to the *Fundamental Theorem of the expectation*, by the absolutely convergent Rieman Integral:

$$E[y = g(x)] = \int_{-\infty}^{+\infty} g(a) f_x(a) da \tag{2.19}$$

Assuming $g = x^k$, Eq. (2.19) gives the *Moments of order k* of the random variable $x$. The Moment of order 1 equals the expectation: it measures a weighted average or the *mean value* $\mu_x$ of $x$; the Moments of higher order describe the dispersion of $x$ around the mean value: therefore *central Moments* of order $k$>1 are better defined relative to the mean value. Particularly important is the *Variance* of $x$, $\sigma^2_x$:

$$\begin{aligned} \sigma^2(x) &= E[y = (x - \mu_x)^2] = \int_{-\infty}^{+\infty} (a - \mu_x)^2 f_x(a) da = \\ &= E[x^2] - \mu_x^2 \end{aligned} \tag{2.20}$$

The square root of the variance is the *standard deviation* $\sigma_x$.

## 2.3 Joint Probability Spaces

Given two probability spaces, $(S_i, \mathcal{S}_i, P_i)$, $i = 1, 2$, the *product* (*or joint*) *probability space*, $(S, \mathcal{S}, P)$, is such that:

$$(i)\ S := \{S_1 \times S_2\}; \tag{2.21}$$

$$(ii)\ \mathcal{S} \text{ is the } \sigma\text{-algebra generated by } \mathcal{C} := \{A_1 \times A_2 : A_i \in \mathcal{S}_i\}; \tag{2.22}$$

$$(iii)\ P(A_1 \times S_2) = P_1(A_1)\ ;\ P(S_1 \times A_2) = P_2(A_2) \tag{2.23}$$

Condition (2.23) is called *marginal* (or *addition*) *rule*, and does not uniquely determine $P$. Spaces $(S_i, \mathcal{S}_i, P_i)$ are called *marginal probability* spaces.

Let $P_i$ be a probability of elementary events on $S_i = \{ s_i^j : j = 1,\ldots, n_i\}$, and let $\mathbf{p}_i$ be a $n_i$–column vector whose $j$-th entry is $P_i(s_i^j)$. Let $P$ be a known probability of joint elementary events on $S_1 \times S_2 = S$, and let $\mathbf{P}$ be a $n_1 \times n_2$ matrix with $(j, k)$-th entry $P\left(s_1^j, s_2^k\right)$. Eq. (2.23) entails (marginal rule)

$$P_1\left(s_1^j\right) = \sum_{s_2^k \in S_2} P\left(s_1^j, s_2^k\right)\ \ ;\ P_2\left(s_2^j\right) = \sum_{s_1^k \in S_1} P\left(s_1^k, s_2^j\right) \tag{2.24}$$

Thus, $P_1\left(s_1^j\right)$ is given by the sum of $j$-th row of $\mathbf{P}$ ($\mathbf{p}_1 = \mathbf{P} \cdot \mathbf{1}_{(n_2)}$), and $P_2\left(s_2^j\right)$ is given by the sum of $j$-th column of $\mathbf{P}$ ($\mathbf{p}_2 = \mathbf{P}^{\mathrm{T}} \cdot \mathbf{1}_{(n_1)}$), where a superscript "T" denotes transposition, and $\mathbf{1}$ is a vector of unit components of proper length.

Provided $P_l\left(s_l^k\right) > 0$, the probability of $s_i^j$ in $S_i$ conditioned on elements $s_l^k$ in $S_l$ can be easily calculated using Eq. (2.4):

$$P_{1|2}\left(s_1^j \mid s_2^k\right) = P\left(s_1^j, s_2^k\right) / P_2\left(s_2^k\right) \tag{2.25}$$

For a given element $s_2^k$, $P_{1|2}$ is thus obtained by dividing the $k$-th column of $\mathbf{P}$ by $P_2\left(s_2^k\right)$. Likewise, for a given element $s_1^k$, $P_{2|1}$ is obtained by dividing the $k$-row of $\mathbf{P}$ by $P_1\left(s_1^k\right)$. Eq. (2.24) yields

$$\sum_{j=1}^{n_1} P_{1|2}\left(s_1^j \mid s_2^k\right) = \frac{1}{P_2\left(s_2^k\right)} \sum_{j=1}^{n_1} P\left(s_1^j, s_2^k\right) = 1 \tag{2.26}$$

and thus $P_{1|2}$ is a probability distribution of elementary events on $S_1$. Likewise for $P_{2|1}$. Let $\mathbf{P}_{1|2}$ be the $n_1 \times n_2$ matrix with $(j, k)$-th entry $P_{1|2}\left(s_1^j \mid s_2^k\right)$, and let $\mathbf{P}_{2|1}$ be the $n_1 \times n_2$ matrix with $(j, k)$-th entry $P_{2|1}\left(s_2^k \mid s_1^j\right)$.

Given the joint probability distribution $\mathbf{P}$, one can calculate: two marginal probabilities, $\mathbf{p}_j$ , by using Eq. (2.24); and then two conditional probabilities, $\mathbf{P}_{1|2}$ and $\mathbf{P}_{2|1}$, by using Eq. (2.25).

On the other hand, given one marginal probability, say $\mathbf{p}_2$, and the conditional probabilities, $\mathbf{P}_{1|2}$, then one can determine $\mathbf{P}$ by using the definition of conditional probability (2.5):

$$\mathbf{P} = \mathbf{P}_{1|2}\, Diag(\mathbf{p}_2) \tag{2.27}$$

where $Diag(.)$ is a diagonal matrix whose $i$-th diagonal element is the $i$-th element of the argument vector. The marginal probabilities $\mathbf{p}_1$ can be either calculated using Eq. (2.24) or directly using the theorem of Total Probability (2.7):

$$\mathbf{p}_1 = \mathbf{P}_{1|2}\, \mathbf{p}_2 \tag{2.28}$$

Likewise, given $\mathbf{p}_1$, and the conditional probabilities, $\mathbf{P}_{2|1}$:

$$\mathbf{P} = Diag(\mathbf{p}_1)\, \mathbf{P}_{2|1} \tag{2.29}$$

and:

$$\mathbf{p}_2^{\mathrm{T}} = \mathbf{p}_1\, \mathbf{P}_{2|1} \tag{2.30}$$

The marginal probability spaces $(S_i, \mathcal{S}_i, P_i)$ are called independent if the joint $P$ is the product measure of $P_1$ and $P_2$, i.e. it satisfies $P = P_1 \otimes P_2 : \mathcal{C} = \{U_1 \times U_2 : U_i \in \mathcal{S}_i\} \to [0,1]$ with:

$$P_1 \otimes P_2 \left(U_1 \times U_2\right) := P_1\left(U_1\right) \cdot P_2\left(U_2\right) \tag{2.31}$$

and Carathéodory Extension Theorem then allows one to extend $P$ to any subset in the σ-algebra $\mathcal{S}$ generated by $\mathcal{C}$.

This definition is coherent with Eq. (2.6) because, if we let $T_1 = U_1 \times S_2$ and $T_2 = S_1 \times U_2$, then $T_1 \cap T_2 = U_1 \times U_2$ and:

$$P(T_1) = P_1 \otimes P_2(U_1 \times S_2) = P_1(U_1) \cdot P_2(S_2) = P_1(U_1) \cdot 1 = P_1(U_1) \tag{2.32}$$

$$P(T_2) = P_1 \otimes P_2(S_1 \times U_2) = P_1(S_1) \cdot P_2(U_2) = 1 \cdot P_2(U_2) = P_2(U_2) \tag{2.33}$$

$$P(T_1 \cap T_2) = P_1 \otimes P_2(U_1 \times U_2) = P_1(U_1) \cdot P_2(U_2) \tag{2.34}$$

Eq. (2.6) follows by putting Eqs. (2.32) and (2.33) into (2.34).

For the probability distribution of the joint elementary events the hypothesis of independence gives:

$$P(s_1^i, s_2^k) = P_1(s_1^i) \cdot P_2(s_2^k)\, ; \mathbf{P} = \mathbf{p}_1 \mathbf{p}_2^{\mathrm{T}} \tag{2.35}$$

## 2.4 Random Vectors

In the two-dimensional space, the set of pairs **a** of real numbers, $\mathbb{R}^2$, the σ-algebra generated by $\mathcal{C}$={(- ∞, **a**]: **a** ∈ $\mathbb{R}^2$ } is again called the *Borel σ-algebra*, $\mathcal{B}_2$ and contains all two-dimensional intervals of $\mathbb{R}^2$.

Let ($S$, $\mathcal{S}$, $P$) be a probability space and **x** a *Borel-measurable* real function **x** : $S \to \mathbb{R}^2$, defined on $S$: **x**($s$) is a *random vector* on ($S$, $\mathcal{S}$, $P$). It means that, for every **a**∈ $\mathbb{R}^2$, {$s$: **x**($s$) ≤ **a**}∈ $\mathcal{S}$, where inequalities are meant to hold component-wise.

In the dependent two-dimensional probability space ($\mathbb{R}^2$, $\mathcal{B}_2$, $P_{\mathbf{x}}$) again $P_{\mathbf{x}}(T)= P(\mathbf{x}^{-1}(T))$ is given by the CDF of the random vector **x:** $F_{\mathbf{x}}(\mathbf{a}) = P_{\mathbf{x}}(\{\mathbf{x}: \mathbf{x} \le \mathbf{a}\}) = P(\{s: \mathbf{x}(s) \le \mathbf{a}\})$.

When $F_{\mathbf{x}}(\mathbf{a})$ is absolutely continuous, it can be expressed as integral of the *joint probability density* $f_{\mathbf{x}}(\mathbf{a})$ of the random vector **x**:

$$f_{\mathbf{x}}(\mathbf{a}=(a_1,a_2)) = \frac{\partial^2 F_x(\mathbf{a})}{\partial a_1 \partial a_2} \tag{2.36}$$

Otherwise, let $B \subset \mathbb{R}^2$ the subset (finite or infinitely numerable) of discontinuity points of **x** and let $p_{\mathbf{x}}$ the discontinuity jump at **a**∈$B$. If:

$$\sum_{\mathbf{a}\in B} p_{\mathbf{x}}(\mathbf{a}) = 1 \tag{2.37}$$

then **x** is a *discrete random vector* and $p_{\mathbf{x}}$ is the *joint mass distribution*.

If $B$ is finite, $p_{\mathbf{x}}$ allows one to calculate the probability of any subset of $B$ in a way similar to the probability of elementary events.

The notions of marginal and conditional mass distributions are related to the joint mass distributions by means of matrix operations equivalent to the operations defined for the joint elementary events of product spaces in Section 2.3.

For an absolutely continuous random vector **x** = ($x_1$, $x_2$) analogous definitions and relations could be given in terms of probability density (pdfs). For example the *conditional pdf* of $x_1$ given $x_2$ is:

$$f_{x_1|x_2}(x_1,x_2) = \frac{f_{\mathbf{x}}(x_1,x_2)}{f_{x_2}(x_2)} \tag{2.38}$$

from which we also have:

$$f_{\mathbf{x}}(x_1,x_2) = f_{x_1|x_2}(x_1,x_2)\, f_{x_2}(x_2) \tag{2.39}$$

Moreover the marginal pdfs can be derived by an integral extension of the Theorem of total probability:

$$f_{x_1}(x_1) = \int_{-\infty}^{+\infty} f_{\mathbf{x}}(x_1, x_2)\, dx_2 = \int_{-\infty}^{+\infty} f_{x_1|x_2}(x_1, x_2)\, f_{x_2}(x_2)\, dx_2 \tag{2.40}$$

When the probability density of **x** exists, the expectation can be given by the extension of the *Fundamental Theorem of the expectation*, through the absolutely convergent Riemann Integral:

$$E[y = g(\mathbf{x})] = \int_{-\infty}^{+\infty} g(\mathbf{a})\, f_x(\mathbf{a})\, d\mathbf{a} \tag{2.41}$$

while for a discrete random vector:

$$E[y = g(\mathbf{x})] = \sum_{\mathbf{a} \in B} g(\mathbf{a})\, p_x(\mathbf{a}) \tag{2.42}$$

Assuming g = $x_1^{k}\, x_2^{j}$, eq. (2.41) or (2.42) give the *Moments of type (k, j) and order k+j* of the random vector *x*. The Moments of order 1 (type (1,0) and (0,1)) equal the *mean values* ($\mu_{x1}$ , $\mu_{x2}$) of the single random variables in the vector*;* the central Moments of order 2 define the *matrix of Covariance* of **x** $\sigma_{\mathbf{x}}$: the diagonal of the matrix ( types (2,0): $\sigma^2_{x1}$ and (0,2): $\sigma^2_{x2}$) contain the Variance of the single random variables, while the other coefficients of the symmetrical squared matrix gives the *Covariance of the couple of random variables* (Type (1,1): $\sigma_{x1,x2}$):

$$\sigma_{\mathbf{x}}(x_1, x_1) = \begin{Vmatrix} E[(x_1 - \mu_{x_1})^2] & E[(x_1 - \mu_{x_1})(x_2 - \mu_{x_2})] \\ Sym & E[(x_2 - \mu_{x_2})^2] \end{Vmatrix} \tag{2.43}$$

Since the determinant of the matrix cannot be negative, the *coefficient of correlation* $\rho_{x1,x2} = \sigma_{x1,x2}/(\sigma_{x1}\ \sigma_{x2})$ must be in the interval [-1, 1]. This coefficient synthetically measures the sign and the weight of a linear correlation between the two random variables. When $|\rho_{x1,x2}| = 1$ the variables are totally (positively or negatively) correlated; when $\rho_{x1,x2} = 0$ the variables are uncorrelated.

Uncorrelation does not mean statistical independence of the single random variables of a random vector. The latter refers to the relations between joint, conditional and marginal pdfs or mass distributions, as specified in § 2.3. Considering for example absolutely continuous random vectors the joint pdf in (2.39) is directly determined by the product of the marginals.

Statistical independence implies uncorrelation, but uncorrelation does not imply independence, because a non linear statistical (or also deterministic) dependence between the two random variables could be present.

# Chapter 3
# Random Sets and Imprecise Probabilities

The idea of random sets is introduced by showing that three different extensions to the classical probabilistic information lead to an equivalent mathematical structure. A formal definition is then given, followed by different ways to describe the same information.

A random set gives upper and lower bounds on the probability of subsets in a space of events. These non-additive and monotone (with respect to inclusion) set functions can be described within a more general framework by resorting to the theory of imprecise probabilities, Choquet capacities, and convex sets of probability distributions. The chapter highlights specific properties, advantages and limitations of random sets with special emphasis on evaluating function expectation bounds and on updating the available information when new information is acquired. To avoid mathematical complications, sets and spaces of finite cardinality are generally considered.

## 3.1 Extension of Probabilistic Information

### *3.1.1 Multi-valued Mapping from a Probability Space*

This extension was proposed in (Dempster 1967) and is summarized in Figure 3.1. Let $(X, \mathcal{X}, P_x)$ be a probability space (for example a random variable with cumulative distribution function $F_x(x)$) and let $G$: $X \rightarrow \mathcal{S}$ be a multi-valued mapping to a measurable space $(S, \mathcal{S})$ (for example $G(x)$ is the interval in the grey area in Figure 3.1). For a set $T \in \mathcal{S}$, let

$$T^* = \{x \in X \mid G(x) \cap T \neq \varnothing\}; \;\; T_* = \{x \in X \mid G(x) \subseteq T\} \tag{3.1}$$

$S^* = S_*$ is the domain of $G$, here assumed to be equal to $X$; hence $P_x(S^*) = P_x(S_*) = 1$.

$G$ is a *strongly measurable* function if for any set $T \in \mathcal{S}$, $T^* \in \mathcal{X}$ (and consequently $T_* \in \mathcal{X}$ (Miranda 2003)). The exact value of the probability of $T$ (in the probability space $(S, \mathcal{S}, P)$) cannot be computed, but it can be bounded by the probabilities of $T_*$ and of the inclusive set $T^*$:

$$\frac{P_x(T_*)}{P_x(S_*)} = P_x(T_*) \;\leq\; P(T) \;\leq\; P_x(T^*) = \frac{P_x(T^*)}{P_x(S^*)} \tag{3.2}$$

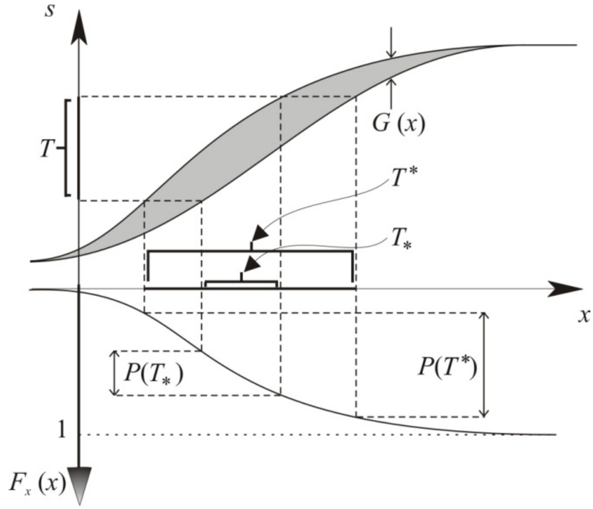



**Fig. 3.1** Probability bounds from a multi-valued mapping

**Example 3.1.** The characteristic compressive strength of a masonry wall ($f_k$) can be derived through a function of the unit ($f_b$) and mortar ($f_m$) strengths. According to (CEN 2005) for plain solid (one head) masonry made with clay, group 1 units and general purpose mortar $f_k = 0.55\, f_b^{0.7} f_m^{0.3} = g\,(f_b\,, f_m)$. Assume that only an interval of possible values is known for the mortar strength, while a precise probability distribution has been derived for the units by testing. The upper and lower bounds of probability for each interval $T$ of values of masonry strength can then be computed as follows.

Assuming: $f_m$= [20, 30] MPa and the Normal cumulative distribution function N(40 MPa, 8 MPa) for the random variable $x = f_b$, $G(x) = [g\,(x, 20), g\,(x, 30)]$, the probability of $T = [25, \infty)$ (i.e. the probability that the masonry strength could be above 25 MPa) is bounded by:

$$P_x(T_* = \{x > g^{-1}(25, 20)\}) = 1 - F_x\,(g^{-1}(25, 20)) = 0.0010$$
$$P_x(T^* = \{x > g^{-1}(25, 30)\}) = 1 - F_x\,(g^{-1}(25, 30)) = 0.0368.$$

## *3.1.2 Theory of Evidence*

In the finite space $S$ (a "*body of evidence*" (Shafer 1976)), a "*probabilistic assignment*" $m$ is given on the power set of $S$ ($\mathcal{P}(S)$: the set of all subsets of $S$; if $|S|$ is the cardinality of $S$, then $|\mathcal{P}(S)| = 2^{|S|}$, including $\varnothing$ and $S$). The probabilistic assignment is given according to the axioms of probability theory, and therefore $m(\varnothing) = 0$, $\Sigma\, m = 1$.

**Example 3.2.** An expert is asked to define the cause of a structural deficiency in a building by choosing among a given list of options $c$ listed in Table 3.1 ($S = \{c^1, c^2, c^3, c^4\}$). Based on his past experience and current observations, the expert could measure the different causes $c$ (first column), and attach subjective probabilities $m$

(second column) not only to single causes, but to sets of causes. In his opinion, some observed symptoms point to single causes, but other symptoms are compatible with more causes, or with all listed causes.

The probability of the single causes or of a set of causes can easily be calculated: for example, the probability of $c^2$ is at least $m^2$ (10%), but could increase to $m^2+m^5+m^6$ (30%); the probability of ($c^1$ or $c^2$: $c^1 \cup c^2$) is at least $m^1+m^2$ (70%) but could be higher, and up to $m^1+m^2+ m^5+m^6$ (90%).

**Table 3.1** Expert's subjective probabilities in a structural diagnosis

| | |
|---|---|
| $c^1$ - Failure of soil | $m^1 = m(c^1) = 60\%$ |
| $c^2$ - Overloading | $m^2 = m(c^2) = 10\%$ |
| $c^3$ - Temperature variations | $m^3 = m(c^3) = 5\%$ |
| $c^4$- Other | $m^4 = m(c^4) = 5\%$ |
| | $m^5 = m(c^2, c^3, c^4) = 10\%$ |
| | $m^6 = m(c^1, c^2, c^3, c^4) = 10\%$ |
| | Total : = 100% |

The original information is described by a family $\mathcal{F}$ of pairs of $n$ nonempty subsets $A^i$ ("*focal elements*") and attached $m^i = m(A^i) > 0$, $i \in I = \{1, 2, \ldots n\}$, with the condition that the sum of $m^i$ is equal to 1. The (total) probability of any subset $T$ of $S$ can therefore be bounded by means of the additivity rule. Shafer suggested the words *Belief* (*Bel*) and *Plausibility* (*Pla*) for the lower and upper bound, respectively. Formally:

$$\begin{aligned}
&\mathcal{F} = \mathcal{F}\left\{ (A^i, m^i),\ i \in I \mid \sum_{i \in I} m^i = 1 \right\} \\
&Pla(T) = \sum_i m^i \mid A^i \cap T \neq \varnothing, \quad \forall T \subset S; \\
&Bel(T) = \sum_i m^i \mid A^i \subseteq T, \qquad \forall T \subset S; \\
&Bel(T) \leq P(T) \leq Pla(T)
\end{aligned} \tag{3.3}$$

### *3.1.3 Inner/Outer Extension of a Probability Space*

It is well known that a probability measure can be given only for a measurable space: i.e. the probability can be attached only to particular families of subsets on a space $S$ (an algebra on finite spaces; a σ-algebra on infinite spaces). The key property is the closure of the family with respect to complementation and (numerable) union (and therefore the (numerable) intersection). For example, the σ-algebra could be generated by a finite partition of $S$. But given a precise probability measure, it is legitimate to ask about bounds of the probability of any other subset $T$ of $S$ (Halpern and

Fagin 1992). The reply can be obtained searching for the best members of the σ-algebra that give an inner approximation ($T_{in} \subseteq T$), and an outer approximation ($T \subseteq T_{out}$) to $T$.

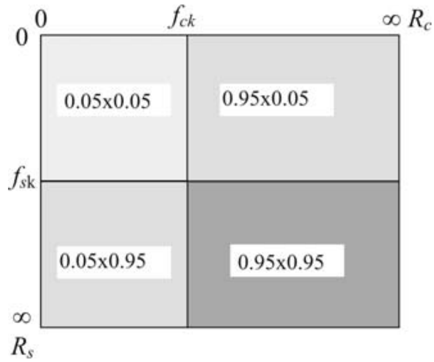



**Fig. 3.2** Partition of the Cartesian product space $S = R_c \times R_s$

**Example 3.3.** Let us suppose the characteristic (reliable at 95%) value of the strength of concrete ($R_c = f_{ck}$ = 30 MPa) and steel ($R_s = f_{sk}$ = 400 MPa) is known in a reinforced concrete (r. c.) frame structure. A partition of 4 elementary events is therefore defined on the Cartesian product space $S = R_c \times R_s = \mathbb{R} \times \mathbb{R}$ ; moreover, supposing stochastic independence between $R_c$ and $R_s$, the probability of the elementary events and 16 members of the algebra generated by the partition (the union of any subsets of elementary events plus the empty set) can easily be derived (Figure 3.2). We now wish to bound the probability of the event $T = \{(R_c, R_s) \mid R_c \leq 40$ MPa; $R_s \leq f_{sk}\}$, clearly not included in the algebra.

The inner approximation is $T_{in} = \{(R_c, R_s) \mid R_c \leq f_{ck} = 30$ MPa; $R_s \leq f_{sk}\}$, with $P(T_{in})$= 0.05x0.05 = 0.0025, while the outer approximation also includes the elementary event $\{(R_c, R_s) \mid R_c > f_{ck} = 30$ MPa; $R_s \leq f_{sk}\}$. Therefore $P(T_{out})$ = 0.0025 + 0.95x0.05 = 0.0500. If additional information is received that $R_c$ is a Gaussian random variable with mean equal to 45 MPa, the exact value of $P(T)$ can be calculated because the standard deviation of $R_c$ is equal to:

$$(45 - 30)/\, N^{1}(\,0,\,1,\,0.95) = 15/1.644 = 9.12 \text{ MPa},$$

and hence $P(T) = 0.05 \text{ x } N(45, 9.12, 40) = 0.05 \text{ x } 0.2917 = 0.01459$.

## 3.2 Random Sets

### *3.2.1 Formal Definition of Random Sets*

The strong formal and substantial analogy between the three formulations given above is self-evident.

In this book priority is given to a direct reference to the second formulation, originally proposed by Shafer within the so-called Evidence Theory, and therefore particularly connected to a subjective view of the probability concept. However, we prefer the term "*Random Sets*", following an idea originally developed within stochastic geometry (Robbins 1944; Robbins 1945; Matheron 1975), to underline that the formulation is compatible with both objective and subjective uncertainty.

Formally, a random set on the space $S$ is a family $\mathcal{F}$ of $n$ focal elements $A^i \subseteq S$ and attached weights of the basic probabilistic assignment $m(A^i)$ that satisfies the conditions: $m(\varnothing) = 0$; $\Sigma_i\, m(A^i) = 1$. See Eq. (3.3).

The weight $m(A^i)$ expresses the extent to which all available and relevant evidence supports the claim that a particular element of $S$ belongs to the set $A^i$ alone (i.e. exactly to set $A^i$) and does not imply any additional claims regarding subsets of $A^i$; if there is any additional evidence supporting the claim that the element belongs to a subset $B$ of $A^i$, it must be explicitly expressed by another value $m(B)$. The main difference between a probability distribution function and a basic assignment is that the former is defined on $S$, whereas the latter is defined on the power set of $S$, $\mathcal{P}(S)$.

As a consequence, the following properties hold:

1) it is not required that $m(S) = 1$;
2) it is not required that $m(A) \leq m(B)$ when $A \subset B$;
3) no relationship between $m(A)$ and $m(A^C)$ is required ($A^C$ is the complementary set of $A$).

Each focal element $A$ must be treated as an object "*per se*"; $m(A) \leq m(B)$ means that object $A$ is less probable than object $B$. It should be noted that:

a) If $m(S) = 1$, there is a unique focal element and this is $S$ itself (maximum ignorance).
b) Conversely, if there is a unique focal element $A \subset S$, then $m(A) = 1$ and $m(S) = 0$. If moreover $|A| = 1$ all uncertainty disappears.
c) If there are two or more focal elements, then $m(S) < 1$.

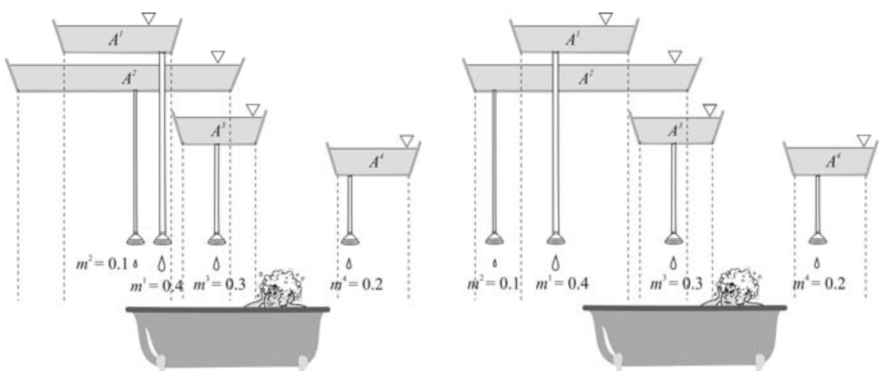
It should be stressed that the definition of random set refers to distinct non-empty subsets of $S$. If these distinct non-empty subsets are singletons (single elements, thus non-overlapping, of $S$) and each one has a probability assignment, then we have a probability distribution on $S$. Note that when processing real world information, the non-empty subsets may be overlapping (see Chapter 1).

**Example 3.4** (Reservoirs-bathtub analogy)**.** As depicted in Figure 3.3, consider a set of reservoirs (focal sets) $A^i$, whose outward flow rate (basic probability assignment) is $m(A^i)$. This outward flow can only be vertical down (positive); as for the $i$-th reservoir, any number of vertical pipes can be located anywhere and arranged in any fashion on the footprint of the reservoir, but their total flow rate is always equal to $m(A^i)$. Pipes are not allowed to discharge into other reservoirs, and the total flow rate from all reservoirs is normalized to 1. No water may come from a source different than a reservoir ($m(\emptyset)=0$).

One can calculate the maximum possible flow rate enjoyed by a bather in a bathtub $T$ (call it $Pla(T)$) by arranging single pipes so that all reservoirs whose vertical projection hits the bathtub actually discharge into it. In Figure 3.3a, the maximum flow rate is 0.8. The minimum flow rate (call it $Bel(T)$) is obtained by placing single pipes outside of the bathtub projection unless a reservoir projects completely into the bathtub, in which case there is no choice but to discharge into $T$. In Figure 3.3b, the minimum flow rate is 0.3. Notice that there may be more than one pipe arrangement that yields the maximum or minimum flow rate into the bathtub. Any other arrangement of the pipes will yield a flow rate into $T$ (Probability of $T$) that will be comprised between these $Bel(T)$ and $Pla(T)$ (Eq. (3.3d)). In precise probability theory, reservoirs are restricted to a single point in space, and thus only one pipe carrying the entire flow $m(A^i)$ can be attached to the $i$-th reservoir, and only one pipe arrangement is possible.

As a consequence, each possible single pipe arrangement that fits in the reservoirs of a random set corresponds to a probability distribution (called Selector, see Section 3.2.3.2 on page 35).

On the other hand, several pipes may be attached to the $i$-th reservoir. Without loss of generality, the pipes attached to the $i$-th reservoir may have a unit total flow rate and may be fitted with flow rate reducers; a flow rate reducer will reduce the flow rate in each single pipe by a factor equal to $m(A^i)$. Each set of pipes of unit flow rate attached to the $i$-th reservoir may be interpreted as a probability distribution on $A^i$. These pipe arrangements over the entire set of reservoirs make up the probability distributions compatible with the random set (Section 3.2.3 on page 34).




(a) (b)

**Fig. 3.3** Reservoir-bathtub analogy: (a) *plausible* flow rate gives an optimistic outlook; (b) *believed* flow rate gives a pessimistic outlook

### *3.2.2 Equivalent Representations of Random Sets*

When a random set is given on the space $S$, Eq. (3.3) shows how upper/lower probability bounds can be attached to each subset $T \subset S$. In this book, the words *Belief* and *Plausibility* suggested by Shafer will be used for upper/lower probability bounds, but without any particular reference to the meaning of these words in normal language.

For the entire space $S$, $Bel(S) = P(S) = Pla(S) = 1$. However, when two complementary sets $T$ and $T^{\mathrm{C}}$ are considered, the sums of their Beliefs or Plausibilities are not required to be equal to 1, but they are related by the following weaker conditions:

$$\begin{aligned}
&Bel(T) + Bel(T^{\mathrm{C}}) = \sum_i m^i \mid A^i \subseteq T + \sum_i m^i \mid A^i \subseteq T^{\mathrm{C}} \\
&\le \sum_i m^i \mid A^i \subseteq (T \cup T^{\mathrm{C}}) \;=\; Bel(T \cup T^{\mathrm{C}} = S) = 1 \\
&Pla(T) + Pla(T^{\mathrm{C}}) = \sum_i m^i \mid A^i \cap T \ne \varnothing + \sum_i m^i \mid A^i \cap T^{\mathrm{C}} \ne \varnothing \\
&\ge \sum_i m^i \mid A^i \cap (T \cup T^{\mathrm{C}}) \ne \varnothing \;=\; Pla(T \cup T^{\mathrm{C}} = S) = 1
\end{aligned} \tag{3.4}$$

It is easy to check the first formula: consider that some focal elements may not be included in either $T$ or $T^{\mathrm{C}}$. In the second formula, some focal elements may intersect both $T$ and $T^{\mathrm{C}}$, and are therefore counted twice: once in $Pla(T)$, and once in $Pla(T^{\mathrm{C}})$.

More generally, given two sets $T^1$ and $T^2$ and considering that:

- Some focal elements included in $T^1 \cup T^2$ may not be included in $T^1$, $T^2$ and $T^1 \cap T^2$;
- Some focal elements may not intersect $T^1 \cap T^2$ but may intersect both $T^1$, $T^2$ (and therefore $T^1 \cup T^2$),

one obtains:

$$\begin{aligned}
&Bel(T^1 \cup T^2) \ge Bel(T^1) + Bel(T^2) - Bel(T^1 \cap T^2) \\
&Pla(T^1 \cap T^2) \le Pla(T^1) + Pla(T^2) - Pla(T^1 \cup T^2)
\end{aligned} \tag{3.5}$$

Even more generally, given $k$ sets $T^1$, $T^2$… $T^k$:

$$
\begin{aligned}
&Bel(T^1 \cup T^2 ... \cup T^k) \geq \sum_{\varnothing \subset K \subseteq \{1,2..k\}} (-1)^{|K|+1} Bel\left(\bigcap_{i\in K} T^i\right) = \\
&\sum_i Bel(T^i) - \sum_{i\neq j} Bel(T^i \cap T^j) + ... + (-1)^{k+1} Bel(T^1 \cap T^2 ... \cap T^k) \\
&Pla(T^1 \cap T^2 ... \cap T^k) \leq \sum_{\varnothing \subset K \subseteq \{1,2..k\}} (-1)^{|K|+1} Pla\left(\bigcup_{i\in K} T^i\right) = \\
&\sum_i Pla(T^i) - \sum_{i\neq j} Pla(T^i \cup T^j) + ... + (-1)^{k+1} Pla(T^1 \cup T^2 ... \cup T^k)
\end{aligned}
\tag{3.6}
$$

*Bel* and *Pla* satisfy Eq. (3.6) for any $k > 2$. Formulas (3.4), (3.5) and (3.6) generalize stronger relations of equality that must be satisfied by a probability set function $P$, as a consequence of the additivity axiom for the probability of the union of disjoint sets (Eq. (2.1)).

It is easy to demonstrate that a duality relation intimately connects the set functions $Bel(T)$ and $Pla(T)$: indeed, the condition ($A^i \subseteq T$) implies the negation of condition ($A^i \cap T^{\mathrm{C}} \neq \varnothing$), and therefore:

$$
\begin{aligned}
&Bel(T) + Pla(T^{\mathrm{C}}) = \\
&\sum_i \left(m^i \mid A^i \subseteq T\right) + \sum_i \left(m^i \mid A^i \not\subset T\right) = \sum_i m^i = 1
\end{aligned}
\tag{3.7}
$$

When the set function *Bel* (or *Pla*): $\mathcal{P}(S) \to [0, 1]$ has been evaluated for every $T \subset S$, Eq. (3.7) gives *Pla* (or *Bel*, respectively). Additionally, the original set function $m(A)$ can be reconstructed through the *Möbius transform* of the set function $Bel(T)$:

$$
{}^{Bel}m(A) = \sum (-1)^{|A-T|} Bel(T) \mid T \subseteq A \tag{3.8}
$$

The Möbius transform of a set function $\mu$ is a one-to-one invertible set function ${}^{\mu}m$: $\mathcal{P}(S) \to \mathbb{R}$, which is defined by replacing $\mu$ for *Bel* in Eq. (3.8). Its inverse is:

$$
{}^{m}\mu(\mathrm{T}) = \sum m(A) \mid A \subseteq T, \qquad \forall T \subset S \tag{3.9}
$$

When Eq. (3.8) is applied to a *Belief* function, $m(A) \in [0,1]$ and $\Sigma\, m(A) = 1$, i.e. $m(A)$ is a probabilistic assignment. Conversely, when $m(A)$ is a probabilistic assignment, Eq. (3.9) gives the corresponding *Belief* function (in fact ${}^{m}\mu$ in (3.9) is equal to *Bel* in (3.3)).

Therefore, the information given by a random set on the space $S$ is completely described by a *Belief* set function; but not every set function ${}^{m}\mu$ is equivalent to a random set. The following Section 3.3.3 on page 69 gives the conditions for the set function ${}^{m}\mu$ to be a *Belief* function, and therefore to be equivalent to a random set.

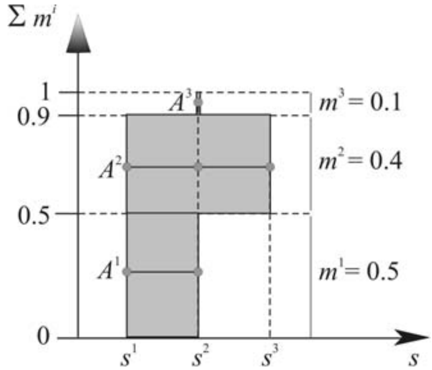



**Fig. 3.4** Graphical representation of the random set in Example 3.5

**Example 3.5.** Let $S=\{s^1, s^2, s^3\}$, and let $\mathcal{F}=\{(\{s^1, s^2\}, 0.5), (\{s^1, s^2, s^3\}, 0.4), (\{s^2\}, 0.1)\}$ be a random set on $S$. A graphical representation of the random set is displayed in Figure 3.4 as a pile of boxes. The width of each box covers a focal element, and the box height is equal to the relevant probabilistic assignment. The power set $\mathcal{P}(S)$ contains the $2^3 = 8$ subsets $A^i$ listed in Table 3.2 and identified by the indicator function $I_{A^i} = \{\chi(s^1), \chi(s^2), \chi(s^3)\}$. The set functions $m(A^i)$, $Bel(A^i)$, $Pla(A^i)$, $Pla(A^{i,C})$ are displayed in the same table; it is therefore possible to check that formulas (3.5) are always satisfied. The reader can also check that formulas (3.4) are satisfied as well. Finally, Eq. (3.8) was used to calculate ${}^{Bel}m$ through *Bel*, and ${}^{Bel}m$ is found to be identical to $m$. For example:

$$
\begin{aligned}
{}^{Bel}m(A^3) &= (-1)^{|A^3-A^3|}Bel(A^3) = (-1)^0\, 0.1 = 0.1\\
{}^{Bel}m(A^5) &= (-1)^{|A^5-A^3|}Bel(A^3) + (-1)^{|A^5-A^5|}Bel(A^5) = (-1)^1\, 0.1 + (-1)^0\, 0.6 = 0.5\\
{}^{Bel}m(A^7) &= (-1)^{|A^7-A^3|}Bel(A^3) + (-1)^{|A^7-A^7|}Bel(A^7) = (-1)^1\, 0.1 + (-1)^0\, 0.1 = 0\\
{}^{Bel}m(A^8) &= (-1)^{|A^8-A^3|}Bel(A^3) + (-1)^{|A^8-A^7|}Bel(A^7) + (-1)^{|A^8-A^8|}Bel(A^8) =\\
&= (-1)^2\, 0.1 + (-1)^1\, 0.6 + (-1)^1\, 0.1 + (-1)^0\, 1.0 = 0.1 - 0.6 - 0.1 + 1.0 = 0.4
\end{aligned}
$$

**Table 3.2** Set functions in Example 3.5

| $i$ | $\chi(s^1)$ | $\chi(s^2)$ | $\chi(s^3)$ | $m(A^i)$ | $Bel(A^i)$ | $Pla(A^i)$ | $Pla(A^{i,C})$ | $Bel(A^i)+Pla(A^{i,C})$ | ${}^{Bel}m(A^i)$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 | 1 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0.9 | 1.0 | 1 | 0 |
| 3 | 0 | 1 | 0 | 0.1 | 0.1 | 1.0 | 0.9 | 1 | 0.1 |
| 4 | 0 | 0 | 1 | 0 | 0 | 0.4 | 1.0 | 1 | 0 |
| 5 | 1 | 1 | 0 | 0.5 | 0.6 | 1.0 | 0.4 | 1 | 0.5 |
| 6 | 1 | 0 | 1 | 0 | 0 | 0.9 | 1.0 | 1 | 0 |
| 7 | 0 | 1 | 1 | 0 | 0.1 | 1.0 | 0.9 | 1 | 0 |
| 8 | 1 | 1 | 1 | 0.4 | 1.0 | 1.0 | 0 | 1 | 0.4 |

### *3.2.3 Probability Distributions Compatible with a Random Set*

Consider the measurable space $(S, \mathcal{S})$. It is possible to consider a random set as a class $\Psi$ of probability measures $P(T)$ for the sets $T \in \mathcal{S}$. For $S$ finite, with cardinality $|S|$, $\mathcal{S}$ could be the maximum algebra on $S$, i.e. the power set of $S$: the class of the $2^{|S|}$ subsets of $S$ (containing $S$ and the empty set $\varnothing$). In a more general way, let $\mathcal{S}$ be the algebra generated by a finite partition of $S$ that contains only elementary events, in the following named "singletons", $s^j$. For all $T \in \mathcal{S}$, $P(T) = \Sigma\,(P(s^j)= P(\{s^j\})|\, s^j \in T)$, and $P(s^j)$ is the probability distribution attached to the singletons $s^j$.

Class $\Psi$ can be defined in two different ways:

1. Class $\Psi^E$ of probability distributions $P$ such that for all $T \in \mathcal{S}$ $P(T)$ is bounded by $Bel(T)$ and $Pla(T)$ (see for example (Dempster 1967)):

$$\Psi^E = \left\{P \;\; | \forall T \in \mathcal{S} : Bel(T) \le P\left(T\right) \le Pla(T)\right\} \tag{3.10}$$

2. Recall that the probability assignment $m^i$ of any focal element $A^i$ is attached to $A^i$ without any other specification about its distribution on the singletons $s^j$ contained in $A^i$. Then, one may consider the class $\Psi^{RS}$ of the probability distributions $P(s^j)$ that correspond to the infinite ways whereby $m^i$ may be *selected* to be distributed on $A^i$ (Fetz and Oberguggenberger 2004). Formally, indicating with $I^i$ the set of the indexes of the singletons in $A^i$ and with $\Psi^i$ the set of all probability distributions on the sub-space $A^i = \{s^j \,|\, j \in I^i\}$:

$$\Psi^{RS} = \left\{ \sum_{i=1}^{n} \left(m^i \cdot P^{i,j}\right) | \, P^{i,j} = P^i\left(s^j\right) \in \Psi^i, j \in I^i ; P^{i,j} = 0, j \notin I^i \right\} \tag{3.11}$$

In the case of finite spaces, the equivalence of the two definitions (i.e. $\Psi = \Psi^E = \Psi^{RS}$) has been demonstrated in (Dempster 1967). In the following, we focus on the second definition, which suggests constructive procedures to evaluate class $\Psi$. Each member $P$ of $\Psi$ takes on values as follows:

$$\forall s^j \in S : P\left(s^j\right) = \sum_{i=1}^{n} \left(m^i \cdot P^{i,j}\left(s^j\right)\right), \tag{3.12}$$

and depends on $k = \Sigma| I^i |$ values $P^{i,j}$, subject to the conditions:

$$\forall i : \sum_{j \in I^i} P^{i,j} = 1 \tag{3.13}$$

Sets $\Psi^i$ are convex because the convex combination of any pair of probability distributions in $\Psi^i$ is a probability distribution on $A^i$, and hence it is in $\Psi^i$. Convexity of the sets $\Psi^i$ and Eq. (3.12) imply that the set $\Psi$ is convex. Let us now underline the special meaning of some members of $\Psi$.

#### 3.2.3.1 White Distribution

If the Principle of Indifference is applied to each focal element, a uniform distribution is obtained along each focal element:

$$\forall j \in I^i : P^{i,j}\left(s^j\right) = \frac{1}{\left|I^i\right|} \;\; ; \;\; i = 1 \text{ to } n \tag{3.14}$$

In the following, this particular distribution will be termed "*white*" (Bernardini 1995) to underline some analogy with the concept of spectral uniformity ("*white noise*", "*white light*"). The same notion was previously suggested by Dubois and Prade (Dubois and Prade 1982; Dubois and Prade 1990).

#### 3.2.3.2 Selectors

In the following, we assume that $S$ is a finite set of real numbers, so that the mass distribution of the simple real function $y = \Sigma_j\, s^j\, \chi(s^j)$ coincides with the probability distribution of the singletons $P(s^j)$ (see Chapter 2.2).

According to (Yager 1991; Miranda, Couso et al. 2002) a random set is a strongly measurable multi-valued function (see §3.1.1), and it can be considered as generated by a random variable on $(X, \mathcal{X}, P_x)$ whose pointwise realizations on $S$ cannot be precisely observed, so we can only say that each realization is a particular but unknown singleton of the focal element. The mass distributions corresponding to the different point-valued mapping, compatible with the multi-valued mapping that generate focal elements, are called "*selectors*" of the random set: when focal element $A^i$ is observed, the realization is an unknown but specific singleton $\left\{s^{k^i}\right\}$ of the focal element (see Figure 3.5). Therefore:

$$P^{i,j} = \begin{cases} 1 & \text{if} \quad j = k^i \in I^i \\ 0 & \text{if} \quad j \neq k^i \end{cases} \tag{3.15}$$

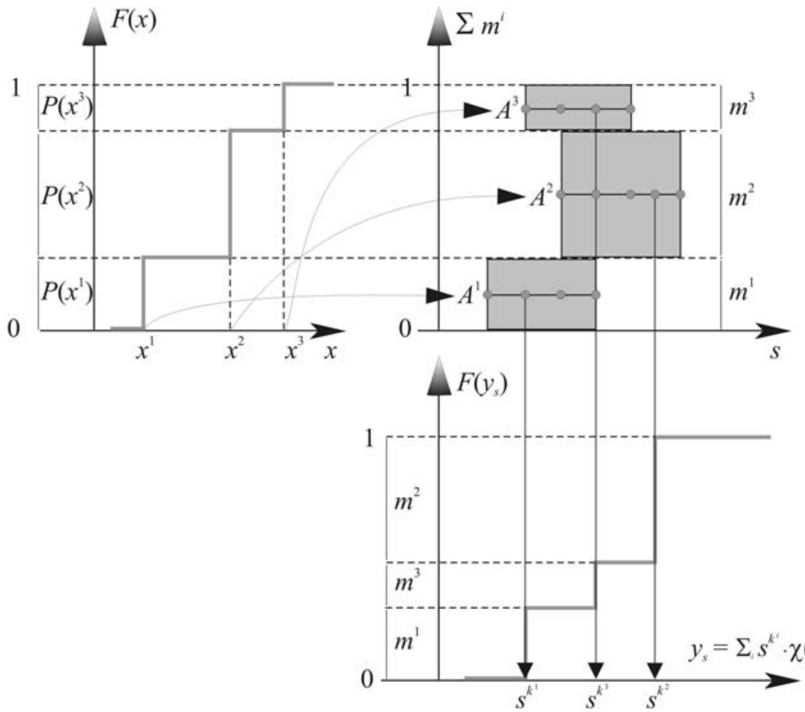



**Fig. 3.5** Selector from a random set

The set of selectors *SCT* is a subset of $\Psi$ because, with reference to Eq. (3.11), *SCT* is generated by considering all Dirac delta probability distributions (measures) in $\Psi^i$ (Section 2.1 on page 15), i.e. probability distributions concentrated at one point in $A^i$. The number of selectors, $|SCT|$, is thus at most equal to the product of the cardinalities $|I^i|$ (some selectors could coincide). If $X$ and $S$ are finite, the number of focal elements and cardinalities $|I^i|$ are finite: therefore *SCT* is a finite (and hence non-convex) set strongly included in the convex set $\Psi$.[N 3-1]

#### 3.2.3.3 Upper and Lower Distributions

Recall that $I^i$ is the set of indices of the singletons in $A^i$. Let us order each set $I^i$ according to the increasing real values of the singletons. Two particular selectors can be generated by taking $k^i$ equal to the upper bound of $I^i$ ($k^i = \max(I^i)$), and equal to the lower bound of $I^i$ ($k^i = \min(I^i)$). The corresponding (discrete) cumulative distribution functions (CDF) of $s$, or, equivalently, of the simple function $y_s = \Sigma_j s^j \chi(s^j)$ are, respectively:

$$\begin{aligned} F_{UPP}(y_s) &= \sum_{s^{\min(I^i)} \le y_s} m(A^i) = Pla\left(\left\{s^j \mid s^j \le y_s\right\}\right) \\ F_{LOW}(y_s) &= \sum_{s^{\max(I^i)} \le y_s} m(A^i) = Bel\left(\left\{s^j \mid s^j \le y_s\right\}\right) \end{aligned} \qquad (3.16)$$

They bound the CDF $F(y_s)$ given by any other probability distribution in $\Psi$. Figure 3.6 shows the upper and lower distributions of the random set considered in Example 3.5 and displayed in Figure 3.4.

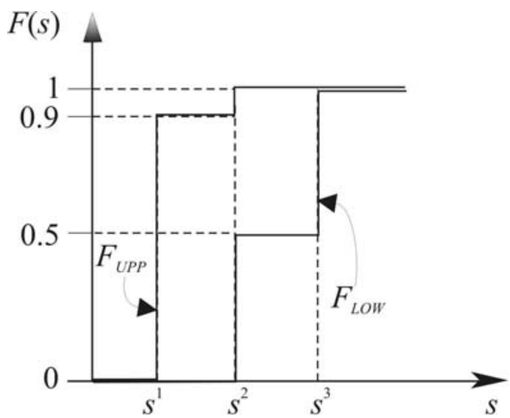



**Fig. 3.6** Upper and lower distributions of the random set in Example 3.5

More generally, if $f$ is a point-valued monotonically increasing or decreasing mapping on $S$ to $Y = \mathbb{R}$, the same upper and lower distribution functions extended to $Y$ can be used to bound the CDF $F(y)$ (Figure 3.7) and the expectation $E[y]$.

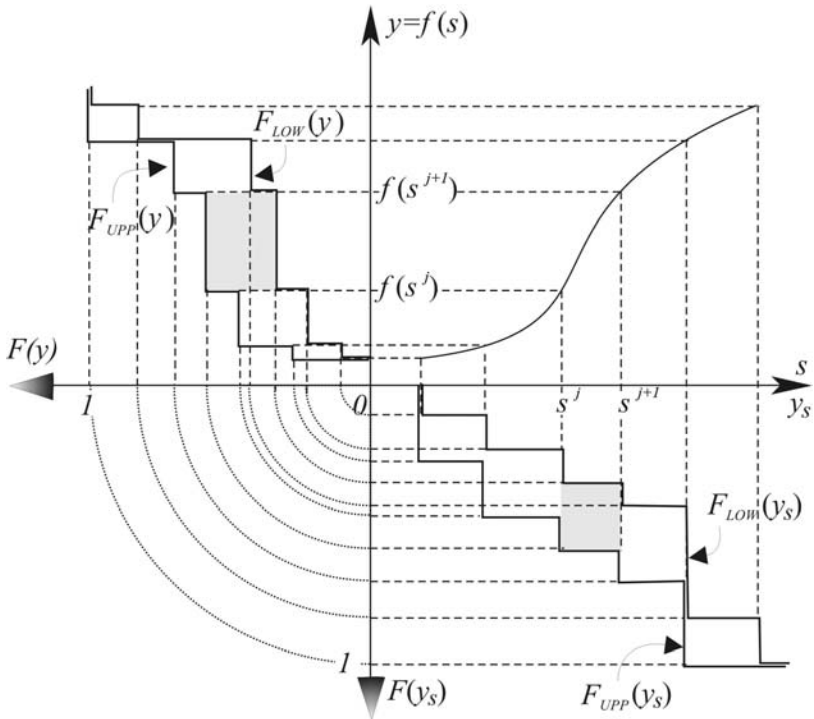



**Fig. 3.7** Extension of upper and lower distributions by a monotonic point-valued function

Hence:

$$
\begin{aligned}
F_{UPP}(y) &= \sum_{s^{\min(I^i)} \leq f^{-1}(y)} m(A^i) &= Pla(\{s^j \mid s^j \leq f^{-1}(y)\}) \\
F_{LOW}(y) &= \sum_{s^{\max(I^i)} \leq f^{-1}(y)} m(A^i) &= Bel(\{s^j \mid s^j \leq f^{-1}(y)\})
\end{aligned}
\tag{3.17}
$$

### 3.2.3.4 Extreme Distributions

Section 3.2.3.3 identified the two selectors that generate upper and lower CDFs for $s$, the simple function $y_s = \Sigma_j\, s^j\, \chi(s^j)$, or monotonic functions of $s$. These two selectors can also be interpreted as the selectors that, for a given $s^*$, give upper and lower probabilities of the event (set) $\{s : s \leq s^*\}$. Given any set $T \subseteq S$, one may wonder if $\Psi$ contains a distribution $P^*$ that maximizes (or minimizes) the probability of $T$, i.e. such that $P^*(T) = \max\{P(T) \mid P \in \Psi\}$ (or $P^*(T) = \min\{P(T) \mid P \in \Psi\}$), and, if it exists, how one can determine it. Such a distribution is called *extreme distribution*. One may then ask the same question for all sets $T \subseteq S$, i.e. determine all the distributions in $\Psi$ that minimize or maximize the probability of some event $T \subseteq S$. This is the set of the extreme distributions, *EXT*.

Likewise, for a general function $y = f(s)$, one may wonder if there is a distribution in $\Psi$ that gives the upper (or lower) CDF of $y$, and, if so, how to determine it. One may then ask the same question for all possible functions, and one may want to determine *all* of the distributions that give upper (or lower) CDFs of $y$ for some function $y = f(s)$. Are these distributions the same as in EXT? Let us start from examining the exteme distributions.

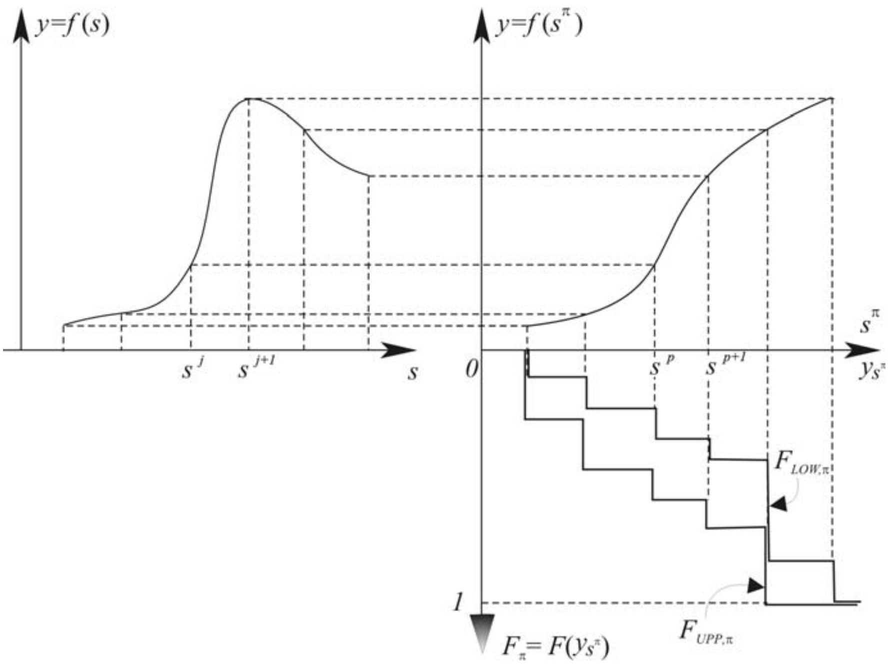



**Fig. 3.8** Extreme distributions and extension by a non monotonic point-valued function

Taking into account Eq. (3.12) subjected to the conditions in Eq. (3.13), more general extreme distributions can be constructed by considering permutations $\pi(j)$, $j = 1$ to $|S|$ of the indexes $j$, and progressively assigning the quantities $P^{i,\pi(j)}$ in such a way as to obtain an extreme (maximum or minimum) of each $P^{\pi(j)} = \Sigma_i\ m^i\ P^{i,\pi(j)}$. Selecting for example the minimum, the criterion corresponds to selecting systematically $P^{i,\pi(j)} = 0$ until, to satisfy conditions Eq. (3.13), $P^{i,\pi(j)} = 1$ must be assumed. By applying the maximum condition to the same permutation or, in an equivalent way, the minimum condition to the reversed permutation (i.e. the permutation $\pi_R(j) = |S| + 1 - \pi(j)$ ), a dual extreme distribution is obtained.

Formally, dual extreme cumulative distributions can be associated to any permutation $\pi(j)$, $j$ = 1 to $|S|$ of the indexes $j$, by extending Eq. (3.16) through the following formulas ($I^i_\pi = \{\pi(j) | j \in I^i\}$):

$$F_{UPP,\pi}(y_s) = \sum_{s^{\min(I_\pi^i)} \le y_s} m(A^i) = Pla(\{s^{\pi(j)} | s^{\pi(j)} \le y_s\})$$
$$F_{LOW,\pi}(y_s) = \sum_{s^{\max(I_\pi^i)} \le y_s} m(A^i) = Bel(\{s^{\pi(j)} | s^{\pi(j)} \le y_s\}) \tag{3.18}$$

Each extreme distribution $P_{EXT}(s^j)$ can finally be derived from each cumulative $F_\pi$ (upper and lower respectively):

$$P_\pi\left(s^{\pi(j)=k}\right) = \left\langle \begin{array}{ll} F_\pi(y_s = s^{\pi(j)=1}) & | \pi(j) = 1 \\ F_\pi(y_s = s^{\pi(j)=k}) - F_\pi(y_s = s^{\pi(j)=k-1}) & | \pi(j) > 1 \end{array} \right.$$
$$P_{EXT}\left(s^j\right) = P_\pi\left(s^{\pi^{-1}\pi(j)}\right) \tag{3.19}$$

This constructive definition shows that:

- extreme distributions are selectors: $EXT \subseteq SCT$;
- upper and lower distributions are particular extreme dual distributions corresponding to the pair of reversed permutations $\pi(j) = j$ and $\pi(j) = |S|+1-j$;
- the number $|EXT|$ of extreme distributions is at most equal to the number of permutations, i.e. $|S|!$; however the number can be lower because some permutations could give the same extreme;
- if the cardinalities $|I^i|$ reduce to 1 (the focal elements reduce to singletons), then $|EXT| = |\Psi|=1$, corresponding to a precise probability distribution;
- let $y_j = f(s^j)$ be a point-valued mapping on $S$ to $Y$ and $\pi_{LOW}(j)$, $\pi_{UPP}(j)$ a pair of reversed permutations so that $f(s^{\pi_{LOW}(j)})$ and $f(s^{\pi_{UPP}(j)})$ are an ordered list of monotonically increasing and decreasing values, respectively (Figure 3.8): the corresponding extension to $Y$ of the dual extreme distributions, obtained by formulae (3.19), can be used to bound the CDF $F(y)$ and the expectation $E[y]$;

- given that the probability of any event $A$ is equal to the expectations of its indicator function $I_A$, $Bel(A)$ and $Pla(A)$ are lower and upper bounds of $E[y = I_A]$ calculated with the corresponding dual extreme distributions; the same dual extreme distributions give bounds of $E[y = I_{A^C}]$ for the complementary event $A^C$, respectively upper $Pla(A^C)$ and lower $Bel(A^C)$;
- We already noticed that Eqs. (3.11) and (3.12) imply that $\Psi$ is a convex set. Since probabilities and expectations are linear functions of probabilities in $\Psi$, probabilities and expectations attain their maximum and minimum values at the vertices of $\Psi$. Therefore, *EXT* is the set of vertices of $\Psi$.
- $\Psi$ is equal to the convex hull of *EXT*, $\Psi^{EXT}$, i.e. the set of all convex combinations of extreme distributions (all extreme distributions are selectors of the random set and hence belong to $\Psi$).

**Example 3.6.** Let $S=\{s^1, s^2, s^3\}$, and $\mathcal{F}=\{(\{s^1, s^2\}, 0.5), (\{s^2\}, 0.3), (\{s^3\}, 0.2)\}$. Therefore: $|S| = 3$, $n = 3$, $I^1 = \{1,2\}$, $I^2 = \{2\}$, $I^3 = \{3\}$, $|I^1|\cdot|I^2|\cdot|I^3| = 2\cdot1\cdot1 = 2$. The set *SCT* contains two selectors:

$k^1 = 1$ gives: $P^1(s^1) = 0.5$, $P^1(s^2) = 0.3$, $P^1(s^3) = 0.2$;
$k^1 = 2$ gives: $P^2(s^1) = 0$, $P^2(s^2) = 0.3 + 0.5 = 0.8$, $P^2(s^3) = 0.2$.

The set $\Psi$ is the one-dimensional interval between the selectors. Of course they are also extreme points, corresponding to the upper and lower distributions (to the identity permutation and its reversal) respectively. The centre of the interval gives the white distributions:

$P_{WHITE}(s^1)= (0 + 0.5) /2 = 0.25$, $P_{WHITE}(s^2)= (0.3 + 0.8) /2 = 0.55$, $P_{WHITE}(s^3)= 0.2$ 0.

**Example 3.7.** Let $S=\{s^1, s^2, s^3\}$, and $\mathcal{F}=\{(\{s^1, s^2\}, 0.5), (\{s^1, s^2, s^3\}, 0.4), (\{s^2\}, 0.1)\}$ the random set considered in Example 3.5. Therefore: $|S| = 3$, n = 3, $I^1 = \{1,2\}$, $I^2 = \{1,2,3\}$, $I^3 = \{2\}$, $I = |I^1|\cdot|I^2|\cdot|I^3|=2\cdot3\cdot1=6$.

The white distribution can be calculated as follows:

$$P_{WHITE}(s^1) = 0.5/2 + 0.4/3 = 0.3833$$
$$P_{WHITE}(s^2) = 0.5/2 + 0.4/3 + 0.1/1 = 0.4833$$
$$P_{WHITE}(s^3) = 0.4/3 = 0.1333$$

The six selectors are listed in Table 3.3. For example, considering $r = 6$, $k^1 = 2$, $k^2 = 3$, $k^3 = 2$:

$$P_6(s^1) = 0\,,\qquad P_6(s^2) = m^1 + m^3 = 0.6,\qquad P_6(s^3) = m^2 = 0.4$$

**Table 3.3** Selectors in Example 2.5

| $r$ | $k^1$ | $k^2$ | $k^3$ | $P^r(s^1)$ | $P^r(s^2)$ | $P^r(s^3)$ | Total |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 2 | 0.9 | 0.1 | 0 | 1 |
| 2 | 1 | 2 | 2 | 0.5 | 0.5 | 0 | 1 |
| 3 | 1 | 3 | 2 | 0.5 | 0.1 | 0.4 | 1 |
| 4 | 2 | 1 | 2 | 0.4 | 0.6 | 0 | 1 |
| 5 | 2 | 2 | 2 | 0 | 1 | 0 | 1 |
| 6 | 2 | 3 | 2 | 0 | 0.6 | 0.4 | 1 |

**Table 3.4** Extreme distributions in Example 3.7

| $j$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| $\pi(j)$ | 1 | 2 | 3 | $m^i$ |
| $i$=1 | 0 | 1 | 0 | 0.5 |
| $i$=2 | 0 | 0 | 1 | 0.4 |
| $i$=3 | 0 | 1 | 0 | 0.1 |
| $P^1(s^{\pi(j)})$ | 0 | 0.6 | 0.4 | |
| $P^1{}_{EXT}(s^j)$ | 0 | 0.6 | 0.4 | |

| $j$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| $\pi(j)$ | 3 | 2 | 1 | $m^i$ |
| $i$=1 | 0 | 0 | 1 | 0.5 |
| $i$=2 | 0 | 0 | 1 | 0.4 |
| $i$=3 | 0 | 1 | 0 | 0.1 |
| $P^2(s^{\pi(j)})$ | 0 | 0.1 | 0.9 | |
| $P^2{}_{EXT}(s^j)$ | 0.9 | 0.1 | 0 | |

| $j$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| $\pi(j)$ | 2 | 1 | 3 | $m^i$ |
| $i$=1 | 0 | 1 | 0 | 0.5 |
| $i$=2 | 0 | 0 | 1 | 0.4 |
| $i$=3 | 1 | 0 | 0 | 0.1 |
| $P^3(s^{\pi(j)})$ | 0.1 | 0.5 | 0.4 | |
| $P^3{}_{EXT}(s^j)$ | 0.5 | 0.1 | 0.4 | |

| $j$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| $\pi(j)$ | 3 | 1 | 2 | $m^i$ |
| $i$=1 | 0 | 0 | 1 | 0.5 |
| $i$=2 | 0 | 0 | 1 | 0.4 |
| $i$=3 | 0 | 0 | 1 | 0.1 |
| $P^4(s^{\pi(j)})$ | 0 | 0 | 1 | |
| $P^4{}_{EXT}(s^j)$ | 0 | 1 | 0 | |

| $j$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| $\pi(j)$ | 2 | 3 | 1 | $m^i$ |
| $i$=1 | 0 | 0 | 1 | 0.5 |
| $i$=2 | 0 | 0 | 1 | 0.4 |
| $i$=3 | 1 | 0 | 0 | 0.1 |
| $P^5(s^{\pi(j)})$ | 0.1 | 0 | 0.9 | |
| $P^5{}_{EXT}(s^j)$ | 0.9 | 0.1 | 0 | |

| $j$ | 1 | 2 | 3 | |
|---|---|---|---|---|
| $\pi(j)$ | 1 | 3 | 2 | $m^i$ |
| $i$=1 | 0 | 0 | 1 | 0.5 |
| $i$=2 | 0 | 0 | 1 | 0.4 |
| $i$=3 | 0 | 0 | 1 | 0.1 |
| $P^6(s^{\pi(j)})$ | 0 | 0 | 1 | |
| $P^6{}_{EXT}(s^j)$ | 0 | 1 | 0 | |

In Table 3.4, rows 3 through 5 and columns 2 through 4 give values $P^{i,\pi(j)}$ used to construct extreme distributions; each row corresponds to a focal element. Framed cells identify focal elements where $P^{i,\pi(j)}$ can be selected between 0 and 1 when $i$ = 1 or 2; in all cases, $P^{3,\pi(j)=2} = 1$, because $|I^3| = 1$, and outside the framed cells $P^{i,\pi(j)} = 0$. In each row, the assignment of $P^{i,\pi(j)}$ values starts from the first column and continues on the second and third columns by inserting 0 in the framed cells until the last framed column is encountered, and 1 is assigned to this column.

The six sub-tables given in Table 3.4 show how at the most 3! = 6 extreme distributions $P^r_{EXT}$ are calculated by starting from the 3 possible pairs of reversed permutations. Four extreme points are identified, which correspond to selectors $P^1$, $P^3$, $P^5$, and $P^6$ in Table 3.3. $P^1 = P^1_{EXT}$ and $P^6 = P^2_{EXT}$ are the upper and lower distributions.

The set $\Psi$ is the $|S|$-1 = 2-dimensional shaded polyhedron shown in Figure 3.9a) in the $|S|$ = 3 dimensional space of the probabilities $P(s^j)$, $j$ = 1 to $|S|$ = 3. Figure 3.9b) depicts the projection of $\Psi$ in the $|S|$-1 = 2 dimensional space of the probabilities $P(s^j)$, $j$ = 1 to $|S|$-1 = 2 (at every point, $P(s^{|S|}) = 1 - \Sigma_j P(s^j)$). $\Psi$ is a quadrangle whose vertexes are the four extreme distributions. Since the selectors $P^2$ and $P^4$ are not extreme distributions, they lie on the boundary of $\Psi$, and more precisely on the edge that connects extreme distributions $P^1$ and $P^5$. Therefore, selectors $P^2$ and $P^4$ may be obtained as convex combinations of the two extreme distributions $P^1$ and $P^5$.[N 3-2]

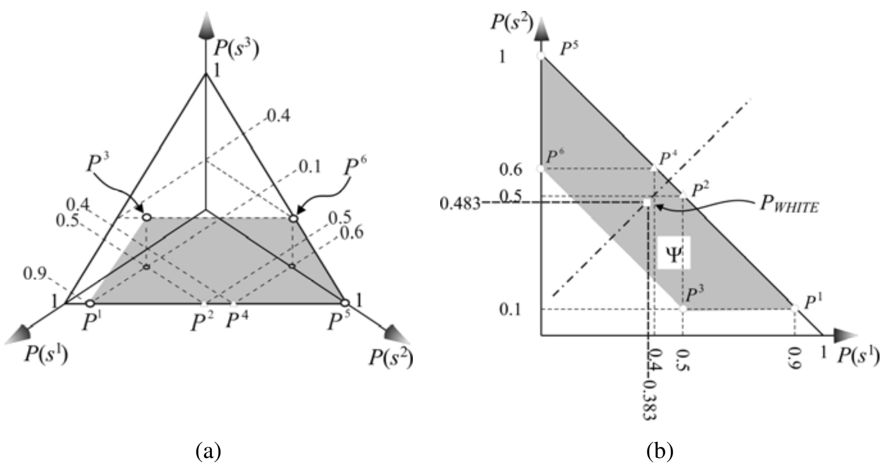



**Fig. 3.9** Set of probability distributions, selectors, extreme and white distributions in Example 3.7: in the 3-dimensional space (a) and its projection in the 2-dimensional sub-space (b)

The reader is encouraged to check that the white distribution is equal to the mean value of all selectors (extreme distributions) in *SCT* (*EXT*) weighted by the multiplicities of the coincident selectors (extreme distributions).

Let us now consider the following function that maps $S$ to $Y$: ($y_1 = f(s^1) = 80$; $y_2 = f(s^2) = 160$; $y_3 = f(s^3) = 140$). Bounds on $E[y]$ can easily be evaluated through the permutation: $\pi(j{=}1) = 1$, $\pi(j{=}2) = 3$, $\pi(j{=}3) = 2$, and the reversed one, which lead to monotonically increasing and decreasing functions, respectively. Therefore, the expectation bounds are the expectations obtained through the probability distributions $P^6{}_{EXT} = P^5$ and $P^5{}_{EXT} = P^1$:

$$E_{LOW}[y] = 0.9 \text{ x } 80 + 0.1 \text{ x } 160 + 0 \text{ x } 140 = 88$$
$$E_{UPP}[y] = 0 \text{ x } 80 + 1 \text{ x } 160 + 0 \text{ x } 140 = 160$$

By considering the indicator function of any subset $T \subseteq S$ as a particular mapping, one can check that *Bel* and *Pla* are probability bounds. For example, let $T = \{s^1, s^2\}$, $T^C = \{s^3\}$ and therefore $I_T$: ($y_1 = \chi(s^1) = 1$; $y_2 = \chi(s^2) = 1$; $y_3 = \chi(s^3) = 0$), $I_{T^C}$: ($y_1 = \chi(s^1) = 0$; $y_2 = \chi(s^2) = 0$; $y_3 = \chi(s^3) = 1$). Observe that:

$$E_{LOW}[y{=}I_T] = E_{P=P^1EXT}[y{=}I_A] = 0 \text{ x } 1 + 0.6 \text{ x } 1 + 0.4 \text{ x } 0 = 0.6 = Bel(T)$$

$$E_{UPP}[y{=}I_T] = E_{P=P^1EXT}[y{=}I_A] = 0.9 \text{ x } 1 + 0.1 \text{ x } 1 + 0 \text{ x } 0 = 1.0 = Pla(T)$$

$$E_{LOW}[y{=}I_{T^C}] = E_{\mathrm{P=P^2EXT}}[y{=}I_{A^C}] = 0.9 \text{ x } 0 + 0.1 \text{ x } 0 + 0 \text{ x } 1 = 0 = Bel(T^C)$$

$$E_{UPP}[y{=}I_{T^C}] = E_{P=P^1EXT}[y{=}I_{A^C}] = 0 \text{ x } 0 + 0.6 \text{ x } 0 + 0.4 \text{ x } 1 = 0.4 = Pla(T^C)$$

### 3.2.3.5 Algorithm to Calculate Extreme Distributions

Let: $s = |S|$; $n = |\mathcal{F}|$; $\mathbf{B}_{s\times n}$ matrix with entries: $B_{j,i} = 1$ if $s^j \in A^i$, 0 otherwise; $\mathbf{m}_{n\times 1}$ = column vector with $i$-th component equal to $m(A^i)$; $\mathbf{p}_r$ = $r$-th probability distribution column vector; $\mathbf{P}_{\pi(j)}$ = row permutation matrix (identity matrix with rows rearranged according to permutation $\pi(j)$). Since permutation matrices are orthogonal, $\mathbf{P}_{\pi(j)}\mathbf{P}_{\pi(j)}{}^{\mathrm{T}} = \mathbf{I}$, and the reverse permutation is effected by $\mathbf{P}_{\pi(j)}{}^{\mathrm{T}}$. Example 3.7 leads us to the following algorithm:

Algorithm 1
DO $r = 1$ to $s!$
  $\mathbf{C} = \mathbf{P}_{\pi(j)}\,\mathbf{B}$
  $\mathbf{C} \leftarrow$ Set columns of $\mathbf{C}$ equal to zero, except for their last non-zero component
  $\mathbf{p}_r = \mathbf{P}_{\pi(j)}{}^{\mathrm{T}}\,\mathbf{C}\,\mathbf{m}$
END DO

The extreme distributions are the vertices of the convex hull ($Conv(\mathbf{p})$) in the space of all possible probability distribution vectors $\mathbf{p}_r$ so calculated. Computational Geometry (O'Rourke 1998; de Berg, van Kreveld et al. 2000; Sack and Urrutia 2000; Goodman and O'Rourke 2004) has devised very efficient algorithms for calculating convex hulls (e.g., Qhull (The Geometry Center 2007), and libraries LEDA (Algorithmic Solutions 2007) and CGAL (Pion 2007)).

Before adopting any of these algorithms, the following must be borne in mind: if a focal element contains only one element, say $s^k$, which is not in any other focal element, then the probability of $s^k$ is precise and equal to $m\left(\left\{s^k\right\}\right)$. As a consequence, along the *precise direction* on the space of probability distributions $P(s^k)$, $Conv(\mathbf{p})$ degenerates onto a single point of coordinate $m\left(\left\{s^k\right\}\right)$ and there is no need to include $s^k$ in the permutations of Algorithm 1.

Let us re-order the numbering of elements in $S$ so that the first $s_{imp}$ elements generate imprecise probability directions, and let us restrict ourselves to the *reduced space of probability distributions* $S_{imp} = \{s^j, j=1,.., s_{imp}\} \subseteq S$. Let $m_{imp}=\sum_{i=1,\ldots,s_{imp}} m\left(A^i\right)$ be the total probability weight in $S_{imp}$.

Since all points of coordinates $\mathbf{p}_r$ lie on plane $\sum_{j=1,\ldots,s_{imp}} P\left(s^j\right) = m_{imp}$, $Conv(\mathbf{p})$ is a subset of this plane, and is thus degenerate in the $s_{imp}$-dimensional space. For example, Figure 3.9a depicts $Conv(\mathbf{p})$ as the polygon $P^1P^5P^6P^3$ in the three-dimensional space $(P(s^1), P(s^2), P(s^3))$.

Under these circumstances, convex hull algorithms may fail. It is thus necessary to:

- Work in the reduced possibility space $S_{imp}$.
- Project points $\mathbf{p}_r$ orthogonal to any direction (e.g., $\mathbf{p}_r \equiv (p_{r,1},\ldots, p_{r,s_{imp}}) \rightarrow (p_{r,1},\ldots, p_{r,s_{imp}-1})$. Since the plane $\sum_{j=1,\ldots,s_{imp}} P\left(s^j\right) = m_{imp}$ makes the same angle with all directions, there is no preferred projection direction. The remaining coordinates project $\Psi$ in a space that will be termed the *projected imprecise space of probability distributions* as exemplified by $(P(s^1), P(s^2))$ in Figure 3.9b.
- Calculate $Conv(\mathbf{p})$ in the projected imprecise space of probability distributions.

The general algorithm is as follows:

```
DO i= 1 to s
      IF row i of B contains only one 1 at location j, AND column j
         contains only one 1, THEN
         Drop row i and column j of B
         Store i and j array "index"
         Drop i-th element from m
         n←n-1
         s←s-1
      IF i==s, EXIT
END DO
DO r= 1 to s!
   C = P_π(j) B
   C ←Set columns of C equal to zero, except for their last non-zero
      component
   p_r = P_π(j)^T Cm
   Drop last component from p_r
END DO
Calculate Conv(p)
Reconstruct vertices of Conv(p) in the initial  space using array "index"
```

An alternative algorithm to calculate the set of the extereme distributions will be presented in the next Chapter 4.2 (page 114).

**Example 3.8.** Let $S=\{s^1, s^2, s^3, s^4\}$, and $\mathcal{F} = \{(\{s^1, s^3\}, 0.25), (\{s^3, s^4\}, 0.30), (\{s^2\}, 0.10), (\{s^1, s^3, s^4\}, 0.15), (\{s^4\}, 0.20)\}$. The original matrix **B** is as follows:

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Row 2 contains only one 1 at location 3, and column 3 contains only one 1; row 2 and column 3 are dropped. No other row contains only one 1 and the search is over (column 5 contains only one 1 at location 4, but row 4 contains 3 1s). Permutations in the projected imprecise space are given in Table 3.5. Vertices of *Conv*(**p**) are permutations 1 through 4 and 6, and are given in Table 3.6.

**Table 3.5** Example 3.8: Distributions in the projected imprecise space of probability distributions

| No. | Permutation $\pi(1)$ | $\pi(2)$ | $\pi(3)$ | Distribution |
|---|---|---|---|---|
| 1 | 1 | 2 | 3 | {0., 0.25, 0.65} |
| 2 | 1 | 3 | 2 | {0., 0.7, 0.2} |
| 3 | 2 | 1 | 3 | {0.25, 0., 0.65} |
| 4 | 2 | 3 | 1 | {0.4, 0., 0.5} |
| 5 | 3 | 1 | 2 | {0., 0.7, 0.2} |
| 6 | 3 | 2 | 3 | {0.4, 0.3, 0.2} |

**Table 3.6** Example 3.8: Extreme distributions in the original space of probability distributions

| No. | Permutation | Distribution |
|---|---|---|
| 1 | 1 | {0.00, 0.10, 0.25, 0.65} |
| 2 | 2 | {0.00, 0.10, 0.70, 0.20} |
| 3 | 3 | {0.25, 0.10, 0.00, 0.65} |
| 4 | 4 | {0.40, 0.10, 0.00, 0.50} |
| 5 | 6 | {0.40, 0.10, 0.30, 0.20} |

### *3.2.4 Consonant Random Sets*

Let us consider the particular case of a random set whose $n$ focal elements are *nested*, i.e. can be ordered in such a way that:

$$A^1 \subseteq A^2 \subseteq .... \subseteq A^n \tag{3.20}$$

A random set satisfying (3.20) is called *consonant*, according to a definition introduced by Shafer (Shafer 1976). This definition highlights that the given information is clearly centered around $A^1$, at least with probability $Bel(A^1) = m(A^1)$ and possibly with probability $Pla(A^1) = 1$, and surely contained in $A^n$ ($Bel(A^n) = Pla(A^n) = 1$). Every focal element $A^i$ does not conflict (i.e. includes and therefore strengthens) with the information given by $A^1, A^2, \ldots A^{i-1}$. Of course every random set with only one focal element ($n$=1) is consonant, and describes a fully non-probabilistic set uncertainty about a variable in the space $S$.

Given a set $T \subset S$ it is possible to find $k$ such that: for $i = 1$ to $k$: $T \cap A^i = \emptyset$, $A^i \subset T^C$ ; for $i = k+1$ to $n$: $T \cap A^i \neq \emptyset$, $A^i \not\subset T^C$ (Figure 3.10). Therefore:

$$\begin{aligned} Pla(T) &= 1 - Bel\left(T^C\right) = \sum_{i=k+1}^{n} m(A^i) \\ Bel\left(T^C\right) &= 1 - Pla(T) = \sum_{i=1}^{k} m(A^i) \end{aligned} \tag{3.21}$$

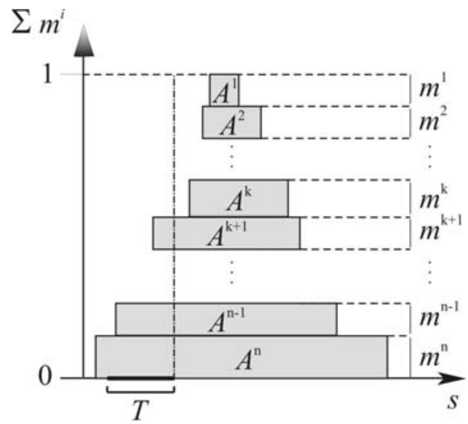



**Fig. 3.10** *Pla* and *Bel* of a set *T* by a consonant random set

It follows that, given two sets, $T^1$ and $T^2$, both $\subset S$, consonant random sets satisfy the fundamental properties (Figure 3.11):

$$\begin{aligned} Pla\left(T^1 \cup T^2\right) &= \sum_{i=(k=\min(k_1,k_2))+1}^{n} m(A^i) = \max\left(Pla(T^1, Pla(T^2)\right) \\ Bel\left(T^{1,C} \cap T^{2,C}\right) &= Bel\left((T^1 \cup T^2)^C\right) = \sum_{i=1}^{k=\min(k_1,k_2)} m(A^i) = \min\left(Bel(T^{1,C}), Bel(T^{2,C})\right) \end{aligned} \tag{3.22}$$

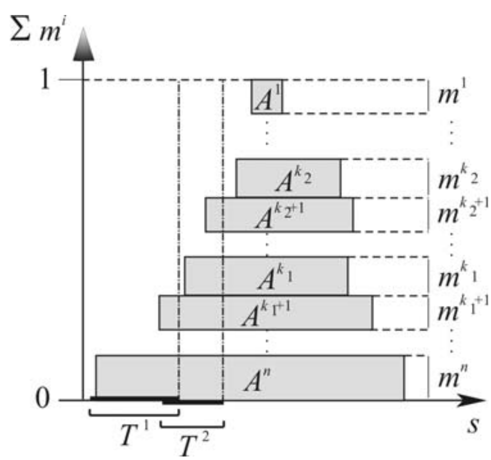



**Fig. 3.11** *Pla* and *Bel* of the union or intersection of two sets $T^1$ and $T^2$ by a consonant random set

Of course the second equation in (3.22) implies that, for any pair of sets $T^1$ and $T^2$, both $\subset S$:

$$Bel\left(T^1 \cap T^2\right) = \min\left(Bel(T^1), Bel(T^2)\right); \tag{3.23}$$

The first equation in (3.22) demonstrates that Plausibility measures of consonant random sets satisfy, as well as classical Probability measures, the "*decomposability property*": the measure of uncertainty of the union of any pair of disjointed sets is dependent solely on the measures of uncertainty of the individual sets. Therefore, in the case of a consonant random set, the point-valued *contour function* (Shafer 1976) μ: $S \to [0, 1]$:

$$\mu(s^j) = Pla\left(\left\{s^j\right\}\right) \tag{3.24}$$

completely defines the information on the measures of any subset $T \subset S$, exactly how the probability distribution $P(s^j) = P(\{s^j\})$ defines, although through a different rule (the additivity rule), the probability of every subset $T$ in the algebra generated by the singletons. In fact:

$$Pla(T) = \max_{s^j \in T} \mu\left(s^j\right); \qquad Bel(T) = 1 - \max_{s^j \in T^c} \mu\left(s^j\right) \tag{3.25}$$

Moreover the Möbius inversion (3.8) of the set function *Bel* allows the (nested) family of focal elements to be recognised through the *m* set function.

More directly, let us assume:

$$\begin{aligned}
&\alpha^1 = \max_j \left(\mu\left(s^j\right)\right) = 1 \\
&\alpha^i = \max_{j \mid \mu\left(s^j\right) < \alpha^{i-1}} \left(\mu\left(s^j\right)\right) \\
&\alpha^n = \max_{j \mid \mu\left(s^j\right) < \alpha^{n-1}} \left(\mu\left(s^j\right)\right) = \min_j \left(\mu\left(s^j\right)\right) \\
&\alpha^{n+1} = 0
\end{aligned} \tag{3.26}$$

The family of the focal elements and related probabilistic assignments (summing up to 1) are given by:

$$\begin{aligned}
&A^i = \left\{s^j \in S \mid \mu\left(s^j\right) \geq \mu\left(s^k\right) = \alpha^i > \mu\left(s^{k+1}\right) = \alpha^{i+1}\right\}; \\
&m^i = \alpha^i - \alpha^{i+1}
\end{aligned} \tag{3.27}$$

The number of focal elements, $n$, related to the $n$ α-levels $\alpha^i$, is therefore equal to the cardinality of the image of $S$ through μ; of course $\leq |S|$, considering that some singletons could map to the same value of plausibility. Hence the set $M$ of the α-levels is given by:

$$M = \left\{ \alpha^1 = 1, \ \ldots \ , \ \alpha^i = \sum_{j:A^j \supseteq A^i} m^j\left(A^j\right) > \alpha^{i+1}, \ \ldots \ , \ \alpha^n = m^n \right\} \tag{3.28}$$

**Example 3.9.** Let $S = \{s^1, s^2, s^3\}$, and $\mathcal{F} = \{(\{s^1, s^2\}, 0.5), (\{s^1, s^2, s^3\}, 0.4), (\{s^2\}, 0.1)\}$ the random set considered in Example 3.5 and Example 3.7. The focal elements can be reordered as follows:

$$A^1 = \{s^2\} \subset A^2 = \{s^1, s^2\} \subset A^3 = \{s^1, s^2, s^3\}$$

demonstrating that the random set is consonant, with contour function:

$$\mu(s^2) = m(A^1)+m(A^2)+m(A^3) = 1 > \mu(s^1) = m(A^2)+m(A^3) = 0.9 > \mu(s^3) = m(A^3) = 0.4$$

Starting from the contour function the random sets can be reconstructed assuming: $\alpha^1 = 1$; $\alpha^2 = 0.9$ , $\alpha^3 = 0.4$; therefore, according to Eq. (3.27):

$$A^1 = \{s^j | \mu(s^j) \geq \alpha^1\} = \{s^2\}, \qquad m(A^1) = \alpha^1 - \alpha^2 = 1 - 0.9 = 0.1$$

$$A^2 = \{s^j | \mu(s^j) \geq \alpha^2\} = \{s^1, s^2\}, \qquad m(A^2) = \alpha^2 - \alpha^3 = 0.9 - 0.4 = 0.5$$

$$A^3 = \{s^j | \mu(s^j) \geq \alpha^3\} = \{s^1, s^2, s^3\}, \quad m(A^3) = \alpha^3 = 0.4$$

For example, compare with values in Table 3.2 that:

$$Pla(\{s^1, s^3\} = \max\ (\mu(s^1), \mu(s^3)) = 0.9$$

$$Bel(\{s^1, s^3\} = 1 - \max\ (\mu(s^2)) = 1 - 1 = 0$$

There is a narrow correspondence between consonant random sets and other decomposable measures of uncertainty: fuzzy sets and theory of possibility.

In a context totally separate from the research of Dempster and Shafer, the idea of fuzzy set was developed by L. Zadeh in the 1960s (Zadeh 1965) as an extension of classical set theory. He suggested that the membership to a subset $A$ of a universal set $S$ (finite or infinite, for example the real numbers $\mathbb{R}$) could not always be a crisp property, absolutely verified or non-verified, but also, in some cases, partially verified. So the boundaries between $A$ and $A^C$ should be considered as separated by a fuzzy zone, where the *Law of excluded middle* is no more valid and the singleton $\{s\}$ partly belongs both to $A$ and $A^C$.

Formally the extension was obtained through the generalization of the indicator function of a classical crisp subset $A \subseteq S$, $\chi$: $S \rightarrow \{0, 1\}$, to the *membership function*: $\mu$: $S \rightarrow [0, 1]$.

We are thus facing a logic (fuzzy logic) in which propositions exist with a degree of truth that goes from 0 (false) to 1 (true), every gradation being permitted in between. This logic seems to closely match human thinking when a true-false judgment has to be given about complex propositions relative to the real world, as will be seen in the following Chapter 6.

The *height* of a fuzzy subset $A$ is the largest membership grade obtained by any singleton: $h(A) = \max \mu(s) \mid s \in S$. A fuzzy set is called *normal* when its height is equal to 1, or else *subnormal*.

But here, on the contrary, interest is focused on the connection between concept of fuzzy set and concept of random set, as uncertain measures of a family of subsets.

This connection can be clearly envisaged using the dual representation of a fuzzy set through their $\alpha$-cuts ${}^{\alpha}A$. They are classical subsets of $S$ defined, for any selected value $\alpha$ of membership, by the formula:

$$ {}^{\alpha}A = \{ s \in S \mid \mu(s) \geq \alpha \} \; ; \tag{3.29} $$

When a fuzzy set is implicitly given through the (finite or infinite) nested sequence of its $\alpha$-cuts ${}^{\alpha}$A for some $\alpha$-levels $\alpha$, its membership function can be reconstructed through the following equation (*decomposition* theorem):

$$ \mu\left(s^j\right) = \max_{\alpha} \quad \min\left(\alpha, \chi_{{}^{\alpha}A}\left(s^j\right)\right) \tag{3.30} $$

where $\chi_{{}^{\alpha}A}(s)$ is the indicator function of the classical subset ${}^{\alpha}$A.

**Example 3.10.** Let $S = \{1, 2, 3, 4, 5\}$ and $A$ be a fuzzy subset of $S$ defined by the membership grades: $\mu_A(1) = \mu_A(5) = 0$; $\mu_A(2) = \mu_A(4) = 0.5$; $\mu_A(3) = 1$.

Then: ${}^{0}A = S = \{1, 2, 3, 4, 5\}$; ${}^{0.5}A = \{2, 3, 4\}$; ${}^{1}A = \{3\}$

From (3.30) we obtain, for example:

$$ \mu_T(1) = \max(\min(0, 1), \min(0.5, 0), \min(1, 0)) = 0 $$
$$ \mu_T(2) = \max(\min(0, 1), \min(0.5, 1), \min(1, 0)) = 0.5 $$
$$ \mu_T(3) = \max(\min(0, 1), \min(0.5, 1), \min(1, 1)) = 1 $$

So it is clear that the $\alpha$-cuts ${}^{\alpha}A$ give the non-specificity of uncertainty, as do the focal elements of a random set, while the associated values of $\alpha$ are related to the strife between the different $\alpha$-cuts. Moreover, comparing formula

(3.29) with (3.27), it is clear that the α-cuts $^{\alpha}A$ of any given normal fuzzy set are a nested sequence of subsets of set *S*, and therefore the family of focal elements of an associated consonant random set: the membership function of normal fuzzy sets gives the contour function of the corresponding random sets, and the basic probabilistic assignment (for a finite sequence of α-cuts) is given by $m\left(A^i = {}^{\alpha^i}A\right) = \alpha^i - \alpha^{i+1}$.

Considering Eq. (3.25) from this point of view, the membership function of a fuzzy subset *A* allows measures of Plausibility and Belief to be attached to every classical subset $T \subseteq S$; this very different interpretation of fuzzy sets was recognized by Zadeh himself in 1978 (Zadeh 1978), as the basis of a theory of *Possibilities,* defined by a *possibility distribution*

$$\pi(s) = \mu_A(s) \tag{3.31}$$

later extensively developed by other authors, in particular Dubois and Prade (Dubois and Prade 1988). In a comparison between Evidence Theory and Possibility Theory, *Necessity* (*Nec*(*T*)) and *Possibility* (*Pos*(*T*)) measures of any classical subset $T \subseteq S$ coincide respectively with Belief (*Bel*(*T*)) and Plausibility (*Pla*(*T*)) deriving from the associated consonant random set through Eq. (3.25).

This comparison suggests a probabilistic (objective or subjective) content of information summarized by a fuzzy set, as we will see in section 6.4. Of course white distribution, selectors, upper and lower distributions, and extreme distributions can also be evaluated for any consonant random set or the corresponding fuzzy set.

**Example 3.11.** Let $S=\{s^1, s^2, s^3\}$, and $\mathcal{F}=\{(A^1=\{s^2\}, 0.8), (A^2=\{s^1, s^2\}, 0.1), (A^3=\{s^1, s^2, s^3\}, 0.1)\}$. Table 3.7 lists 5 selectors (one with double multiplicity). Original matrixes **B** and **m** defined in §3.2.3.5 are as follows:

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{m} = \begin{pmatrix} 0.8 \\ 0.1 \\ 0.1 \end{pmatrix}$$

They cannot be reduced; the procedure allows the four extreme distributions listed in Table 3.8 to be derived (two with double multiplicity). The selector $P_4$ is not extreme: in fact it is the mean between $P_1$ and $P_5$. The mean value of the columns in Table 3.7 and Table 3.8 are coincident and equal to the white distribution, confirming the property given in §3.2.3.4, Example 3.7.

**Table 3.7** Selectors in Example 3.11

| $r$ | $k^1$ | $k^2$ | $k^3$ | $P^r(s^1)$ | $P^r(s^2)$ | $P^r(s^3)$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 1 | 0.2 | 0.8 | 0 |
| 2 | 2 | 1 | 2 | 0.1 | 0.9 | 0 |
| 3 | 2 | 1 | 3 | 0.1 | 0.8 | 0.1 |
| 4 | 2 | 2 | 1 | 0.1 | 0.9 | 0 |
| 5 | 2 | 2 | 2 | 0 | 1 | 0 |
| 6 | 2 | 2 | 3 | 0 | 0.9 | 0.1 |
| Mean | | | | 0.0833 | 0.8833 | 0.0333 |

**Table 3.8** Extreme distributions in Example 3.11

| No. | π(1) | π(2) | π(3) | $P_{EXT}(s^1)$ | $P_{EXT}(s^2)$ | $P_{EXT}(s^3)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 0.2 | 0.8 | 0 |
| 2 | 1 | 3 | 2 | 0.2 | 0.8 | 0 |
| 3 | 2 | 1 | 3 | 0 | 1 | 0 |
| 4 | 2 | 3 | 1 | 0 | 1 | 0 |
| 5 | 3 | 1 | 2 | 0.1 | 0.8 | 0.1 |
| 6 | 3 | 2 | 1 | 0 | 0.9 | 0.1 |
| Mean | | | | 0.0833 | 0.8833 | 0.0333 |

### 3.2.5 Conditioning

According to Bayes' Theorem, when an event $B$ with *a priori* positive probability $P(B)$ is observed, the *posterior* probability changes to:

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A\cap B)}{P(A\cap B)+P\left(A^C\cap B\right)} \tag{3.32}$$

It is quite natural for a random set to assume:

$$Bel(A/B)=\inf_{P\in\Psi} P(A/B);\qquad Pla(A/B)=\sup_{P\in\Psi} P(A/B) \tag{3.33}$$

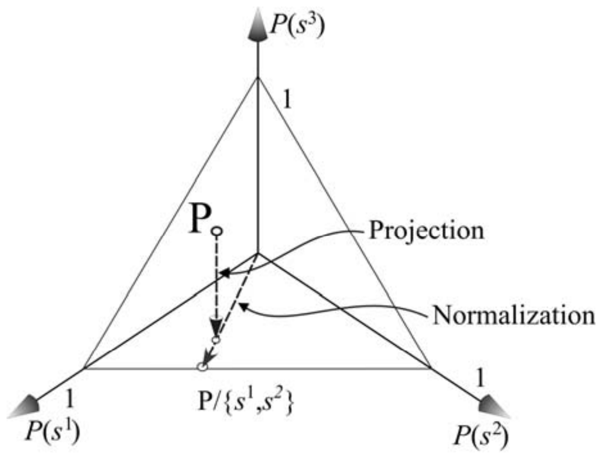
For a random set in a finite space and therefore with a finite number of extreme distributions $EXT$:

$$Bel(A/B)=\min_{P\in EXT} P(A/B);\quad Pla(A/B)=\max_{P\in EXT} P(A/B) \tag{3.34}$$

An explicit solution can easily be derived observing that the same extreme distribution gives the lower bound of $P(A\cap B)$ and the upper bound of $P(A^C\cap B)$ in the numerators and denominators appearing in Eq. (3.32); moreover the dual extreme distribution gives the corresponding upper bound of $P(A\cap B)$ and lower bound of $P(A^C\cap B)$. Therefore:

$$Bel(A/B)=\frac{Bel(A\cap B)}{Bel(A\cap B)+Pla(A^C\cap B)}$$
$$Pla(A/B)=\frac{Pla(A\cap B)}{Pla(A\cap B)+Bel(A^C\cap B)} \tag{3.35}$$

**Fig. 3.12** Conditioning a probability distribution $P/B = \{s^1, s^2\}$




Conditioning can also be defined by a rule to derive the posterior set $\Psi/B$ from the set $\Psi$ of the probability distributions compatible with the a priori random set. For the probabilities of the singleton, Bayes' Theorem gives directly, for any $P\in\Psi$ (Figure 3.12):

$$P\left(\left\{s^j\right\}/B\right)=\left\langle \begin{array}{ll} 0 \text{ if } s^j \notin B & \text{(projection)} \\ P\left(s^j\right)/P(B) & \text{(normalization)} \end{array} \right| \tag{3.36}$$

**Example 3.12.** Let $S=\{s^1, s^2, s^3\}$, and $\mathcal{F}=\{(\{s^1, s^2\}, 0.5), (\{s^1, s^2, s^3\}, 0.4), (\{s^2\}, 0.1)\}$ the random set considered in Example 3.5 and Example 3.7. Suppose that the event $B = \{s^1, s^2\}$ has been observed (event $\{s^3\}$ must be excluded). We are interested in evaluating bounds of the posterior probability of the event $T = \{s^1\}$:

$T \cap B = \{s^1\} \cap \{s^1, s^2\} = \{ s_1\}$ ;  $T^C \cap B = \{s^2, s^3\} \cap \{s^1, s^2\} = \{s^2\}$;

$Bel(T \cap B) = Bel(\{s^1\}) = 0$ ;  $Pla(T \cap B) = Pla(\{s^1\}) = 0.9$

$Bel(T^C \cap B) = Bel(\{s^2\}) = 0.1$;  $Pla(T^C \cap B) = Pla(\{s^2\}) = 1.0$

Therefore:

$Bel(\{s^1\}/\{s^1, s^2\}) = 0/(0+1) = 0$;  $Pla(\{s^1\}/\{s^1, s^2\}) = 0.9/(0.9+0.1) = 0.9$

The same results can be obtained applying Bayes' Theorem respectively to the extreme distributions $P^6_{EXT}$ and $P^5_{EXT}$:

$P^6_{EXT}(\{s^1\}/B) = P^6_{EXT}(\{s^1\} \cap B)/ P^6_{EXT}(B) = 0 / 1 = 0$

$P^5_{EXT}(\{s^1\}/B) = P^5_{EXT}(\{s^1\} \cap B)/ P^5_{EXT}(B) = 0.9 / 1 = 0.9$

In a similar way:

$Bel(\{s^2\}/\{s^1, s^2\}) = 0.1/(0.1+0.9) = 0.1$; $Pla(\{s^2\}/\{s^1, s^2\}) = 1/(1+0) = 1$

and of course:

$Bel\ (T=\{s^1, s^2\}/B = \{s_1, s_2\}) = Pla(T=\{s^1, s^2\}/B = \{s^1, s^2\}) = 1$

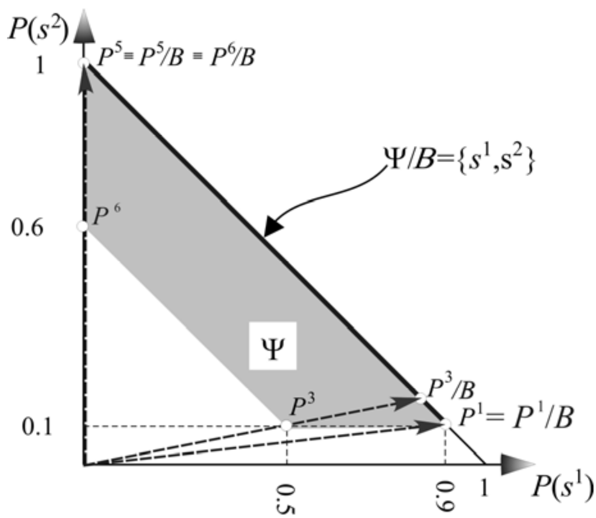



**Fig. 3.13** Conditioning random set in Example 3.7 and Example 3.12 with respect to $B = \{s^1, s^2\}$

Conditional *Bel* and *Pla* define a random set on the space $B=\{s^1, s^2\}$, whose probabilistic assignment can evaluated through the Möbius inversion (Eq. (3.8)). In the considered example:

$m(\{s^2\}/B) = 0.1 = m(\{s^2\})$; $m(\{s^1, s^2\}/B) = 0.9 = m(\{s^1, s^2\}) + m(\{s^1, s^2, s^3\})$.

Figure 3.13 shows how the conditional set $\Psi/B$ can be derived by simple geometric rules normalizing the extremes of $\Psi$ projected on the sub-space ($P(s^1)$, $P(s^2)$): the posterior bounds to $P(s^1)$ and $P(s^2)$ can therefore be easily checked.

In the simple case given in Example 3.12 the probabilistic assignment of each focal element $A$ of the a priori random set is transferred to the focal element $A \cap B$ of the posterior random set conditional to $B$. But a general closed rule giving the posterior probabilistic assignment corresponding directly to Eq. (3.35) cannot be given. For example: $A \cap B$ could be an empty set $\varnothing$, and an empty set cannot be a focal element for the posterior random set. This question will be considered later in the book, and an approximate rule to solve the problem will be given (Section 6.3.1).

Moreover the coincidence between the set $\Psi/B$ obtained by the Bayes' Rule (or procedure (3.36) ) and the set of the probability distributions $\Psi^{Bel/B}$ compatible with the random set defined by the belief function (3.35) is not in any case guaranteed. The set $\Psi/B$ could be not coinciding but included by $\Psi^{Bel/B}$. The reasons will be justified by the discussion about the more general theory of imprecise probabilities in the next Section 3.5 and later in the Section 6.2 (see Example 6.6).

## 3.3 Imprecise Probabilities and Monotone Non-additive Measures

### *3.3.1 Introduction*

In the previous sections we have seen that a random set $\mathcal{F} = (A^i \subseteq S, m(A^i))$, $i = 1$ to $n$, determines in a unique way the set functions $Bel(T)$ and $Pla(T)$ bounding $P(T)$ for any $T \subseteq S$ and the set $\Psi$ of the compatible probability distributions. Following the ideas and terms given in (Walley 1991) the bounds $Bel(T)$ and $Pla(T)$ and set $\Psi$ are the *natural extension* of the information supplied by the random set.

To better appreciate the simplicity and related computational advantages of random sets it is important to observe that when the available information is directly given through upper ($\mu_{UPP}$) and lower ($\mu_{LOW}$) bounds of the probability of an algebra of events on the space $S$, it cannot in any case be described by an equivalent random set. For example a probabilistic assignment cannot be reconstructed by the lower bounds through the Möbius transform Eq. (3.8), and hence $\mu_{LOW}$ is not a Belief set function, as shown in the following Example 3.14 and Example 3.15.

**Example 3.13** (Dempster 1967). Let $S=\{s^1, s^2, s^3\}$; $\mu_{LOW}(\{s^1\}) = \mu_{LOW}(\{s^2\}) = \mu_{LOW}(\{s^3\}) = 0$; $\mu_{LOW}(\{s^1, s^2\}) = \mu_{LOW}(\{s^2, s^3\}) = \mu_{LOW}(\{s^1, s^3\}) = 0.5$.

Observe that Eq. (3.6) is not satisfied for $k = 3$ : $T^1 = \{s^1, s^2\}$, $T^2 = \{s^2, s^3\}$, $T^3 = \{s^1, s^3\}$. In fact:

$$\mu_{LOW}(T^1 \cup T^2 \cup T^3) = \mu_{LOW}(S) = 1 \text{ not} \geq$$

$$\mu_{LOW}(T^1) + \mu_{LOW}(T^2) + \mu_{LOW}(T^3) - \mu_{LOW}(T^1 \cap T^2 = \{s^2\}) - \mu_{LOW}(T^2 \cap T^3 = \{s^3\}) - \mu_{LOW}(T^3 \cap T^1 = \{s^1\}) + \mu_{LOW}(T^1 \cap T^2 \cap T^3 = \varnothing) =$$

$$= 0.5 + 0.5 + 0.5 - 0 - 0 - 0 + 0 = 1.5$$

Every bound is respected in a half space of the 3-dimensional space $(P(s^1), P(s^2), P(s^3))$ defined by the plane where it is respected as equality. The intersection of this plane with the equilateral triangle $(P(s^1) + P(s^2) + P(s^3) = 1)$ determines a line giving a bound for the set $\Psi$ of compatible probability distributions on the triangle.

In this example the set $\Psi$ is the equilateral triangle displayed in Figure 3.14, with three extreme points $P^1_{EXT} = (0.5, 0, 0.5)$, $P^2_{EXT} = (0, 0.5, 0.5)$, $P^3_{EXT} = (0.5, 0.5, 0)$.

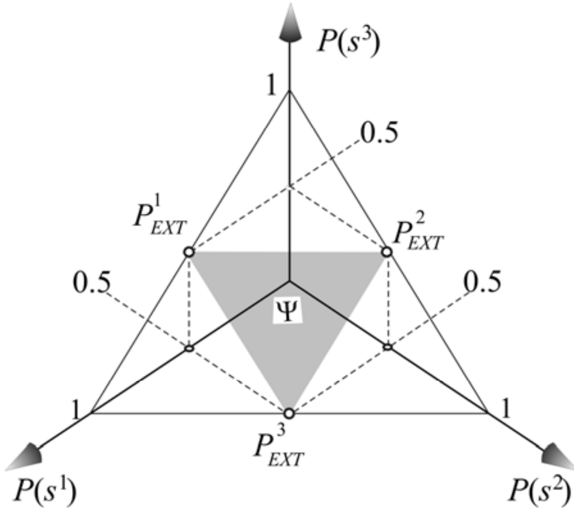



**Fig. 3.14** Set Ψ in Example 3.13

Möbius transform (3.8) of the set function $\mu_{LOW}(T)$ gives:

$m(\{s^1\}) = m(\{s^2\}) = m(\{s^3\}) = 0$; $m(\{s^1, s^2\}) = m(\{s^2, s^3\}) = m(\{s^1, s^3\}) = 0.5$

$m(S=\{s^1, s^2, s^3\}) = 0 + 0 + 0 - 0.5 - 0.5 - 0.5 + 1 = -0.5 < 0.$

The sum of the weights is 1 but one of the weights is negative: so the probabilistic model cannot be a random set.

In some other applications the available information could explicitly suggest a set Ψ of probability distributions for a variable, or implicitly define a set Ψ through a finite set of extreme distributions (see Example 3.15), or some more general restrictions to the values of the probabilities $P(s^j)$ of the singletons $s^j$, not corresponding to bounds of events on the space $S$, as displayed in the following Example 3.15).

This set Ψ univocally determines upper and lower bounds of the probability of any event (subset) $T$ on $S$:

$$\mu_{LOW}(T) = \min_{P\in\Psi} \sum_{s^j\in T} P\left(s^j\right)$$
$$\mu_{UPP}(T) = \max_{P\in\Psi} \sum_{s^j\in T} P\left(s^j\right) \qquad (3.37)$$

But the set function $\mu_{LOW}$ is not necessarily a Belief function.

Moreover it is possible to show that the set $\Psi^E$ of probability distributions compatible with (the natural extension of) the bounds:

$$\Psi^E = \left\{P : \mu_{LOW}(T) \le P(T) \le \mu_{UPP}(T), \forall T \subseteq S\right\} \qquad (3.38)$$

does not necessarily coincide with the set $\Psi$ (which the bounds $\mu_{UPP}$ and $\mu_{LOW}$ have been determined from), as displayed in Example 3.14).

**Example 3.14** (modified from (Walley 2000))**.** The simply-supported beam shown in Figure 3.15 is loaded by a concentrated force in the middle of the span. The intensity of the force can assume the values $\{s^1, s^2, s^3\}$, depending on the combination of the permanent load $W$ and two non-compatible accidental loads $L_1$ and $L_2$:

$s^1 = W = 100$ kN; $\qquad s^2 = \mathrm{W} + L_1 = 500$ kN; $\qquad s^3 = \mathrm{W} + L_2 .= 200$ kN.

The probability of observing these values is not exactly known, but described, according to expert opinions, by the following judgments:

$P(s^1) \leq 0.5, \qquad P(s^2) \leq P(s^1); \qquad P(s_3) \leq P(s^2).$

Taking into account that the sum of probabilities is equal to 1, the third judgement implies:

$$2\,P(s^2) \geq 1 - P(s^1)$$

The projection of set $\Psi$ on plane $(P(s^1), P(s^2))$ in Figure 3.15 shows a triangle with three extreme distributions $P^1{}_{EXT} = (1/2, 1/2, 0)$, $P^2{}_{EXT} = (1/3, 1/3, 1/3)$, $P^3{}_{EXT} = (0.5, ¼, ¼)$; Table 3.9 summarizes the lower/upper probabilities of the subsets and in the last column the Möbius transform of the lower bound. The negative values for the set $S = \{s^1, s^2, s^3\}$ demonstrates that the given information is not modelled by a random set.

Observe that Eq. (3.6) is again not satisfied for $k = 3$ : $T^1 = \{s^1, s^2\}$, $T^2 = \{s^2, s^3\}$, $T^3 = \{s^1, s^3\}$. In fact:

$\mu_{LOW}(T^1 \cup T^2 \cup T^3) = \mu_{LOW}(S) = 1$ not $\geq$

$\mu_{LOW}(T^1) + \mu_{LOW}(T^2) + \mu_{LOW}(T^3) - \mu_{LOW}(T^1 \cap T^2 = \{s^2\}) - \mu_{LOW}(T^2 \cap T^3 = \{s^3\}) - \mu_{LOW}(T^3 \cap T^1 = \{s^1\}) + \mu_{LOW}(T^1 \cap T^2 \cap T^3 = \varnothing) =$

$= 2/3 + 0.5 + 0.5 - 1/3 - 1/4 - 0 + 0 = 1.$

The set $\Psi^E$ of probability distributions compatible with the lower/upper probabilities is displayed in Figure 3.16, and clearly does not coincide with $\Psi$; there are five extreme distributions: the 3 extremes of the set $\Psi$ and two new extremes. Hence $\Psi \subset \Psi^E$. The set $\Psi^E$ is defined by the probability bounds on each singleton in Table 3.9, rows $i = 2, 3, 4$ (or by the bounds of the complementary sets, rows $i = 5, 6, 7$). On the projected plane, these bounds generate pairs of parallel lines with normal vectors (1,0), (0,1) and (1,1).

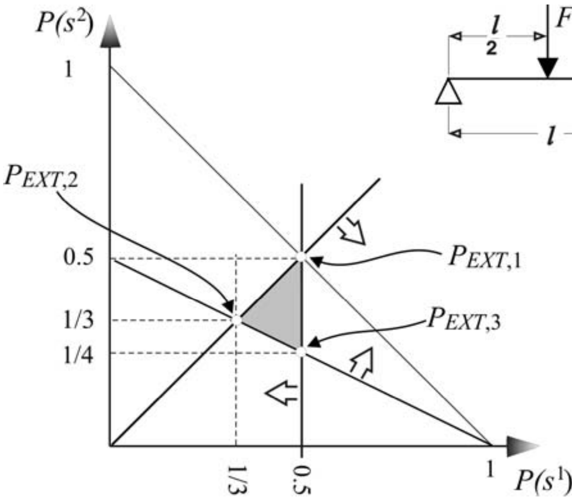
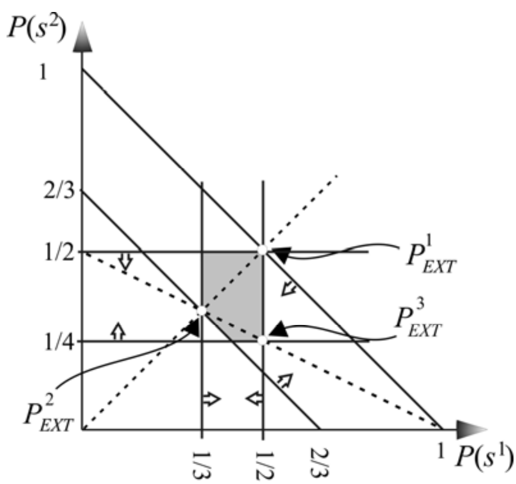



**Fig. 3.15** Set Ψ,and extreme joint probability distributions in Example 3.14

**Table 3.9** Set functions in Example 3.14

| $i$ | $\chi(s^1)$ | $\chi(s^2)$ | $\chi(s^3)$ | $\mu_{LOW}(A^i)$ | $\mu_{UPP}(A^i)$ | $m(A^i)$ |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 1/3 | 1/2 | 1/3 |
| 3 | 0 | 1 | 0 | 1/4 | 1/2 | 1/4 |
| 4 | 0 | 0 | 1 | 0 | 1/3 | 0 |
| 5 | 1 | 1 | 0 | 1-1/3=2/3 | 1-0=1 | 1/12 |
| 6 | 1 | 0 | 1 | 1-1/2=1/2 | 1-1/4=3/4 | 2/12 |
| 7 | 0 | 1 | 1 | 1-1/2=1/2 | 1-1/3=2/3 | 3/12 |
| 8 | 1 | 1 | 1 | 1.0 | 1.0 | -1/12 |

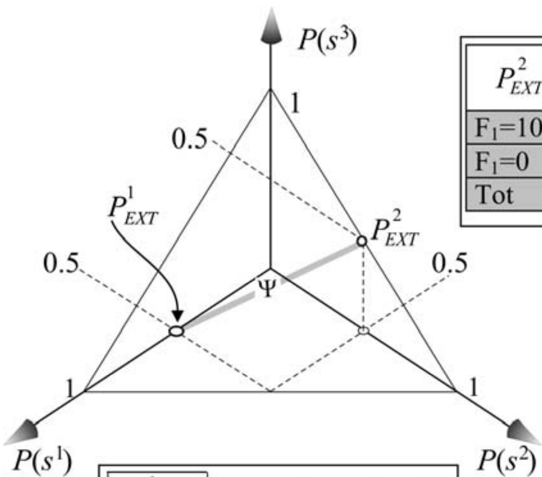
**Fig. 3.16** Set $\Psi^E$ in Example 3.14



**Example 3.15** (modified from (Walley 2000))**.** The simply-supported beam shown in Figure 3.17 is loaded by two concentrated accidental forces $F_1$ and $F_2$, both of precisely-known intensity $F$. It is also known that loading and unloading are equally likely ( $P(F_i = F) = P(F_i = 0) = 0.5$) for each of the loads ($i$ =1 and 2), but we have no information about their dependence or independence. Let

$$S = \{s^1 = (F_1 = F) \cap (F_2 = F),\quad s^2 = (F_1 = F) \cap (F_2 = 0),$$
$$s^3 = (F_1 = 0) \cap (F_2 = \mathrm{F}),\quad s^4 = (F_1 = 0) \cap (F_2 = 0)\}$$

be the space of the possible joint outcomes. Of course, for example, $P(F_1{=}F){=}0.5$ is the marginal value equal to $P(\{s^1\}){+}\ P(\{s^2\})$. The available information is given by precise marginal distributions, but set $\Psi$ of the compatible joint distributions $P(s^j)$ is not precisely determined. If independence is assumed, we obtain $P(s^j){\equiv}0.25$, but cannot exclude two opposite extreme distributions $P^1_{EXT} = (1, 0, 0, 1)$ and $P^2_{EXT} = (0, 1, 1, 0)$. In the first case some hidden mechanism constrains the loads to the same value, in the second the loading of one force implies unloading of the other. The two extreme joint distributions and projection of set $\Psi$ (a one-dimensional interval) on the sub-space ($P(s^1)$, $P(s^2)$, $P(s^3)$) are shown in Figure 3.17.




| $P^2_{EXT}$ | $F_2$=100 | $F_2$=0 | Tot |
|---|---|---|---|
| $F_1$=100 | 0 | 0.5 | 0.5 |
| $F_1$=0 | 0.5 | 0 | 0.5 |
| Tot | 0.5 | 0.5 | 1 |

| $P^1_{EXT}$ | $F_2$=100 | $F_2$=0 | Tot |
|---|---|---|---|
| $F_1$=100 | 0.5 | 0 | 0.5 |
| $F_1$=0 | 0 | 0.5 | 0.5 |
| Tot | 0.5 | 0.5 | 1 |

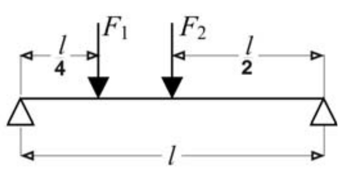



**Fig. 3.17** Set $\Psi$, and extreme joint probability distributions in Example 3.15

The lower/upper bounds of every subset or prevision can be evaluated by Eq. (3.37), simply considering minimum and maximum on the two extremes.

Observe that Eq. (3.6) is not satisfied for $k = 2 : T^1 = \{s^1, s^2\}, T^2 = \{s^1, s^3\}$. In fact:

$$\mu_{LOW}(T^1 \cup T^2) = \mu_{LOW}(\{s^1, s^2, s^3\}) = 0.5 \ \text{ not} \geq$$
$$\mu_{LOW}(T^1) + \mu_{LOW}(T^2) - \mu_{LOW}(T^1 \cap T^2 = \{s^2\}) = 0.5 + 0.5 - 0 = 1.$$

Möbius transform (3.8) of the set function $\mu_{LOW}(T)$ gives negative values for all subsets with cardinality equal to 3. For example:

$$m(\{s^1, s^2, s^3\} = \mu_{LOW}(\{s^1, s^2, s^3\}) - \mu_{LOW}(\{s^1, s^2\}) - \mu_{LOW}(\{s^1, s^3\}) -$$
$$\mu_{LOW}(\{s^2, s^3\}) + \mu_{LOW}(\{s^1\}) + \mu_{LOW}(\{s^2\}) + \mu_{LOW}(\{s^3\}) =$$

$$= 0.5 - 0.5 - 0.5 - 0 + 0 + 0 + 0 = -0.5$$

Summation of all $m(T)$ again gives 1.

These examples show that random sets are particular cases of a more general theoretical model of *imprecise probabilities*, covering a more extensive list of modes to describe the uncertainty of variables or parameters in engineering problems. Although this conclusion apparently reduces the attraction of random sets in applications, it does highlight the profit when the problem of interest can be modeled, at least with some approximation, by random sets, as will be seen in the next chapters.

Some aspects of this extended theory will be briefly summarized in the following sections.

### 3.3.2 Coherent Upper and Lower Previsions

The foundation of a theory of imprecise probabilities is mainly due to the pioneering work of P. Walley in the 1980s/90s on a new theory of probabilistic reasoning, statistical inference and decision, in conditions of uncertainty, partial information or ignorance. In recent years growing interest in the subject has been demonstrated by applications in many scientific fields, the foundation of an international "Society for Imprecise Probability: Theories and Applications" (SIPTA) and the related biannual ISIPTA Conferences.

In Walley's theory (Walley 1991), and for a concise introduction (Walley 2000), the idea of upper/lower probabilities $\mu_{LOW}(T)$, $\mu_{UPP}(T)$) for a family $\mathcal{S}$ of events $T^i$ on the space $S$ is enlarged to the more general concept of upper/lower *previsions* for a family $\mathcal{K}$ of bounded and point-valued functions $f_i$: $S \rightarrow Y = \Re$. For a specific precise probability distribution $P(s^j)$ the prevision is equivalent to the linear expectation:

$$E[f_i] = \sum_{s^j \in S} f_i\left(s^j\right) P\left(s^j\right) \tag{3.39}$$

Taking into account that the probability of an event $T^i \subset S$ is equal to the expectation of its indicator function $I_{T^i}$:

$$P\left(T^i\right) = \sum_{s^j \in T^i} P\left(s^j\right) = \sum_{s^j \in S} I_{T^i} \cdot P\left(s^j\right) \tag{3.40}$$

upper/lower previsions generalize and hold as particular case upper/lower probabilities.

Assigning an upper bound to the prevision of a function $f_i$ is equivalent to assigning the opposite value as lower bound of the function $-f_i$: this property extends the duality relation for the upper/lower probabilities of the complementary events $T^i$ and $T^{i,C}$ to the previsions. In fact: $-I_{Ti} = I_{Ti,C} - 1$.

**Example 3.16.** Let us again consider the simply-supported beam loaded by one concentrated live load of uncertain intensity $F$ discussed in Example 3.14. Taking into account that the probability of events coincides with the expectation of the indicator function, the original information about intensity $F$ can be equivalently given as follows:

$E_{UPP}[f_1 = I_{\{s^1\}} - 0.5 = (1,0,0) - 0.5 = (0.5, -0.5, -0.5)] = 0$ or :

$E_{LOW}[-f_1 = (-0.5, 0.5, 0.5)] = 0$

$E_{LOW}[f_2 = I_{\{s^1\}} - I_{\{s^2\}} = (1, -1, 0)] = 0$

$E_{LOW}[f_3 = I_{\{s^2\}} - I_{\{s^3\}} = (0, 1, -1)] = 0$

Considering that $E[f] = f(s^1)\,P(s^1) + f(s^2)\,P(s^2) + f(s^3)\,P(s^3)$, clearly the 3 equations for the lower previsions are exactly equivalent to the constraints generating the set $\Psi$ displayed in Figure 3.15. The corresponding lower probabilities however generate the set $\Psi^E \supset \Psi$ in Figure 3.16. It is easy to check that the new extreme distributions appearing in $\Psi^E$ satisfy the lower bounds of the probabilities, but not all given lower previsions.

For example in $P = (P(s^1) = 1/3, P(s^2) = 1/2, P(s^3) = 1 - 1/3 - 1/2 = 1/6)$:

$E[f_2] = 1 \text{ x } 1/3 - 1 \text{ x } 1/2 + 0 \text{ x } 1/6 = -1/6 < 0.$

Let us now focus on the information about the space of events on $S$ given by upper and/or lower previsions, $E_{LOW}[f_i]$, $E_{UPP}[f_i]$, of a family $\mathcal{K}$ of bounded and point-valued functions $f_i$. This is accomplished by the set, $\Psi^E$, of compatible probability distributions $P(s^j)$:

$$\Psi^E = \left\{ P \mid \forall f_i \in \mathcal{K} : E_{LOW}\left[f_i\right] \le E_P\left[f_i\right] \le E_{LOW}\left[f_i\right] \right\} \tag{3.41}$$

We are interested in checking two basic conditions of the suggested bounds.

A preliminary strong condition (§ 3.3.2.1) requires that set $\Psi^E$ should be non-empty. If set $\Psi^E$ is empty, it means there is something basically irrational in the suggested bounds.

A weaker but reasonable condition (§ 3.3.2.2) requires that the given bounds should be coherent with the bounds that can be derived from $\Psi^E$. If the given bounds are not envelopes of the set of values derived by means of each probability distribution in $\Psi^E$, it means they can be restricted without changing the probabilistic content (set $\Psi^E$) of the original information.

#### 3.3.2.1 Non-empty $\Psi^E$

Let us consider first that only lower bounds $E_{LOW}[f_i]$, $i = 1$ to $n$ are given.

Set $\Psi^E$ is non-empty if and only if the following formulae for the unknown compatible distributions $P(s_j)$ admit some solutions:

$$\begin{aligned} &\sum_j f_i\left(s^j\right)P\left(s^j\right) \geq E_{LOW}\left[f_i\right],\ i=1,2,\ ....n && \text{(a)} \\ &\sum_j P\left(s^j\right) = 1 && \text{(b)} \end{aligned} \tag{3.42}$$

Multiply the right sides of constraints (3.42)a by the left side of (3.42)b (identically equal to 1) and add the $n$ relations, each considered $k_i$ times:

$$\sum_j \left(\sum_i k_i\left(f_i\left(s^j\right) - E_{LOW}\left[f_i\right]\right)\right)P\left(s^j\right) \geq 0 \tag{3.43}$$

Taking into account that every $P(s^j)$ cannot be negative, set $\Psi$ is non-empty only if (and only if) at least one (of course the maximum, or the supremum if $f_i$ takes infinitely many values) of the coefficients multiplying $P(s^j)$ is not negative, whenever $n \geq 1$ and $k_i$ are positive integers[N 3-3]:

$$\sup_j \sum_{i=1}^{n} k_i\left(f_i\left(s^j\right) - E_{LOW}\left[f_i\right]\right) \geq 0 \tag{3.44}$$

In the same way we can consider that only upper bounds $\mu_{UPP}(T^i)$, or more generally $E_{UPP}[f_i]$, $i = 1$ to $n$ could be given, obtaining:

$$\sup_j \sum_{i=1}^{n} k_i\left(-f_i\left(s^j\right) + E_{UPP}\left[f_i\right]\right) \geq 0 \tag{3.45}$$

Assigning an upper bound to the prevision of a function $f_i$ is equivalent to assigning the opposite value as lower bound of the function $-f_i$ : therefore upper and lower bounds can be considered together through Eq. (3.44).

Many interesting consequences can be derived for particular cases.

– Eq. (3.44) and (3.45) give, for $n = 1$:

$$E_{LOW}[f] \leq \sup_j f\left(s^j\right); \quad E_{UPP}[f] \geq \inf_j f\left(s^j\right) \tag{3.46}$$

For the probabilities of a set $T$ Eq. (3.46) simply requires that the lower bound is ≤ 1 and the upper ≥ 0.

– When lower and upper bounds are given for the prevision of the same function $f$ (or for the probability of a set $T$), Eq. (3.44) and (3.45) imply that:

$$f\left(s^j\right) - E_{LOW}[f] - f\left(s^j\right) + E_{UPP}[f] \equiv E_{UPP}[f] - E_{LOW}[f] \geq 0 \tag{3.47}$$

i.e. the upper bound must be greater than the lower bound.

– When lower (upper) bounds are given for a family of disjoint sets $T^i$ of a partition of space $S$ ($T^1 \cup T^2 \ldots \cup T^n = S$), in every point $s^j$ the sum of the indicator functions $I_{T^i}$ (respectively $-(-I_{T^i})$) is equal to 1 (each point belongs to only one set of the family of disjoint sets); hence Eq. (3.44) and (3.45) imply that:

$$\begin{aligned} &1 - \sum_i \mu_{LOW}\left[T^i\right] \geq 0 \quad \Rightarrow \sum_i \mu_{LOW}\left[T^i\right] \leq 1 \\ &-1 + \sum_i \mu_{UPP}\left[T^i\right] \geq 0 \quad \Rightarrow \sum_i \mu_{UPP}\left[T^i\right] \geq 1 \end{aligned} \tag{3.48}$$

Formulas (3.4) are a particular case of (3.48), when considering *Bel* and *Pla* set functions of complementary (of course disjoint) sets $T$, $T^C$.

– When lower (upper) bounds are given for all subsets of space $S$ ( $\forall T^i \in \mathcal{P}(S)$ ), in each point $s^j$ the sum of the indicator functions $I_{T^i}$ (respectively $-(-I_{T^i})$) is identically equal to $2^{|S|-1}$ (each point can belong to every subset on the sub-space given by the complementary sets, with cardinality |S| – 1); therefore:

$$\begin{aligned} &2^{|S|-1} - \sum_i \mu_{LOW}\left[T^i\right] \geq 0 \quad \Rightarrow \sum_i \mu_{LOW}\left[T^i\right] \leq 2^{|S|-1} \\ &-2^{|S|-1} + \sum_i \mu_{UPP}\left[T^i\right] \geq 0 \quad \Rightarrow \sum_i \mu_{UPP}\left[T^i\right] \geq 2^{|S|-1} \end{aligned} \tag{3.49}$$

**Example 3.17.** Let us again consider the simply-supported beam loaded by one concentrated live load of uncertain intensity $F$ discussed in Example 3.14 and Example 3.16, but the statistical information about loads is now described, according to expert opinions, by the following judgments:

$$P(s^1) \leq 0.5, \qquad P(s^2) \leq P(s^1)/10; \qquad P(s_3) \leq P(s^2).$$

or equivalently:

$$E\,[f_1 = (-1, 0, 0)] \geq -0.5\ ; \ E\,[f_2 = (0.1, -1, 0)] \geq 0; E\,[f_3 = (0, 1, -1)] \geq 0$$

Assuming for example: $k_1 = 1$, $k_2 = 3$, $k_3 = 1$ Eq. (3.44) gives:

$$\max\,(\ 1\text{x}(-1 + 0.5) + 3\text{x}0.1 + 0,\ 1\text{x}(0.5) + 3\text{x}(-1)+1\text{x}1,\ 1\text{x}0.5 + 0 + 1\text{x}(-1))$$
$$= \max\,(-0.2, -1.5, -0.5) = -0.2 < 0.$$

Clearly the constraints imply that $P(s^1) + P(s^2) + P(s_3) \leq 0.5 + 0.05 + 0.05 = 0.6$: hence the set $\Psi^E$ is empty. See also Figure 3.15, reducing from 1 to 1/10 the inclination of the second constraint (the line through the origin of the Cartesian axes).

### 3.3.2.2 Coherence

When Eq. (3.44) is satisfied, the natural extension of the bounds is a non-empty set $\Psi^E$; from it upper and lower bounds of the probability of any set $T$ or more generally of the prevision of any function $f$ can be evaluated:

$$\begin{aligned} E_{LOW,c}[f] &= \min_{P\in\Psi^E} E_P[f] \\ E_{UPP,c}[f] &= \max_{P\in\Psi^E} E_P[f] \end{aligned} \tag{3.50}$$

It is easy to check that, when Eq. (3.50) is applied to evaluate again the original bounds of the sets or the previsions $E_{LOW}[f_i]$, $E_{UPP}[f_i]$ of functions used to derive $\Psi$, the resulting effective bounds $E_{LOW,c}[f_i]$, $E_{UPP,c}[f_i]$ could be generally different: of course the updated bounds do not contradict the original ones, but could be more restrictive. In fact, taking into account that, according to Eq. (3.50), the bounds are lower or upper envelopes of probability measures, they must satisfy further conditions of *coherence* (Walley 1991) that can be derived from the properties of probability measures[N 3-4].

**Example 3.18.** Let $S=\{s^1, s^2, s^3\}$, $\mu_{UPP}(T^1 = \{s^1\}) = 0.7$, $\mu_{UPP}(T^2 = \{s^1, s^2\}) = 0.3$. The projection of set $\Psi^E$ in the plane $(P(s^1), P(s^2))$ is shown in Figure 3.18a).

The effective bound $\mu_{UPP}(T^1 = \{s^1\}) = 0.3$, implied by $\mu_{UPP}(\{s^1, s^2\}) = 0.3$ is lower than the assigned value of 0.7. In fact the assigned value is not coherent because monotonicity was not respected with respect to the inclusion ($T^1 \subset T^2$).

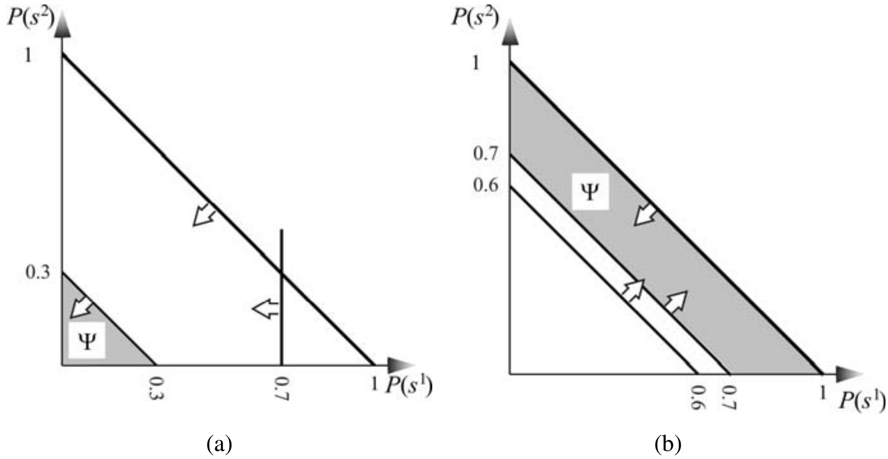



**Fig. 3.18** Set $\Psi^E$ as a natural extension of incoherent bounds given in Example 3.18 (a) and Example 3.19 (b). The arrows display the direction of given constraints

**Example 3.19.** Let $S=\{s^1, s^2, s^3\}$, $\mu_{LOW}(T^1 = \{s^1, s^2\}) = 0.6$, $\mu_{UPP}(T^2 = \{s^3\}) = 0.3$

The projection of set $\Psi$ in the plane $(P(s^1), P(s^2))$ is shown in Figure 3.18b.

The effective bounds $\mu_{LOW}(T^1 = \{s^1, s^2\}) = 0.7$ are higher than the assigned value of 0.6. In fact the assigned value is not coherent because duality of complementary sets ($T^1 = T^{2,C}$) is not respected (see formulae (3.7) for random sets).

Coherent upper/lower bounds of probabilities must satisfy the following (non-exhaustive[N 3-5]) list of necessary conditions:

- $0 \le \mu_{LOW}(T) \le \mu_{UPP}(T) \le 1$
- Monotonicity with respect to inclusion: $T^1 \subseteq T^2 \Rightarrow \mu_{LOW}(T^1) \le \mu_{LOW}(T^2)$
- Super-additivity [N 3-6] of $\mu_{LOW}$ for disjoint sets ($T^1 \cap T^2 = \varnothing$) : $\mu_{LOW}(T^1 \cup T^2) \ge \mu_{LOW}(T^1) + \mu_{LOW}(T^2)$
- Sub-additivity [N 3-7] of $\mu_{UPP}$ for any pair of sets $T^1, T^2$ : $\mu_{UPP}(T^1 \cup T^2) \le \mu_{UPP}(T^1) + \mu_{UPP}(T^2)$
- Duality of $\mu_{LOW}$, $\mu_{UPP}$, for complementary sets ($T \cup T^C = S$): $\mu_{LOW}(T) + \mu_{UPP}(T^C) = 1$

Information given by a random set (through the probabilistic assignment $m(A_i)$ or one of the dual, monotonic with respect to inclusion, set functions *Bel* or *Pla*) is coherent (because $\Psi^E = \Psi$, as shown in § 3.2.3).

However coherent, upper/lower set functions cannot necessarily be assumed as *Bel*/*Pla* set functions of an associated random set. In fact formulas (3.5) and (3.6) imply super-additivity of *Bel* and sub-additivity of *Pla,* but are stronger relations not required by coherence.

The crucial point is the observation that, when upper/lower probabilities $\mu_{LOW}(T)$, $\mu_{UPP}(T)$) for a family of events $T^i$ in space $S$ are widened to the more general concept of upper/lower previsions for a family of bounded and point-valued functions $f_i$: $S \rightarrow Y$, there is a one-to-one correspondence between coherent lower (or upper) previsions and set $\Psi = \Psi^E$ (or the extreme distributions $EXT_\Psi = EXT \cap \Psi$ of set $\Psi$) of compatible probability distributions.

A simple geometrical interpretation of this property can be obtained by observing that a lower bound, $\mu_{LOW}(T)$, for the probability of an event $T$ defines a half space of compatible probability distributions (points $P$ of the $|S|$-dimensional space $(P(s^1), P(s^2), \ldots P(s^{|S|}))$, bounded by the hyper-plane:

$$
\begin{aligned}
&a_1 P\left(s^1\right) + \ldots . a_j P\left(s^j\right) + \ldots . a_{|S|} P\left(s^{|S|}\right) - \mu_{LOW}\left(T\right) = 0 \\
&a_j = \begin{cases} 1 & \text{if } s^j \in T \\ 0 & \text{if } s^j \notin T \end{cases}
\end{aligned}
\tag{3.51}
$$

The normal vector to this hyper-plane is $\mathbf{a}^T = (a_1, \ldots, a_{|S|})^T$ and its 0-1 components are the values of the characteristic function for $T$. As a consequence, this set of hyper-planes (with normal components either 0 or the same value) is not rich enough to describe the boundary of a general polyhedron, i.e. the convex hull of a general set of extreme points. For instance, set $\Psi$ in Figure 3.15 is not bounded by lines whose normals are $(1,0)^T$, $(0,1)^T$, and $(1,1)^T$. Hence many (infinite) polyhedrons, with different faces and extreme points, could be compatible with probability bounds assigned on a family of events and the correspondence between probability bounds and polyhedrons is not one-to-one. The natural extension of the bounds, $\Psi^E$, (Eq. (3.41)) is uniquely determined because it is defined as the largest of this set of polyhedrons. Set $\Psi^E$ in Figure 3.16 is indeed bounded by lines whose normals are $(1,0)^T$, $(0,1)^T$, and $(1,1)^T$, and is the largest set that yields the probability bounds listed in Table 3.9. The extreme distributions for $\Psi^E$ will be derived in Example 3.21.

When $|S| = 3$, a probability bound on an event $T$, $P(T) \geq \mu_{LOW}\left(T\right)$, is always equivalent to a probability bound on a singleton. If $T$ is composed of a singleton, there is nothing to show. If $|T| = 2$, then $|T^C| = 1$ and $P(T^C) \geq 1 - \mu_{LOW}\left(T\right)$. As a consequence, when $|S| = 3$, one can assume that all bounding planes have normals (1,0,0), (0,1,0) or (0,0,1), i.e. they are parallel to coordinate planes. This property also holds for $|S| = 2$, but it is not true for any other dimension because in general $|T^C| \neq 1$.

Now, suppose that the normal components $a_j$ in Eq. (3.51a) are not just 0-1, but may be selected from the set of the real numbers, $\mathbb{R}$, and that each non-zero component may be different from one another. In this case, the family of indicator functions must be enlarged to all bounded real functions *f*, probability of events become expectations (Eq. (3.39)), probability bounds become expectation bounds, $E_{LOW}\left[f\right]$, and Eq. (3.51a) becomes:

$$E[f]-E_{LOW}[f]=0 \Rightarrow \sum_{j=1}^{|S|} f\left(s^j\right)P\left(s^j\right)-E_{LOW}[f]=0 \tag{3.52}$$

For any given polyhedron face *F*, one can always find at least one function *f* such that plane (3.52) contains *F*, i.e. such that the *j*-th normal component is $f(s^j)$ and such that its known term is $E_{LOW}\left[f\right]$. On the other hand, given a plane (3.52), there is a polyhedron whose face is contained in plane (3.52). Therefore, there is a one-to-one correspondence between coherent expectation bounds of generic functions (upper/lower previsions) and polyhedrons. A rigorous proof is contained in Walley (1991, Theorem 3.6.1, page 145).

### *3.3.3 Choquet and Alternating Choquet Capacities of Order k*

An important criterion for classifying monotonic (with respect to inclusion) measures was introduced in the 1950s, by Choquet in his theory of *capacities* (Choquet 1954).

Given a finite space *S,* a *Capacity of order k*, integer > 1, is a set function

$$C_{[k]} : \mathcal{P}(S) \rightarrow [0, 1]$$

that satisfies requirements: $C_{[k]}(\emptyset) = 0$, $C_{[k]}(S)= 1$, and moreover formulas similar to the first of Eq. (3.6) (first of Eq. (3.5) when $k = 2$), when $C_{[k]}$ is substituted for *Bel*. Capacity of order *k* satisfies formulas similar to Eq. (3.6)a for any other *k’* < *k*, but not generally for any other *k’* > *k*. Hence a capacity of order 2 is the most general.

*Bel* set function of a random set and *P* (probability measure on the algebra of the power set of space *S*) are Choquet capacities for all $k \geq 2$, and are therefore termed *Choquet capacities of infinite order* (although *k* cannot surpass $2^{|S|}$). Probability measures *P* satisfy Eqs. (3.6)a and b as equalities, due to additive axiom, while generally a weaker property of super-additivity is implied for all other Choquet capacities. For every *k*, Choquet capacities are monotone with respect to inclusion:

$$T^1 \subseteq T^2 \Rightarrow C_{[k]}(T^1) \leq C_{[k]}(T^2)$$

Monotone set functions not satisfying Eq. (3.5)a are termed *Choquet capacities of order 1*.

The duality property for complementary sets allows to define, for any Choquet capacity of order $k$, a dual, monotone and sub-additive set function $A_{[k]}$ termed *Alternating Choquet capacity of order k* satisfying relation (3.6)b ((3.5)b when $k = 2$), when $A_{[k]}$ is substituted for *Pla*.

*Pla* set function of a random set and $P$ are *Alternating Choquet capacities of infinite order* (for $P$ the dual functions coincide).

Choquet and dual Alternating Choquet capacities of order $k > 1$ are coherent lower and upper probabilities respectively (Walley, 1991). However coherent super-additive lower probabilities are not necessarily Choquet capacities of order $k > 1$, as will be shown in Example 3.20.

There is a strong relation between the order $k$ and the properties of the set function $m$ that can be obtained from $C_{[k]}$ or $A_{[k]}$ through the Möbius transform (3.8). See for example (Chateauneuf and Jaffray 1989; Klir 2005). The more interesting properties are the following:

1. a set function μ is monotone ($k \geq 1$) if and only if:

$$m(\varnothing) = 0; \qquad \sum_{T \in \mathcal{P}(S)} {}^{\mu}m(T) = 1;$$
$$\forall T \in \mathcal{P}(S): \sum_{A \subseteq T} {}^{\mu}m(A) \geq 0 \tag{3.53}$$

   and, therefore, $\forall\, j$: ${}^{\mu}m(\{s_j\}) \geq 0$.
2. If $\mu(T) = C_{[k]}(T)$ and $|T| \leq k$ then ${}^{\mu}m(T) \geq 0$
3. $\mu(T) = C_{[\infty]}(T)$ if and only if: $\forall\; T \in \mathcal{P}(S)$ : ${}^{\mu}m(T) \geq 0$ and therefore it is the *Bel* set function of a random set

Properties 1 and 2 confirm the results reported in Example 3.13, Example 3.14 (in both cases $\mu_{LOW}$ is a Choquet capacity of the order $k = 2$; ${}^{\mu}m(T) < 0$ for $|T| = 3$; ${}^{\mu}m(T) \geq 0$ for $|T| = 1$ or 2) and Example 3.15 (in this case $\mu_{LOW}$ is simple monotone ($k = 1$): again ${}^{\mu}m(T) < 0$ for all subsets with $|T| = 3$). Summation of all $m(T)$ gives 1 in all cases.

The procedure described in § 3.2.3.4 for random sets can be extended to evaluate the set of extreme distributions *EXT,* when a monotone measure (a Choquet capacity of the order $k \geq 1$) is assumed as lower bound of imprecise probabilities: having assigned or calculated $\mu_{LOW}$ (or $\mu_{UPP}$) of any $T \in \mathcal{P}(S)$, and chosen one of the $|S|!$ permutations $\pi(j)$ of the indexes, the corresponding extreme distribution is given by:

$$P_{EXT}^{\pi(j)=1} = \mu_{LOW}\left(\left\{s^{\pi(\mathrm{j})=1}\right\}\right)$$
$$P_{EXT}^{\pi(j)=r>1} = \mu_{LOW}\left(\left\{s^{\pi(\mathrm{j})=1},\ldots s^{r}\right\}\right) - \mu_{LOW}\left(\left\{s^{\pi(\mathrm{j})=1},\ldots s^{r-1}\right\}\right) \tag{3.54}$$

Recall that, for *Bel* set functions of a random set, the convex hull of *EXT* ($\Psi^{EXT}$) coincides with set $\Psi$ of the probability distributions compatible with the probabilistic assignment (Eq. (3.11)), and also with the set $\Psi^E$ of probability distributions "dominating" $Bel(T)$ for every $T \subseteq S$, (Eq. (3.10)). However, when $\mu_{LOW}$ is used, $\Psi^{EXT}$ is the convex hull of *EXT* generated by Eq. (3.54) and $\Psi^E$ is given by Eq. (3.38), the three sets could be more generally different. Precisely:

- for coherent measures with $k = 1$ the procedure (3.54) could generate probability distributions in *EXT* not satisfying the assigned $\mu_{LOW}$ or $\mu_{UPP}$ bounds; $\Psi^{EXT}$ is larger than the set of probability distributions satisfying the bounds and hence $\Psi^E$ could be strongly included in $\Psi^{EXT}$ (see Example 3.20 );
- for Choquet capacities ($k > 1$) all probability distributions in *EXT* (and in $\Psi^{EXT}$) satisfy the assigned bounds (and thus $\Psi^{EXT} = \Psi^E$); *EXT* coincides with the set of the extreme points (or the *profile*) of the closed convex set $\Psi^E$ (Klir, 2005, pp. 118-9); but $\Psi^E$ could be larger than the original set $\Psi$ describing the information used to evaluate the bounds (see Example 3.21).

In in the next Chapter 4.2 ( page 114 ) an alternative procedure will be given to evaluate in any case (hence for $k = 1$ also) the set of the extreme distributions.

**Example 3.20.** Let us consider set $\Psi$ in Example 3.15 from which lower probabilities $\mu_{LOW}$ were calculated. These lower probabilities are coherent lower probabilities because calculated using Eq. (3.50). However Eq. (3.5) (or Eq. (3.6) for $k = 2$) is not respected and thus they are simple monotone measures ($k = 1$).

The (at most 4!=24) members of *EXT* can be found through Eq. (3.54). For example the identity $\pi(j) = j$ gives:

$\mu_{LOW}(\{s^1\}) = 0$; $\mu_{LOW}(\{s^1, s^2\}) = 0.5$; $\mu_{LOW}(\{s^1, s^2, s^3\}) = 0.5$; $\mu_{LOW}(S) = 1$

$\mu_{UPP}(\{s^1\}) = 0.5$; $\mu_{UPP}(\{s^1, s^2\}) = 0.5$; $\mu_{UPP}(\{s^1, s^2, s^3\}) = 1$; $\mu_{UPP}(S) = 1$

and therefore the extremes $P^3_{EXT} = (0, 0.5, 0, 0.5)$ and $P^4_{EXT} = (0.5, 0, 0.5, 0)$ are identified.

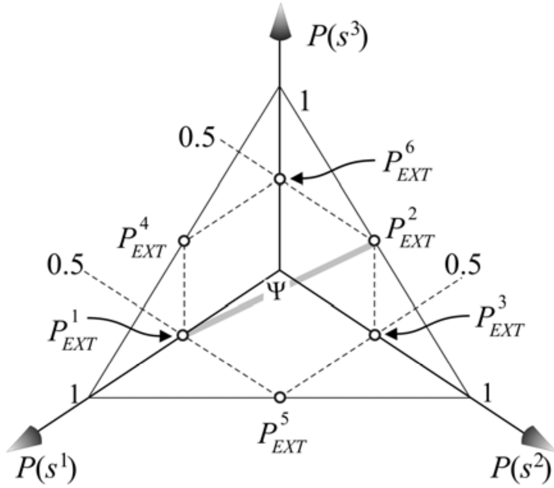



**Fig. 3.19** Set Ψ in Example 3.15 and extremes from the lower probabilities identified in Example 3.20

In the same way, the following 6 extreme distributions (each one with multiplicity equal to 2) displayed in Figure 3.19 can be calculated:

$P^1{}_{EXT} = (0.5, 0, 0, 0.5)$ ; $P^2{}_{EXT} = (0, 0.5, 0.5, 0)$ ; $P^3{}_{EXT} = (0, 0.5, 0, 0.5)$

$P^4{}_{EXT} = (0.5, 0, 0.5, 0)$ ; $P^5{}_{EXT} = (0.5, 0.5, 0, 0)$ ; $P^6{}_{EXT} = (0, 0, 0.5, 0.5)$

Clearly the four new points that do not coincide with extremes of Ψ ($P_{EXT,1}$ and $P_{EXT,2}$ ) satisfy the bounds of the singletons $\{s^j\}$ but not the marginals of the joint distributions, which are equal to the precise value 0.5. The precise marginals derive from the hypothesis in Example 3.15 that loading and unloading are equally likely ($P(F_i = \text{F}) = P(F_i = 0) = 0.5$) for $i = 1$ and 2). Hence: $\Psi^{EXT} \supset \Psi^E = \Psi$.

**Example 3.21.** Let us again consider the problem in Example 3.14. Table 3.10 demonstrates that the procedure (3.54) through the 3!=6 permutations of the indexes identifies that $P^1{}_{EXT}$ has multiplicity 2 and, above all, that two new extreme distributions $\text{P}_{EXT,4}$=(1/3, 1/2, 1/6), $\text{P}_{EXT,5}$=(5/12, 1/4, 1/3) are compatible with the bounds $\mu_{LOW}$ and $\mu_{UPP}$ , as previously discussed in Example 3.14 and displayed in Figure 3.16: the natural extension of the bounds is a set $\Psi^E$ larger than Ψ. In fact every set Ψ′ included by $\Psi^E$ and including Ψ is compatible with the bounds.

**Table 3.10** Extreme distributions from µLOW (Ai) in Example 3.14

| π(1) | π(2) | π(3) | $P(s^1)$ | $P(s^2)$ | $P(s^3)$ | $P^r{}_{EXT}$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 1/3 | 2/3-1/3=1/3 | 1-2/3=1/3 | $P^2$ |
| 3 | 2 | 1 | 1-1/2=1/2 | 1/2- 0 = 1/2 | 0 | $P^1$ |
| 1 | 3 | 2 | 1/3 | 1-1/2 =1/2 | 1/2-1/3=1/6 | $P^4$ (new) |
| 2 | 3 | 1 | 1-1/2=1/2 | 1/4 | 1/2-1/4=1/4 | $P^3$ |
| 2 | 1 | 3 | 2/3-1/4=5/12 | 1/4 | 1-2/3=1/3 | $P^5$ (new) |
| 3 | 1 | 2 | 1/2-0=1/2 | 1-1/2=1/2 | 0 | $P^1$ |

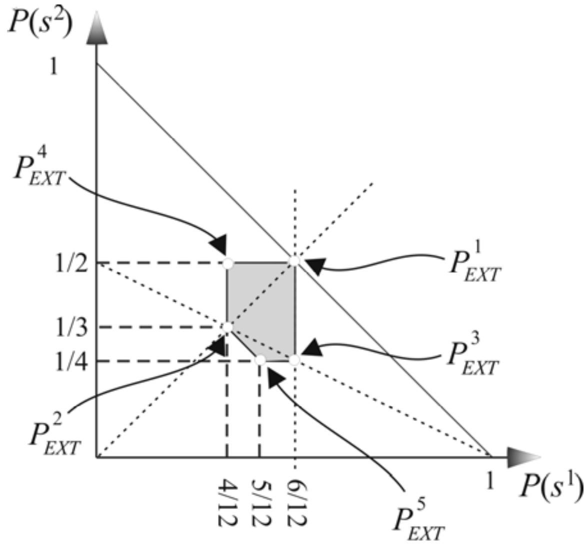



**Fig. 3.20** Set $\Psi^{EXT} = \Psi^E$ and extreme joint probability distributions in Example 3.14. Set $\Psi$ is the triangle with extremes $P^1{}_{EXT}$, $P^2{}_{EXT}$, $P^3{}_{EXT}$

### *3.3.4 Expectation Bounds and Choquet Integral for Real Valued Functions*

Imprecise probabilities imply that the expectation of a function $f : S \rightarrow Y$ of a discrete or continuous variable is imprecise: each member of set $\Psi^E$ gives a specific value, so obtaining a convex interval bounded by the lower and upper prevision, according to Walley's theory. A search for minimum/maximum values must be solved to evaluate the bounds.

Choquet Integral is a mathematical procedure that in some cases allows the bounds to be directly and exactly evaluated through integration (summation for discrete variables) of the function multiplied by lower/upper probabilities of subsets of the measurable space $(S, \mathcal{S})$. In other cases the Choquet Integral does not give exact values, however it identifies bounds of a wider interval including the effective lower/upper previsions.

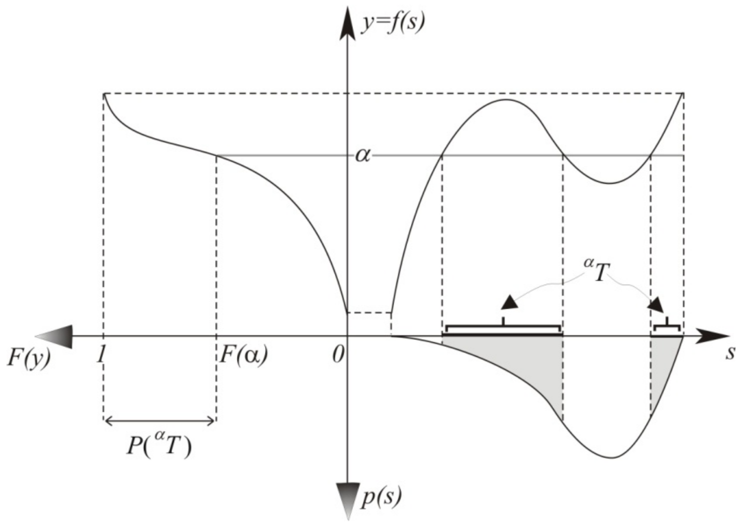



**Fig. 3.21** α-cuts $^{\alpha}T$ of point-valued function $f$

Let us first observe that the expectation of a point-valued function $f$: $S \rightarrow Y=[y_L, y_R] \subset \mathbb{R}$ can be calculated, indicating with $F(y)$ the CDF of the dependant variable $y$, through the Stieltjes Integral (Eq. (2.18)) and equivalent expressions:

$$\begin{aligned} E[y=f] &= \int_{y_L}^{y_R} f \cdot dF = \int_{y_L}^{y_R} y \cdot F'(y)dy \overset{\substack{\text{Integrate} \\ \text{by parts}}}{=} [yF]_{y_L}^{y_R} - \int_{y_L}^{y_R} Fdy = \\ &= y_R - \int_{y_L}^{y_R} Fdy = y_L + \int_{y_L}^{y_R} (1-F)dy = y_L + \int_{y_L}^{y_R} P(f > \alpha)d\alpha = \\ &= y_L + \int_{y_L}^{y_R} P(^{\alpha}T = \{s \in S \mid f(s) > \alpha\})d\alpha \end{aligned} \tag{3.55}$$

The second equality can be found for example in (Kolmogorov and Fomin, 1975; example 2, p. 364)). The geometrical meaning of set $^{\alpha}T$ is shown in Figure 3.21.

Choquet Integral is the direct extension of the last functional expression to a monotonic measure μ given on an appropriate family $\mathcal{C}$ of subsets of $S$:

$$C(f,\mu) = (\mathcal{C})\int f \cdot d\mu = y_L + \int_{y_L}^{y_R} \mu(^{\alpha}T)d\alpha \tag{3.56}$$

When $S$ is a finite space $\{s^j, j = 1 \text{ to } |S|\}$ and the single function $f$ obtains values: $\alpha^1 = y_R > \alpha^2 > \ldots . > \alpha^n = y_L$, Eq. (3.56) becomes:

$$C(f,\mu) = y_L + \sum_{i=1}^{n-1} \mu\left({}^{\alpha^i}T\right)\left(\alpha^i - \alpha^{i+1}\right) \tag{3.57}$$

Eq. (3.57) is equivalent to reordering the space $S$ through a permutation $p = \pi(j)$ of the indexes $j$ in such a way that the function $f$ changes to a monotonically decreasing function: $f(s^{\pi(j)=1}) = y_R \geq f(s^{\pi(j)=2}) \geq \ldots f(s^p) \ldots \geq f(s^{\pi(j)=|S|}) = y_L$, and ${}^{\alpha^i}T = \left\{s^1, \ldots\ldots, f^{-1}\left(\alpha^{i+1}\right)\right\}$ in Eq. (3.57).

Comparing the calculation with the procedures introduced in §3.2.3.4 to evaluate extreme distributions of random sets and extended in §3.3.3 to more general imprecise probabilities (see Eq. (3.54)), it is clear that the Choquet Integral evaluated for $\mu = \mu_{LOW}$ and $\mu = \mu_{UPP}$ is coincident with the expectation of the function $f$ using the particular extreme distribution in the set *EXT* corresponding to the selected permutation.

Moreover:

- For random sets $Bel = \mu_{\mathrm{LOW}}$ is a Choquet capacity of the order $\infty$: *EXT* contains the extremes of the set $\Psi = \Psi^E$: therefore the Choquet Integral gives exactly the lower/upper previsions of any point-valued bounded function;
- When $\mu_{LOW}$ is a Choquet capacity of order $k = 2$, $EXT \in \Psi^E$, but $\Psi$ could be a subset of $\Psi^E$ for specific problems. The Choquet Integral could give values not contained in the interval of the effective lower/upper previsions, but the values are in any case coherent with the upper/lower probabilities corresponding to $\Psi$.
- When $\mu_{LOW}$ is simply a monotone measure the convex hull $\Psi^{EXT}$ of set *EXT* contains probability distributions that could not be contained both in $\Psi$ and $\Psi^E$. Although the inclusion $\Psi \subseteq \Psi^E \subseteq \Psi^{EXT}$ is in any case respected, strict inclusion could appear in specific cases. The Choquet Integral could give values not contained in the interval of the effective lower/upper previsions, and not coherent with the upper/lower probabilities corresponding to $\Psi$.

**Example 3.22.** Let us consider set $\Psi = \Psi^E$ and the coherent monotone lower probabilities ($k = 1$) describing the problem considered in Example 3.15 and Example 3.20.

The lower/upper bounds of the prevision of the bending moment in the middle section of the beam ($M = F_1 l/8 + F_2 l/4$) are given by the expectations of $M$ evaluated by the two extreme distributions $P^1_{EXT}$ and $P^2_{EXT}$ shown in Figure 3.17:

$E_{LOW}[M]$ = min (($F\,l/8$ + F $l/4$) x 0.5 + 0 x 0.5, $F\,l/8$ x 0.5 + $F\,l/4$ x 0.5) =
= min ($3F\,l/16$, $3F\,l/16$) = $3F\,l/16$
$E_{UPP}[M]$ = max ($3F\,l/16$, $3F\,l/16$) = $3F\,l/16$.

In this particular case a precise prevision is therefore obtained.

Table 3.11 summarizes the calculation of the bounds through the Choquet Integral. It is easy to check that the obtained values coincide with the expectation of $M$ according to the probability distributions $P_{EXT,5}$ and $P_{EXT,6}$ shown in Figure 3.19; moreover the lower/upper bounds of the expectation in the set $EXT$ are obtained by the probability distributions $P_{EXT,3}$ ($E[M] = F\,l/16$) and $P_{EXT,4}$ ($E[M] = 5\,F\,l/16$). However, as specified in Example 3.20, these four extreme distributions do not respect the original information that loading or unloading of the forces $F_1$ and $F_2$ are equally likely and consequently the precise marginals of the joint distribution.

**Table 3.11** Choquet Integrals of the bending moment M in Example 3.15

| $j$ | $M$(j) | $p = \pi$(j) | $f(s^p)$- $f(s^{p+1})$ | ${}^{\alpha}T$ | $\mu_{LOW}({}^{\alpha}T)$ | $\mu_{UPP}({}^{\alpha}T)$ |
|---|---|---|---|---|---|---|
| 1 | 3 $Fl/8$ | 1 | $Fl/8$ | $\{s^1\}$ | 0 | 1/2 |
| 2 | $Fl/8$ | 3 | $Fl/8$ | $\{s^1, s^2, s^3\}$ | 1/2 | 1 |
| 3 | 2 $Fl/8$ | 2 | $Fl/8$ | $\{s^1, s^2\}$ | 1/2 | 1/2 |
| 4 | 0 | 4 | | | | |

$$C(M, \mu_{LOW}) = 0 + 0 \text{ x } Fl/8 + 1/2 \text{ x } Fl/8 + 1/2 \text{ x } Fl/8 = \quad Fl/8 = 2 \text{ x } Fl/16$$

$$C(M, \mu_{UPP}) = 0 + 1/2 \text{ x } Fl/8 + 1/2 \text{ x } Fl/8 + 1 \text{ x } Fl/8 = 2 \text{ x } Fl/8 = 4 \text{ x } Fl/16$$

**Example 3.23.** Let us consider set $\Psi$ describing the problem considered in Example 3.14 and Example 3.21. We are interested in evaluating the expectation of the safety margin with respect to the limit state of flexural yielding of the middle section of the beam:

$$z(s^j) = \mathrm{M}_y - F(s^j)\,1/4$$

Assuming $M_y$ = 300 kNm and $l$ = 4 m, the calculation of the bounds through the Choquet Integral is summarized in Table 3.12. The lower bound (-50/3 kNm) equals the expectation according to the probability distributions $P^4{}_{EXT}$: therefore this value respects the lower probabilities (2-monotone Choquet capacities) derived from the original information ($P^4{}_{EXT} \in \Psi^E$) but is not congruent with the probabilistic content of the original information ($P^4{}_{EXT} \notin \Psi$). The effective lower bound of the expectation is positive and equals the expectation according to the probability distributions $P^2{}_{EXT}$:

$$E_{LOW}(z) = 1/3 \text{ x } 200 + 1/3 \text{ x } (-200) + 1/3 \text{ x } 100 = 100/3 \text{ kNm}$$

On the other hand the upper bound (75 kNm) is exactly evaluated by the Choquet Integral and equals the expectation according to $P_{EXT,3}$.

**Table 3.12** Choquet Integrals of the safety margin z in Example 3.14 and Example 3.21. See Table 3.9 for values of µLOW and µUPP

| $j$ | $z(j)$ (kNm) | $p=\pi(j)$ | $f(s^p)-f(s^{p+1})$ | ${}^{\alpha}T$ | $\mu_{LOW}({}^{\alpha}T)$ | $\mu_{UPP}({}^{\alpha}T)$ |
|---|---|---|---|---|---|---|
| 1 | 200 | 1 | 100 | $\{s_1\}$ | 1/3 | ½ |
| 2 | -200 | 3 | | | | |
| 3 | 100 | 2 | 300 | $\{s_1, s_3\}$ | 1/2 | 3/4 |

$C(z, \mu_{LOW})$= -200 + 1/3 x 100 + 1/2 x 300 = - 50/3 kNm
$C(z, \mu_{UPP})$= -200 + 1/2 x 100 + 3/4 x 300 = 75 kNm

### *3.3.5 The Generalized Bayes' Rule*

The numerical examples presented in the previous sections show that lower probabilities (even when coherent) are not sufficiently informative to determine unique lower expectations (lower prevision, according to Walley) of a dependent function. Problems arise mainly from the lack of one-to-one correspondence between lower probabilities and the set $\Psi$ of compatible probability distributions.

The same inadequacy appears when evaluating lower/upper bounds of conditional probabilities or conditional expectation of dependent functions.

Taking into account that:

- the probability of any event (a subset $B$) is equal to the expectation (prevision) of its indicator function;
- the indicator of the intersection of two events (subsets $T$, $B$) is equal to the product of their indicator function,

Bayes Theorem (Eq. (3.32)) can be written as:

$$P(T/B) = E[I_T/B] = \frac{E[I_T \cdot I_B]}{E[I_B]} = \frac{E[I_T \cdot I_B]}{E[I_T \cdot I_B] + E\left[I_{T^C} \cdot I_B\right]} \quad (3.58)$$

More generally, supposing that additional information allows a likelihood function $L(s^j)$ to be specified:

$$P(T/L) = E[I_T/L] = \frac{E[I_T \cdot L]}{E[L]} = \frac{E[I_T \cdot L]}{E[I_T \cdot L] + E\left[I_{T^C} \cdot L\right]} \quad (3.59)$$

(Wasserman and Kadane 1990) demonstrated that, when lower/upper probabilities on the measurable space $S$ are given by monotone set functions $\mu_{LOW}(T)$ and $\mu_{UPP}(T)$ the following inequalities hold:

$$\begin{aligned}
&\frac{C(I_T \cdot L, \mu_{LOW})}{C(I_T \cdot L, \mu_{LOW}) + C(I_{T^C} \cdot L, \mu_{UPP})} \le \frac{E_{LOW}\left[I_T \cdot L\right]}{E_{LOW}\left[I_T \cdot L\right] + E_{UPP}\left[I_{T^C} \cdot L\right]} \le \mu_{LOW}\left(T / L\right) \\
&\mu_{UPP}\left(T / L\right) \le \frac{E_{UPP}\left[I_T \cdot L\right]}{E_{UPP}\left[I_T \cdot L\right] + E_{LOW}\left[I_{T^C} \cdot L\right]} \le \frac{C(I_T \cdot L, \mu_{UPP})}{C(I_T \cdot L, \mu_{UPP}) + C(I_{T^C} \cdot L, \mu_{LOW})}
\end{aligned} \tag{3.60}$$

When $\mu_{LOW}$ $(T)$ and $\mu_{UPP}$ $(T)$ are 2-monotone Choquet capacities and 2-alternating Choquet capacities respectively, and $\Psi = \Psi^E$ (i.e. set $\Psi$ is the larger set compatible with the lower/upper probabilities $\Psi^E$, as it is natural to assume when the original information is effectively given through lower or upper probabilities) the inequalities are respected as equalities.

Equations (3.35) for random sets are particular cases of inequalities (3.60), when conditioning with respect to an event $B$ ($L= I_B$).

**Example 3.24** (modified from (Wasserman and Kadane 1990))**.** Let us again consider the simply-supported beam loaded by two concentrated accidental forces $F_1$ and $F_2$, both of precisely-known intensity F, discussed in Example 3.15. However, although loading and unloading are equally likely for load $F_2$ ( $P(F_2 = \text{F}) = P(F_2 = 0) = 0.5$), no information is given about the probabilities of load $F_1$: it can possibly assume the intensity $F$ in any case, or alternatively it can assume the value 0 in any case. However we only know that the mechanisms generating the actual values of forces $F_1$ and $F_2$ are stochastically (strongly: see Chapter 4) independent, i.e. $P(F_1 \cap F_2) = P(F_1)\, P(F_2)$. Let again

$$\begin{aligned}
S = \{&s^1 = (F_1 = \text{F}) \cap (F_2 = \text{F}),\ s^2 = (F_1 = \text{F}) \cap (F_2 = 0), \\
&s^3 = (F_1 = 0) \cap (F_2 = \text{F}),\ s^4 = (F_1 = 0) \cap (F_2 = 0)\}
\end{aligned}$$

be the space of the possible joint outcomes. The available information is given by imprecise marginal distributions for $F_1$, and determines two opposite extreme distributions $P^5{}_{EXT} = (1/2, 1/2, 0, 0)$ and $P^6{}_{EXT} = (0, 0, 1/2, 1/2)$ (see Figure 3.19). The two extreme joint distributions and projection of set $\Psi$ (a one-dimensional interval) on sub-space ($P(s^1)$, $P(s^2)$, $P(s^3)$) are shown in Figure 3.22.

The lower/upper bounds of every subset or prevision can be evaluated by Eqs. (3.50), simply considering minimum and maximum on the two extremes. Again the Möbius transform (3.8) of the set function $\mu_{LOW}(T)$ gives negative values for all subsets with cardinality equal to 3 and is not 2-monotone.

Supposing that additional information suggests values of the likelihood function $L(s^j)$ proportional to (15, 10, 3, 2), the posterior upper bound of the subset $T = \{s^1\}$ can be evaluated through Eq. (3.60)b:

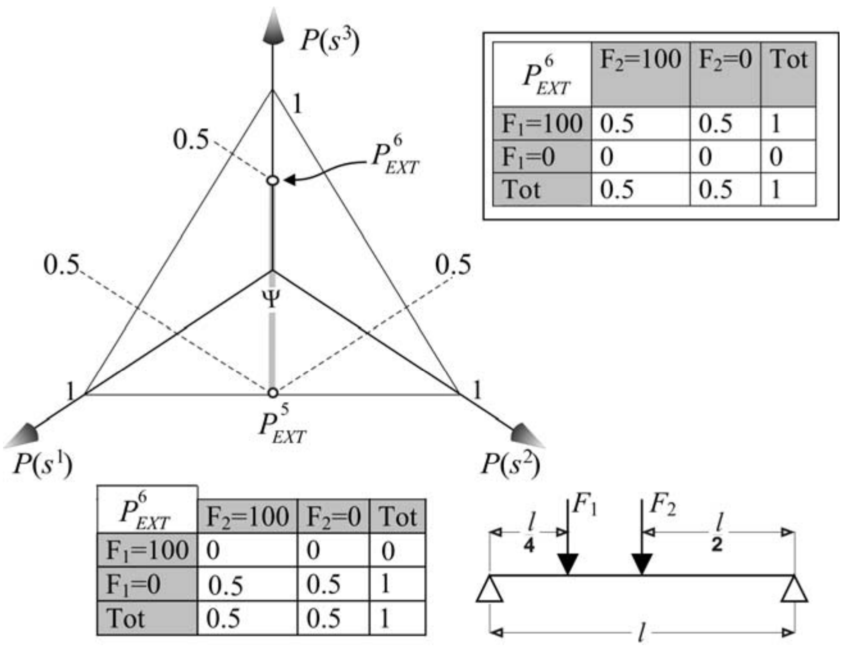
$I_T\, L = (15, 0, 0, 0)$ ; $\qquad I_{T^C}\, L = (0, 10, 3, 2)$

$$\mu_{UPP}(T/L) = \max\left(\frac{15\times(1/2)}{15\times(1/2)+10\times(1/2)}, \frac{15\times(0)}{3\times(1/2)+2\times(1/2)}\right) = 15/25$$

$$E_{UPP}[I_T\, L] = \max(15\text{x } 1/2,\ 15 \text{ x } 0) = 15/2$$
$$E_{LOW}[I_{T^C}\, L] = \min(10 \text{ x } 1/2,\ 3 \text{ x } 1/2 + 2 \text{ x } 1/2) = 5/2$$

$$C_{UPP}[I_T\, L] = 0 + 15 \text{ x } 1/2 + 0 \text{ x } 1/2 + 0 \text{ x } 1 = 15/2$$

$$C_{LOW}[I_{T^C}\, L] = 0 + 7 \text{ x } 0 + 1 \text{ x } 1/2 + 2 \text{ x } 1/2 = 3/2$$




| $P^6_{EXT}$ | $F_2$=100 | $F_2$=0 | Tot |
|---|---|---|---|
| $F_1$=100 | 0.5 | 0.5 | 1 |
| $F_1$=0 | 0 | 0 | 0 |
| Tot | 0.5 | 0.5 | 1 |

| $P^6_{EXT}$ | $F_2$=100 | $F_2$=0 | Tot |
|---|---|---|---|
| $F_1$=100 | 0 | 0 | 0 |
| $F_1$=0 | 0.5 | 0.5 | 1 |
| Tot | 0.5 | 0.5 | 1 |

**Fig. 3.22** Set Ψ,and extreme joint probability distributions in Example 3.24

Therefore:

$$\mu_{UPP}(T/L) = 15/25 < \frac{(15/2)}{(15/2)+(5/2)} = 15/20 < \frac{(15/2)}{(15/2)+(3/2)} = 15/18$$

**Example 3.25.** Let us again consider the simply-supported beam loaded by one concentrated accidental force of uncertain intensity $F$ discussed in Example 3.14. The set Ψ of probability distributions for the possible values of the intensity generates 2-monotone lower probabilities, but is strictly included by the set $\Psi^E = \Psi^{EXT}$ displayed in Figure 3.20. Let us consider the conditional event $T/B = \{s^1\}/\{s^1, s^2\}$. Therefore $L = I_B = (1, 1, 0)$:

$I_T L = (1, 0, 0)$ ; $I_{T^C} L = (0, 1, 0)$

$$\mu_{UPP}(T/L) = \max\left(\frac{1\times(1/2)}{1\times(1/2)+1\times(1/2)}, \frac{1\times(1/3)}{1\times(1/3)+1\times(1/3)}, \frac{1\times(1/2)}{1\times(1/2)+1\times(1/4)}\right)$$
$$= \max(1/2, 1/2, 2/3) = 2/3 = 20/30$$

$$\mu_{LOW}(T/L) = \min(1/2, 1/2, 2/3) = 1/2 = 15/30$$

$E_{UPP}[I_T L]$ = max(1x 1/2, 1 x 1/2, 1 x 1/3)) = 1/2 = 15/30

$E_{LOW}[I_T L]$ = min(1x 1/2, 1 x 1/2, 1 x 1/3)) = 1/3 = 10/30

$E_{UPP}[I_{T^C} L]$ = max(1 x 1/2, 1 x 1/4, + 1 x 1/3) = 1/2 = 15/30

$E_{LOW}[I_{T^C} L]$ = min(1 x 1/2, 1 x 1/4, + 1 x 1/3) = 1/4 = 15/60

$C_{UPP}[I_T L]$ = 0 + 1x 1/2 = 1/2 = 15/30

$C_{LOW}[I_T L]$ = 0 + 1x 1/3 = 1/3 = 10/30

$C_{UPP}[I_{T^C} L]$ = 0 + 1x 1/2 = 1/2 = 15/30

$C_{LOW}[I_{T^C} L]$ = 0 + 1x 1/4 = 1/4 = 15/60

Relations (3.60) are therefore respected:

$$12/30 = 12/30 < \mu_{LOW}(T/L) = 15/30 < \mu_{UPP}(T/L) = 20/30 = 20/30 = 20/30$$

If, given the lower and upper probabilities, it is assumed that $\Psi = \Psi^E$, all strict inequalities change to equalities. The actual maximum and minimum must be searched on the five extreme distributions *EXT* of $\Psi = \Psi^E$; therefore:

$$\mu_{UPP}(T/L) = \max\left(1/2, 1/2, 2/3, \frac{1\times(5/12)}{1\times(5/12)+1\times(1/4)}, \frac{1\times(1/3)}{1\times(1/3)+1\times(1/2)}\right) =$$
$$= \max(1/2, 1/2, 2/3, 5/8, 2/5) = 2/3 = 20/30$$

$$\mu_{LOW}(T/L) = \min(1/2, 1/2, 2/3, 5/8, 2/5) = 2/5 = 12/30$$

We will now introduce a more general formulation for Bayes rule, that in any case directly gives the exact result, overcoming the above-discussed limitations. This formulation was proposed by Walley in his theory of coherent lower/upper previsions (see § 3.3.2 in this Chapter).

For a particular precise probability distribution, the conditional prevision of a function *f* with respect to a likelihood function *L* can be expressed generalizing Eq. (3.59) to the following:

$$E[f \cdot L] - E[f/L] \cdot E[L] = E\left[L\left(f - E[f/L]\right)\right] = 0 \tag{3.61}$$

The unknown conditional prevision $\alpha = E[f/L]$, conditional to the likelihood $L$ of positive prior expectation $E[L]$, can be derived by solving the equation:

$$E\left[L \cdot (f - \alpha)\right] = 0 \tag{3.62}$$

Equation (3.62) has a unique solution α because the left side is strictly decreasing in α if $E[L] > 0$. When $E[L] = 0$, α is undetermined.

**Example 3.26.** Let us consider the particular extreme distribution $P^5_{EXT}$ = (1/2, 1/2, 0,0) in Example 3.24 and again suppose that additional information suggests values of the likelihood function $L(s^j)$ proportional to (15, 10, 3, 2). The expectation $E[L]$ = 15x1/2 + 10x1/2 = 25/2 is positive. We want to calculate the conditional expectation of the function $f$ = (10, 8, 7, -3). The prior expectation $E[f]$ is equal to 10x1/2 + 8 x 1/2 = 9. Eq. (3.62) gives:

15 x(10 - α) x 1/2 + 10 x (8 - α) x 1/2 + 3 x (7 - α) x 0 +2 x (-3 - α) x 0 = 0

α = $E[f/L]$ = (230/2) / (25/2) = 230/25 = 9.2

If alternatively the likelihood function $L(s^j)$ is proportional to (0, 0, 3, 2), $E[L] = 0$ and Eq. (3.62) is satisfied for any value of $\alpha \in (-\infty, \infty)$.

The extension to imprecise coherent lower/upper previsions is quite natural: assuming that $E_{LOW}[L] > 0$[N 3-8] , $\alpha = E_{LOW}[f/L]$ is the unique solution of the following Generalized Bayes' Rule:

$$E_{LOW}\left[L \cdot (f - \alpha)\right] = 0 \tag{3.63}$$

In the calculation of the $E_{LOW}$ operator for the function $L \cdot (f - \alpha)$, Eq. (3.50)a should be solved, or more easily, when the set of the extreme distributions $EXT_\Psi$ is known or has been evaluated by the available information, the minimum of the expectation can be searched on this finite set of probability distributions.

**Example 3.27.** Let us consider the lower probabilities (Choquet capacity of order 2) discussed in Example 3.13, generating the set $\Psi = \Psi^E = \Psi^{EXT}$ shown in Figure 3.14. $EXT = EXT_\Psi$ = {(0.5, 0, 0.5), (0, 0.5, 0.5), (0.5, 0.5, 0)}. For any bounded point valued function $f = (f_1, f_2, f_3)$ the lower prevision is given by:

$E_{LOW}[f] = \min(f_1/2 + f_3/2,\ f_2/2 + f_3/2,\ f_1/2 + f_2/2)$

Let $f = (1, 0, 1)$, and therefore $E_{LOW}[f] = \min(1, 0.5, 0.5) = 0.5$; additional information suggests that the likelihood function $L(s^j)$ is proportional to (4, 1, 2), and hence $E_{LOW}[L] = \min(3, 1.5, 2.5) = 1.5 > 0$. Eq. (3.63) gives:

$$\begin{aligned}\min(&4 \text{ x } (1 - \alpha) \text{ x } 0.5 + 2 \text{ x } (1 - \alpha) \text{ x } 0.5,\\ &1 \text{ x } (0 - \alpha) \text{ x } 0.5 + 2 \text{ x } (1 - \alpha) \text{ x } 0.5,\\ &4 \text{ x } (1 - \alpha) \text{ x } 0.5 + 1 \text{ x } (0 - \alpha) \text{ x } 0.5 \;) = 0\end{aligned}$$

$$\alpha = E[f/L] = \min(3/3, 1/1.5, 2/2.5) = 0.667$$

Of course the same result can be obtained applying Bayes' Theorem separately with the 3 extreme distributions and searching for the minimum. Moreover, observe that $f = (1, 0, 1)$ is the indicator $I_T$ of the set $T = \{s^1, s^3\}$; therefore the same result can be exactly obtained through Eq. (3.60), taking into account that lower probabilities are Choquet capacities of order $k = 2$ and set $\Psi$ coincides with $\Psi^E$:

$$I_T \cdot L = (4, 0, 2);\quad E_{LOW}[I_T \cdot L] = \min(3, \quad 1, \quad 2) = 1$$

$$I_{T^C} \cdot L = (0, 1, 0);\quad E_{UPP}[I_{T^C} \cdot L] = \max(0, 0.5, 0.5) = 0.5$$

$$E[f/L] = E[I_T/L] = \frac{(1)}{(1)+(0.5)} = 0.667$$

**Example 3.28.** Let us again consider Example 3.24, where the extreme distributions $EXT_\Psi = \{(0.5, 0{,}5, 0, 0), (0, 0, 0.5, 0.5)\}$ generate 1-monotone lower probabilities. Let us again evaluate through the Generalized Bayes' Theorem the upper bound of the probability of subset $T = \{s^1\}$ conditional to additional information suggesting values of the likelihood function $L(s^j)$ proportional to (15, 10, 3, 2). For any bounded point valued function $f = (f_1, f_2, f_3, f_4)$ the upper prevision is given by:

$$E_{UPP}[f] = \max(f_1/2 + f_2/2,\ f_3/2 + f_4/2) = \min(-f_1/2 - f_2/2,\ -f_3/2 - f_4/2)$$

The prior upper bound is equal to $E_{UPP}[f = I_T = (1, 0, 0, 0)] = \max(0.5, 0) = 0.5$. $E_{LOW}[L] = \min(12.5, 2.5) = 5/2 > 0$. The posterior is solution $\alpha$ of the equation:

$$\max(15 \text{ x}(1 - \alpha) \text{ x } 0.5 + 10 \text{ x } (0 - \alpha) \text{ x } 0.5\ ,\ 3 \text{ x } (0 - \alpha) \text{ x } 0.5+2 \text{ x } (0 - \alpha) \text{ x } 0.5) = 0$$

Hence: $\alpha = \max(7.5/12.5, 0/2.5) = 15/25$.

Observe that assuming $L(s^j)$ is proportional to (15, 10, 0, 0), $E_{LOW}[L] = \min(12.5, 0) = 0$, although the upper bound is positive. Therefore, taking into account that in this case $\alpha$ is a probability:

$$\alpha = \max(7.5/12.5, [0,1]) = [15/25, 1]$$

## 3.4 Credal Sets

The term "*credal set*" is frequently used when the information is assessed giving upper and lower bounds to the probability of the singletons $\{s^j\}$ or cumulative probabilities of the sets $\{s^1, s^2, \ldots s^j\}$ (i.e. upper/lower bounds of the CDF). In the former case the terms "*Interval valued Probabilities*" or "*Interval Probabilities*" are used in the literature. In the latter (Ferson, Kreinovich et al. 2003) suggested the term "*Probability boxes*", or more simply "*P-boxes*".

In both cases the bounds must respect Eq. (3.46) (the lower bound is ≥ 0; the upper is ≤ 1) and Eq. (3.47) (the upper bound must be greater than the lower one).

Finally a convex set Ψ of the probability distributions can be implicitly defined by a set of parametric probability distributions when a convex set of values of the parameters is assumed or inferred from statistical data.

### *3.4.1 Interval Valued Probabilities*

Given a finite space *S*, or a finite partition of a space *S*, the interval valued probabilities are a set of intervals attached to the singletons, defining a set $\Psi = \Psi^E$ of the probability distributions on the algebra generated by the singletons:

$$\begin{aligned} &I_j = [l_j, u_j],\ j = 1 \ \text{ to} \mid S \mid \\ &\Psi = \left\{P : l_j \le P(s^j) \le u_j,\ j = 1 \ \text{ to} \mid S \mid\right\} \end{aligned} \tag{3.64}$$

According to Eq. (3.48) the set Ψ is non-empty if and only if the bounds satisfy the condition:

$$\sum_{j=1}^{|S|} l_j \le 1 \quad ; \quad \sum_{j=1}^{|S|} u_j \ge 1 \tag{3.65}$$

However stronger further conditions are required for *coherence*: relation (3.50) requires that each upper or lower bound should be "*reachable*": i.e. for every bound, set Ψ should contain a particular distribution *P* equally satisfying the bounds in the second formula in Eq. (3.64).

**Example 3.29.** The macro-seismic scale EMS98 (Grünthal 1998) defines 6 vulnerability classes (from the most vulnerable A to F) of ordinary multi-storey buildings, and for each class and each macro-seismic intensity suggests implicit interval probabilities for 6 qualitative damage levels *d* (from 0 (= no damage) to 5 (total collapse) as a function of a parameter α of credibility (from 0 (minimum) to 1 (maximum) (Bernardini 2005). For α = 1, class B and macro-seismic intensity VI only 2 damage levels (1(negligible) and 2 (moderate)) are possibly expected, and the interval percentages of buildings are specified as follows:

$I_{d=0}$ = [23, 100] ; $I_{d=1}$ = [20, 50] ; $I_{d=2}$ = [0, 10]

But clearly the bounds of intervals $I_{d=0}$ = [23, 100] cannot be reached. In fact at least 20% of buildings suffer damage > 0, and so the percentage of undamaged buildings cannot be more than 80%. Moreover no more than 10+50 = 60% of buildings suffer damage > 0, and therefore the percentage of undamaged buildings cannot be lower than 40%.

The generalization of the procedure used in Example 3.29 leads to the following conditions to assure that the bounds ($l_j$ , $u_j$) are reachable:

$$\forall j : \sum_{i \neq j} l_i + u_j \leq 1 \quad ; \quad \sum_{i \neq j} u_i + l_j \geq 1 \tag{3.66}$$

or equivalently:

$$\forall j : 1 - \sum_{i=1}^{|S|} l_i \geq u_j - l_j \quad ; \quad \sum_{i=1}^{|S|} u_i - 1 \geq u_j - l_j \tag{3.67}$$

$$\forall j : u_j - l_j \leq \min\left(1 - \sum_{i=1}^{|S|} l_i, \sum_{i=1}^{|S|} u_i - 1\right) \tag{3.68}$$

When non-reachable interval valued probabilities are given, the reachable counterpart, i.e. a set of coherent interval valued probabilities can be derived with the formulas:

$$\begin{aligned} \forall j : l_j^{'} &= \max(l_j, u_j - (\sum_{i=1}^{|S|} u_i \text{-} 1)) \\ \mathrm{u}_j^{'} &= \min(u_j, l_j + (1 - \sum_{i=1}^{|S|} l_i)) \end{aligned} \tag{3.69}$$

The reachable interval valued probabilities can be used to evaluate the bounds of the probability of any other set $T$ in the algebra generated by the singletons:

$$\begin{aligned} \mu_{LOW}(T) &= \max(\sum_{s^i \in T} l_i^{'}, 1 - \sum_{s^i \in T^c} u_i^{'}) \\ \mu_{UPP}(T) &= \min(\sum_{s^i \in T} u_i^{'}, 1 - \sum_{s^i \in T^c} l_i^{'}) \end{aligned} \tag{3.70}$$

It is possible to show that lower and upper probabilities given by Eq. (3.70) are 2-monotone and 2-alternating Choquet capacities respectively (Campos, Huete et al. 1994). The set Ψ of compatible probability distributions is directly defined by the intersection of pairs of parallel hyper-planes with the unit hyper-triangle in the |S|-dimensional space of the probabilities of singletons. The set *EXT* of extreme distributions can be derived by index permutations (Eq. (3.54)). However a more efficient recursive algorithm suggested in (Campos, Huete et al. 1994) can be used.

**Example 3.30.** Let us again consider Example 3.29. Table 3.13 shows the reachable bounds and cumulative extreme distributions evaluated through Eq. (3.70), corresponding to extreme distributions $P^1_{EXT}$ = (0.4, 0.5, 0.1) and $P^2_{EXT}$ = (0.8, 0.2, 0) respectively. However 2 other extreme distributions ((0.5, 0.5, 0) and (0.7, 0.2, 0.1)) can be discovered by permutation of the indexes or directly by intersection of pairs of parallel planes with the unit triangle in the 3-dimensional space, or by its projection on the 2-dimensional plane ($p(s^1)$, $p(s^2)$) shown in Figure 3.23. It can be checked that the lower probabilities are 2-monotone Choquet capacities (in this particular case ∞-monotone Choquet capacities) and $\Psi = \Psi^E$: i.e. no other distribution is contained in the set *EXT*.

**Table 3.13** Reachable bounds and lower/upper CDF in Example 3.29

| $d$ | $l$ (%) | $u$ (%) | $l'$ (%) | $u'$ (%) | $\mu_{LOW}(\{s^1,\ldots s^j\})$ | $\mu_{UPP}(\{s^1,\ldots s^j\})$ |
|---|---|---|---|---|---|---|
| $s^1(d=0)$ | 23 | 100 | 40 | 80 | 40 | 80 |
| $s^2(d=1)$ | 20 | 50 | 20 | 50 | 90 | 100 |
| $s^3(d=2)$ | 0 | 10 | 0 | 10 | 100 | 100 |

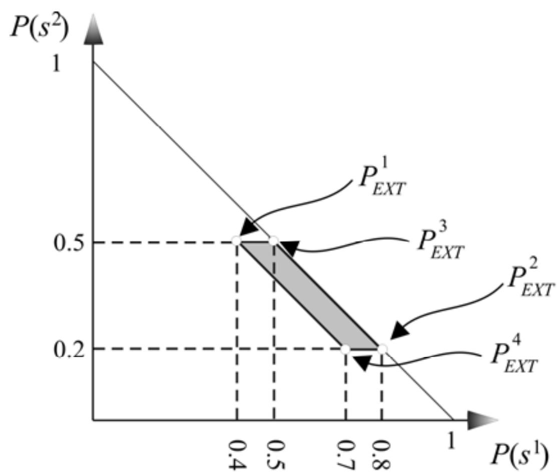



**Fig. 3.23** Set $\Psi = \Psi^E$ and extreme distributions *EXT* in Example 3.30

### 3.4.2 P-Boxes

Given a finite space $S$, a set $\Psi = \Psi^E$ of probability distributions is implicitly defined by lower and upper bounds, $F_{LOW}(s^j)$ and $F_{UPP}(s^j)$, of the cumulative distribution functions $F(s^j)$:

$$\Psi = \left\{ P : F_{LOW}(s^j) \leq F(s^j) = P\left(\left\{s^1, ..., s^j\right\}\right) \leq F_{UPP}(s^j),\ j = 1 \text{ to} \mid S \mid \right\} \quad (3.71)$$

The set $\Psi$ is non-empty if Eq. (3.44) is respected considering the upper bound $F_{UPP}(s^j)$ of the probability of set $\{s^1, s^2, \ldots s^j\}$ and the lower bound of any other subset $\{s^1, s^2, \ldots s^k\}$ with $k \leq j$:

$$1 - F_{LOW}(s^k) - 1 + F_{UPP}(s^j) \geq 0; \quad \text{and therefore:} \quad F_{LOW}(s^k) \leq F_{UPP}(s^j)$$

However coherence clearly requires stronger conditions: the bounds $F_{LOW}(s^j)$ and $F_{UPP}(s^j)$ should be non-negative, non-decreasing in $j$, both equating 1 for $j = |S|$ ((Walley 1991), § 4.6.6).

Explicit evaluation of set $\Psi$ can be obtained solving the constraints (3.71) for the probabilities of the singletons $P(s^j)$:

$$\begin{aligned}
&F_{LOW}(s^1) \leq P\left(s^1\right) \leq F_{UPP}(s^1); \qquad P\left(s^1\right) \geq 0 \\
&F_{LOW}(s^2) \leq P\left(s^1\right) + P\left(s^2\right) \leq F_{UPP}(s^2); \quad P\left(s^2\right) \geq 0 \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
&F_{LOW}(s^j) \leq P\left(s^j\right) + \sum_{i=1}^{j-1} P\left(s^i\right) \leq F_{UPP}(s^j); P\left(s^j\right) \geq 0 \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
&P\left(s^{j=|S|}\right) + \sum_{i=1}^{|S|-1} P\left(s^i\right) = 1; \quad P\left(s^{j=|S|}\right) \geq 0
\end{aligned} \quad (3.72)$$

A simple iterative procedure can be used to solve this equation. For example, the explicit solution of the first two constraints is shown in Figure 3.24: observe that the p-box defines 4 or 5 extreme points of the projection of set $\Psi$ on the two-dimensional space $(P(s^1), P(s^2))$:

– case a): $F_{LOW}(s^2) - F_{UPP}(s^1) \geq 0$:

$$\begin{aligned}
P_1 &= (F_{UPP}(s^1), F_{LOW}(s^2) - F_{UPP}(s^1)), \\
P_2 &= (F_{UPP}(s^1), F_{UPP}(s^2) - F_{UPP}(s^1)), \\
P_3 &= (F_{LOW}(s^1), F_{UPP}(s^2) - F_{LOW}(s^1)), \\
P_4 &= (F_{LOW}(s^1), F_{LOW}(s^2) - F_{LOW}(s^1));
\end{aligned}$$

– case b): $F_{LOW}(s^2)$ -$F_{UPP}(s^1) < 0$:

$$P_1^{'} = (F_{UPP}(s^1), 0)),$$
$$P_1^{''} = (F_{LOW}(s^1), 0)) \ \ (\equiv P_4 \text{ if } F_{LOW}(s^2) = F_{LOW}(s^1)).$$

Moreover the reachable interval bounds for the probability of the singletons are given by the intervals:

$[l_1, u_1] = [F_{LOW}(s^1), F_{UPP}(s^1)]$,
$[l_2, u_2] = [\max(0, F_{LOW}(s^2) - F_{UPP}(s^1)), F_{UPP}(s^2) - F_{LOW}(s^1)]$,

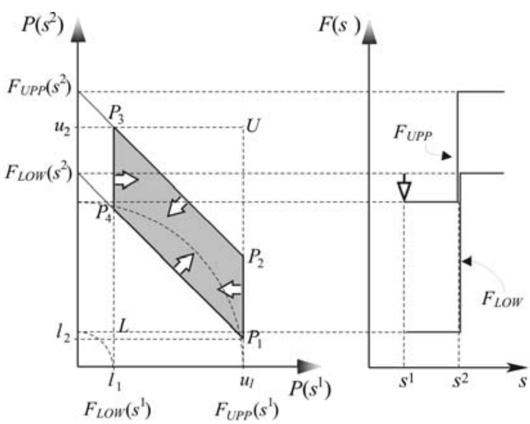
but the set generated by the same (non-interacting) interval probabilities should be much greater (the extreme points $U$=( $u_1$, $u_2$) and $L$=( $l_1$, $l_2$) could appear, if reachable (the last equation in (3.72) must be respected)).

More generally the interval probabilities for singleton $\{s^j\}$ are given by the intervals:

$$[l_j, u_j] = \left[\max\left(0, F_{LOW}\left(s^j\right) - F_{UPP}\left(s^{j-1}\right)\right), F_{UPP}\left(s^j\right) - F_{LOW}\left(s^{j-1}\right)\right] \quad (3.73)$$

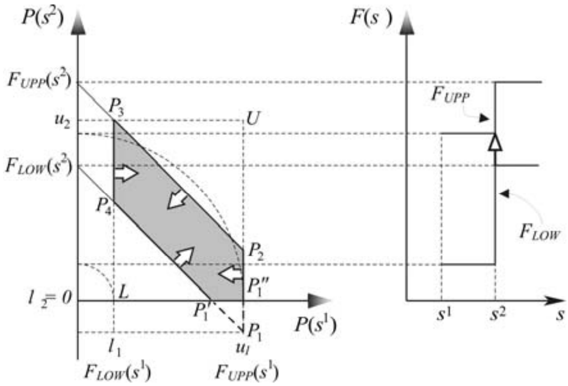
However the extreme distributions obtained by such intervals using the procedure described in §3.4.1 could generally give cumulative distributions functions not contained in the p-box.

The extreme points of the projection of set Ψ on the $j$-dimensional space ($P(s^1)$, …, $P(s^j)$) can therefore be derived from each extreme point on the $j$–1-dimensional space, of course considering that the sum $P(s^1)$+ …+$P(s^j)$ must be bounded by $F_{LOW}(s^j)$ and $F_{UPP}(s^j)$.




Case a

**Fig. 3.24** Explicit solution of the first 2 constraints in Eq. (3.72). Projection of set Ψ is shown hatched. Case a: $F_{LOW}(s^2)$ - $F_{UPP}(s^1) > 0$; case b: $F_{LOW}(s^2)$ - $F_{UPP}(s^1) < 0$




Case b

**Fig. 3.24** (*continued*)

A constructive procedure to evaluate the effective extreme distributions compatible with the information given by a p-box could be obtained by selecting the set *EXT* corresponding to the cumulative (non-decreasing) distribution functions $F$ jumping, at some points $s^j$, from $F_{LOW}(s^j)$ to $F_{UPP}(s^{j+1})$ and, at other points $s^k$, from $F_{UPP}(s^k)$ to $F_{LOW}(s^{k+1})$ (or, if $F_{UPP}(s^k) > F_{LOW}(s^{k+1})$ assuming $F(s^{k+1}) = F_{UPP}(s^k)$ (case b) in Figure 3.24 at point $s^1$).

Of course the set *EXT* contains the distribution functions corresponding to the bounds of the p-box:

$$P_{EXT,LOW}(s^j) = F_{LOW}(s^j) - F_{LOW}(s^{j\text{-}1});\ P_{EXT,UPP}(s^j) = F_{UPP}(s^j) - F_{UPP}(s^{j\text{-}1}).$$

The same set *EXT* (and therefore the same set $\Psi$ of probability distributions) can be given by an equivalent random set, with focal elements and probabilistic assignment derived by the p-box through a rule quite similar to the procedure for deriving an equivalent random set from a normal fuzzy set (considering the membership function as a possibility distribution; see § 3.2.4). The procedure is as follows (with the aid of Figure 3.25):

Let:

$$
\begin{aligned}
&\forall j : F_{LOW}\left(s^{j-}\right) = \lim_{\varepsilon \to 0^+} F_{LOW}\left(s^j - \varepsilon\right) = F_{LOW}\left(s^j\right) - P_{LOW}\left(s^j\right) \\
&\alpha^1 = 1 = \max_j\left(F_{LOW}\left(s^j\right)\right) = \max_j\left(F_{UPP}\left(s^j\right)\right); \\
&\alpha^2 = \max\left(\max_{j \mid F_{LOW}(s^{j-}) < \alpha^1}\left(F_{LOW}\left(s^j\right)\right), \max_{j \mid F_{UPP}(s^j) < \alpha^1}\left(F_{UPP}\left(s^j\right)\right)\right); \\
&\ldots\ldots\ldots \\
&\alpha^i = \max\left(\max_{j \mid F_{LOW}(s^{j-}) < \alpha^{i-1}}\left(F_{LOW}\left(s^j\right)\right), \max_{j \mid F_{UPP}(s^j) < \alpha^{i-1}}\left(F_{UPP}\left(s^j\right)\right)\right) \\
&\ldots\ldots\ldots \\
&\alpha^n = \max\left(\max_{j \mid F_{LOW}(s^{j-}) < \alpha^{n-1}}\left(F_{LOW}\left(s^j\right)\right), \max_{j \mid F_{UPP}(s^j) < \alpha^{n-1}}\left(F_{UPP}\left(s^j\right)\right)\right) = \\
&= \min\left(\min_j\left(F_{LOW}\left(s^j\right)\right), \min_j\left(F_{UPP}\left(s^j\right)\right)\right) \; ; \\
&\alpha^{n+1} = 0
\end{aligned}
\tag{3.74}
$$

and assume:

$$
A^i = \left\{s^j \in S \mid F_{UPP}\left(s^j\right) \geq \alpha^i ; F_{LOW}\left(s^{j-}\right) < \alpha^i\right\}; m(A^i) = \alpha^i - \alpha^{i+1} \tag{3.75}
$$

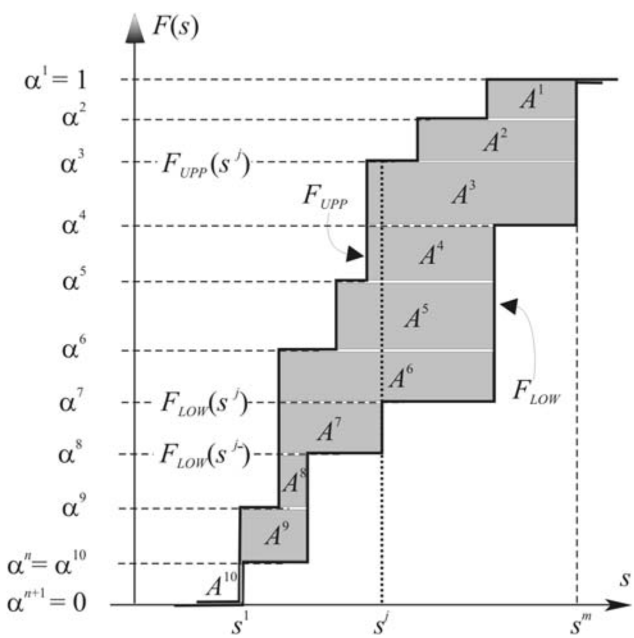



**Fig. 3.25** Random set from a p-box

Consequently:

- the lower/upper probabilities for subsets $T \subseteq S$ are Choquet capacities and Alternate Choquet capacities of infinite order respectively (or Belief and Plausibility set functions respectively);
- the probabilistic assignment of the equivalent random set can alternatively be derived from the Belief function through the Möbius transform;
- the upper bounds $u_j$ of the singletons (Eq. (3.73)) give the contour function of the equivalent random set.

In (Alvarez 2006) the procedure is extended to p-boxes on infinite spaces with general $F_{UPP}$ and $F_{LOW}$, thus deriving equivalent random sets with infinite focal elements given by the α-cuts of the upper/lower CDFs. Tonon (Tonon 2008) deals with inclusion properties for discretizations of upper/lower CDFs.

**Example 3.31.** Let us consider $S = \{s^1, s^2, s^3, s^4\}$ and the p-box defined in the first three columns of Table 3.14. The five extreme points in the two-dimensional space $(P(s^1), P(s^2))$ (case b)) determine 10 extreme points shown in Figure 3.26a for the projection in the three-dimensional space $(P(s^1), P(s^2), P(s^3))$. For example the extreme $P_2 = (0.2, 0.1)$ determines the extremes $(0.2, 0.1, l_3 = 0.4)$ and $(0.2, 0.1, \min(u_3 = 0.9, 1 - 0.2 - 0.1) = 0.7)$ ($P^2_{EXT}$ in Figure 3.26a).

Of course in the four-dimensional space $(P(s^1), P(s^2), P(s^3), P(s^4))$ 10 extreme distributions are obtained by taking $P(s^4)=1- P(s^1) - P(s^2) - P(s^3)$. For example the above indicated pair of extremes determined by $P_2 = (0.2, 0.1)$ give the pair (0.2, 0.1, 0.4, 0.3), (0.2, 0.1, 0.7, 0.0). The extreme points $P^1_{EXT}$ and $P^2_{EXT}$ correspond to the cumulative distribution functions $F_{LOW}(s^j)$ and $F_{UPP}(s^j)$ respectively. Table 3.15 presents the lower probabilities for all the subsets of space $S$ and their Möbius transform $m$, confirming the rules given by Eqs. (3.73) and (3.74) and shown in Figure 3.26b).

**Table 3.14** Reachable bounds and lower/upper CDF in Example 3.31

| $s^j$ | $F_{LOW}(s^j)$ | $F_{UPP}(s^j)$ | $l_j = Bel(\{s^j\})$ | $u_j = Pla(\{s^j\}) = \mu(s^j)$ |
|---|---|---|---|---|
| $s^1$ | 0 | 0.2 | 0 | 0.2 |
| $s^2$ | 0.1 | 0.3 | max(0, 0.1 – 0.2) = 0 | 0.3 – 0 = 0.3 |
| $s^3$ | 0.7 | 1.0 | max(0, 0.7 – 0.3) = 0.4 | 1.0 – 0.1 = 0.9 |
| $s^4$ | 1.0 | 1.0 | max(0, 1 – 1) = 0 | 1.0 – 0.7 = 0.3 |

**Table 3.15** Set functions in Example 3.31

| $i$ | $\chi_i(s_1)$ | $\chi_i(s_2)$ | $\chi_i(s_3)$ | $\chi_i(s_4)$ | $\mu_{LOW}(A^i)$ | $m^i = m(A^i)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 1 | 0 | 0.4 | 0.4 |
| 4 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 1 | 1 | 0 | 0 | 0.1 | 0.1 |
| 6 | 0 | 1 | 1 | 0 | 0.5 | 0.5-0.4=0.1 |
| 7 | 0 | 0 | 1 | 1 | 0.7 | 0.7-0.4=0.3 |
| 8 | 1 | 0 | 1 | 0 | 0.4 | 0.4-0.4=0 |
| 9 | 0 | 1 | 0 | 1 | 0 | 0 |
| 10 | 1 | 0 | 0 | 1 | 0 | 0 |
| 11 | 1 | 1 | 1 | 0 | 0.7 | 0.7-1+0.4=0.1 |
| 12 | 0 | 1 | 1 | 1 | 0.8 | 0.8-1.2+0.4=0 |
| 13 | 1 | 0 | 1 | 1 | 0.7 | 0.7-1.1+0.4=0 |
| 14 | 1 | 1 | 0 | 1 | 0.1 | 0.1-0.1+0=0 |
| 15 | 1 | 1 | 1 | 1 | 1.0 | 1-2.3+1.7-0.4=0 |

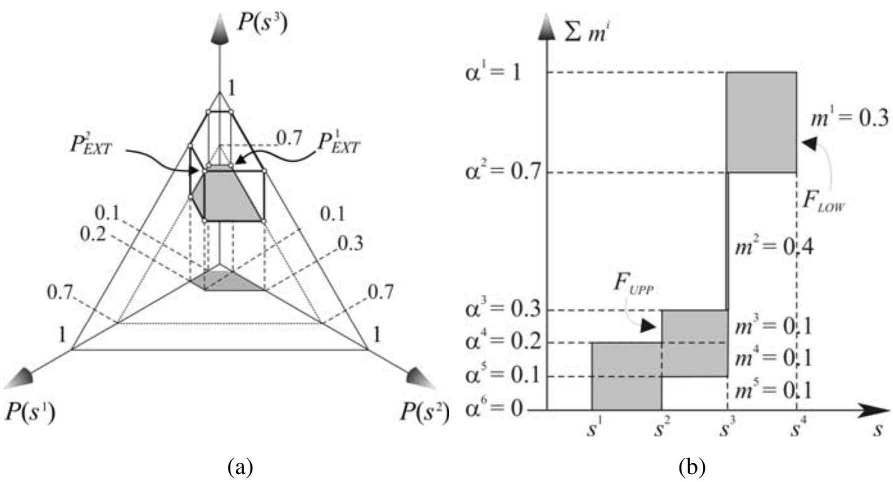



**Fig. 3.26** Example 3.31: (a) extreme points in the 3-dimensional space; (b) equivalent random set

It is easy to show that the random set determined by Eqs. (3.74) and (3.75) is not the only random set compatible with the p-box: indeed, each compatible probability distribution is a particular random set compatible with the bounds of the p-box (with focal elements given by singletons). However it must be considered as the natural extension of the information given by the p-box: the set $\Psi^E$ determined by Eqs. (3.74), (3.75) and (3.10) (or equivalently (3.11)) includes all probability distributions compatible with the p-box and also the set $\Psi$ of any other random set compatible with the p-box.

For example, when the maximum of the contour function $\mu(s^j) = u(s^j)$ defined, through Eq. (3.73), by the p-box is equal to 1, the procedure presented in § 3.2.4) can be used to derive a consonant random set compatible with the p-box: the focal elements are now the $\alpha$-cuts of the contour function and the probabilistic assignment is again defined by the increment of $\alpha$. In other words: the information given by the p-box and additional information suggesting that the structure of the underlying random set is consonant determine a consonant random set and a corresponding set $\Psi_c$ of probability distributions, and of course $\Psi_c \subseteq \Psi^E$.

**Example 3.32.** Let us consider the slowly enlarged p-box with respect to the p-box discussed in Example 3.31, defined by Table 3.16. The 8 extreme points $EXT^{\Psi}$ of set $\Psi = \Psi^E$ (the projection in the two-dimensional space ($P(s^1)$, $P(s^2)$) contains 4 extreme points because $F_{LOW}(s^1) = F_{LOW}(s^2)$) and the underlying non-consonant random set are shown in Figure 3.27 a) and b) respectively. Of course this set $\Psi$ strongly includes the set of probability distributions in Example 3.31, displayed in Figure 3.26. Now $u(s^3) = 1$, so the contour function can be assumed as a possibility distribution determining the consonant random set and corresponding set $\Psi_c$ of compatible distributions shown in Figure 3.28. The set $EXT^{\Psi_c}$, derived using the procedures presented in § 3.2.3.4, contains only 5 (the extremes of a pyramid with vertex in $P^1_{EXT}$ and quadrangular base on the equilateral triangle $P(s^4)= 1 - P(s^1) - P(s^2) - P(s^3) = 0$) of the 8 extremes in set $EXT^{\Psi}$; of course both $EXT^{\Psi_c}$ and $EXT^{\Psi}$ contain the extreme points $P^1_{EXT}$ and $P^2_{EXT}$ corresponding to the cumulative distribution functions $F_{LOW}(s^j)$ and $F_{UPP}(s^j)$ respectively. The same procedure cannot be applied to the p-box discussed in Example 3.31 because the contour function maximum is 0.9 (Table 3.14, last column). Observe that the random set shown in Figure 3.28b gives upper and lower CDFs corresponding to the bounds of the p-box in Example 3.32, and hence outer approximations of the bounds of the p-box in Example 3.31; however the set $\Psi_c$ does not include the set $\Psi^E$ in Example 3.31.

**Table 3.16** Reachable bounds and lower/upper CDF in Example 3.32

| $s^j$ | $F_{LOW}(s^j)$ | $F_{UPP}(s^j)$ | $l= Bel(\{s^j\})$ | $u=Pla(\{s^j\})=\mu(s^j)$ |
|---|---|---|---|---|
| $s^1$ | 0 | 0.2 | 0 | 0.2 |
| $s^2$ | 0 | 0.3 | max(0, 0 – 0.2) = 0 | 0.3 - 0 = 0.3 |
| $s^3$ | 0.7 | 1.0 | max(0, 0.7 – 0.3) = 0.4 | 1.0 – 0 = 1 |
| $s^4$ | 1.0 | 1.0 | max(0, 1 – 1) = 0 | 1.0 – 0.7 = 0.3 |

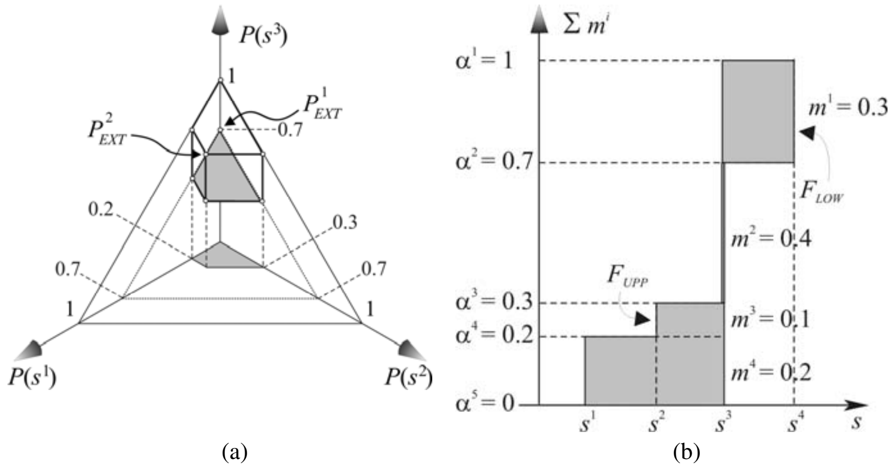



**Fig. 3.27** Example 3.32: (a) extreme points in the 3-dimensional space; (b) equivalent random set

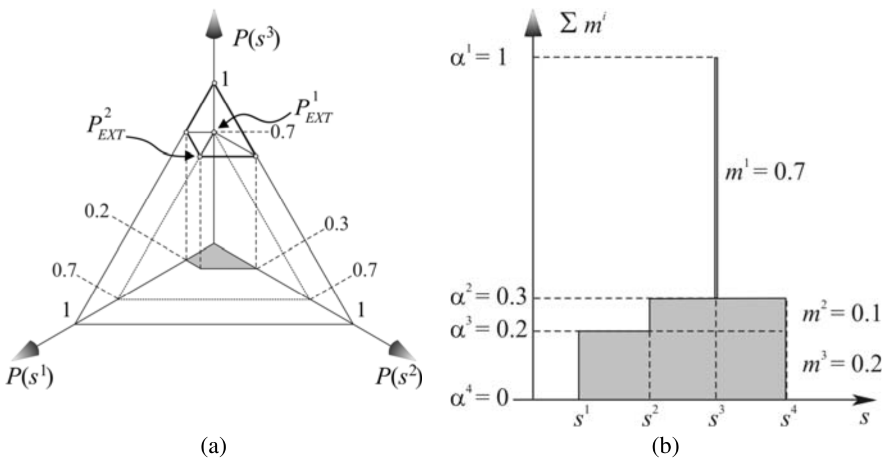



**Fig. 3.28** Consonant random set in Example 3.32: (a) extreme points in the 3-dimensional space; (b) focal elements and probabilistic assignment

Consonant approximations of non consonant random sets measuring variables (for example input variables of engineering systems) are very attractive, as will be shown in Chapter 5.1.3. However, as shown in the above examples, consonant approximations that yield the same (or even outer approximations of ) upper/lower CDFs may not guarantee inclusion of the overall convex sets of compatible probability distributions. Hence, when the reliability of a system is described by non linear function, using such consonant approximations may lead to unsafe predictions of the expected reliability.

**Example 3.33.** Let $S=\{s^1, s^2, s^3, s^4\}$, and consider the point-valued function $f(s^j)$ defined by the mapping: $\{s^1, s^2, s^3, s^4\} \rightarrow \{5, 20, 10, 0\}$. The permutation leading to a monotonic decreasing ordering of the function $f(s^j)$ is the following:

$\pi(s^2) = 1, \pi(s^3) = 2, \pi(s^1) = 3, \pi(s^4) = 4)$.

Table 3.17 shows the corresponding dual extreme distributions giving bounds of the expectation, by the dual set functions *Pla* and *Bel* derived by the p-boxes (and corresponding random sets) considered in Example 3.31 and Example 3.32. Relations (3.18) and (3.19) in these examples give the same extremes, and therefore the same bounds of the expectation:

$$E_{UPP}[f] = 5\times 0.0 + 20\times 0.3 + 10\times 0.7 + 0\times 0.0 = 13.0\ ;$$
$$E_{LOW}[f] = 20\times 0.0 + 10\times 0.5 + 5\times 0.2 + 0\times 0.3 = 6.0$$

On the other hand, considering the information on the space $S$ given by the consonant random set compatible with the p-box in Example 3.32, Table 3.18 shows the extreme distributions that give the expectation bounds of the same function *f*. By comparing with Table 3.17 , $P_{EXT,UPP}(s^j)$ remains unchanged, but $P_{EXT,LOW}(s^j)$ is now different and coincides with $P^1_{EXT}$, hence, again: $E_{UPP}[f] = 13.0$ ; while the lower bound increases to: $E_{LOW}[f] = 20\times 0.0 + 10\times 0.7 + 5\times 0.0 + 0\times 0.3 = 7.0$.

**Table 3.17** Dual extreme distributions for function in Example 3.33 and p-boxes in Example 3.31 and Example 3.32

| $T$ | $Pla(T)$ | $P_{EXT,UPP}(s)$ | $Bel(T)$ | $P_{EXT,LOW}(s)$ |
|---|---|---|---|---|
| $T^1=\{s^2\}$ | 0.3 | $P(s^2) = Pla(T^1)$ $= 0.3$ | 0.0. | $P(s^2)= Bel(T^1)$ $= 0.0$ |
| $T^2=\{s^2,s^3\}$ | 1.0 | $P(s^3) =Pla(T^2) - Pla(T^1)$ $= 0.7$ | 0.5 | $P(s^3)= Bel(T^2) - Bel(T^1)$ $= 0.5$ |
| $T^3=\{s^2,s^3,s^1\}$ | 1.0 | $P(s^1) =Pla(T^3) - Pla(T^2)$ $= 0.0$ | 0.7 | $P(s^1)=Bel(T^3) - Bel(T^2)$ $= 0.2$ |
| $T^4 = S$ | 1.0 | $P(s^4) = Pla(T^4)- Pla(T^3)$ $= 0.0$ | 1.0 | $P(s^4)=Bel(T^4) - Bel(T^3)$ $= 0.3$ |

**Table 3.18** Dual extreme distributions for function in Example 3.33 and the consonant random set included in the p-box considered in Example 3.32

| $T$ | $Pla(T)$ | $P_{EXT,UPP}(s)$ | $Bel(T)$ | $P_{EXT,LOW}(s)$ |
|---|---|---|---|---|
| $T^1=\{s^2\}$ | 0.3 | $P(s^2) = Pla(T^1)$ $= 0.3$ | 0.0. | $P(s^2)= Bel(T^1)$ $= 0.0$ |
| $T^2=\{s^2,s^3\}$ | 1.0 | $P(s^3) =Pla(T^2) - Pla(T^1)$ $= 0.7$ | 0.7 | $P(s^3)= Bel(T^2) - Bel(T^1)$ $= 0.5$ |
| $T^3=\{s^2,s^3,s^1\}$ | 1.0 | $P(s^1) =Pla(T^3) - Pla(T^2)$ $= 0.0$ | 0.0 | $P(s^1)=Bel(T^3) - Bel(T^2)$ $= 0.2$ |
| $T^4 = S$ | 1.0 | $P(s^4) = Pla(T^4)- Pla(T^3)$ 0.0 | $= 0.3$ | $P(s^4)=Bel(T^4) - Bel(T^3)$ $= 0.3$ |

### 3.4.3 Convex Sets of Parametric Probability Distributions

In many applications the available information suggests that a special type of probability distribution can be assumed in modeling a variable, but precise value of parameters $p_1$, $p_2$, …on which the distribution depends (e.g. mean value, variance etc.) are not known: for example a set of values for each parameter (a subset of the space of possible values) is given, or more generally a joint subset $\Psi_{p1,\, p2,\ldots}$ (perhaps convex) of the Cartesian product of the spaces of possible values. The information therefore defines a set $\Psi$ of probability distributions and consequently upper/lower bounds of any event (Eq. (3.37)) or previsions (Eq. (3.50)). However the bounds generally do not correspond to upper/lower bounds (*Bel* / *Pla* set-functions) of an equivalent random set or to 2-monotone Choquet and alternate Choquet capacities.

**Example 3.34.** The damage to residential buildings observed after earthquakes is generally well described by binomial distributions on a finite integer scale $d$ ranging from 0 (no damage) to 5 (total collapse). A binomial distribution is completely defined by its mean value μ, or by the binomial coefficient $p$ in the range between 0 to 1 (equal in this case to μ/5), depending on the seismic intensity and the building structural type.

$$Bin\left(d^{j}, p\right) = \frac{5!}{d^{j}! \cdot (5-d^{j})!} p^{d^{j}} \left(1-p\right)^{5-d^{j}} \mid d^{j} \in \left\{0,\, 1, \ldots, 5\right\} \tag{3.76}$$

Supposing that $p$ is not precisely known, but restricted by the set $\Psi_p = [0.1, 0.4]$, the bounds of each subset of the damage space $D = \{d^1 = 0, d^2 = 1, d^3 = 2, \ldots d^6 = 5\}$ can be evaluated. It is easy to check that the *Bin* function is non-monotonic with respect to $p$: for any value $d^j$ an extreme maximum value is obtained for $p = d^j/5$.

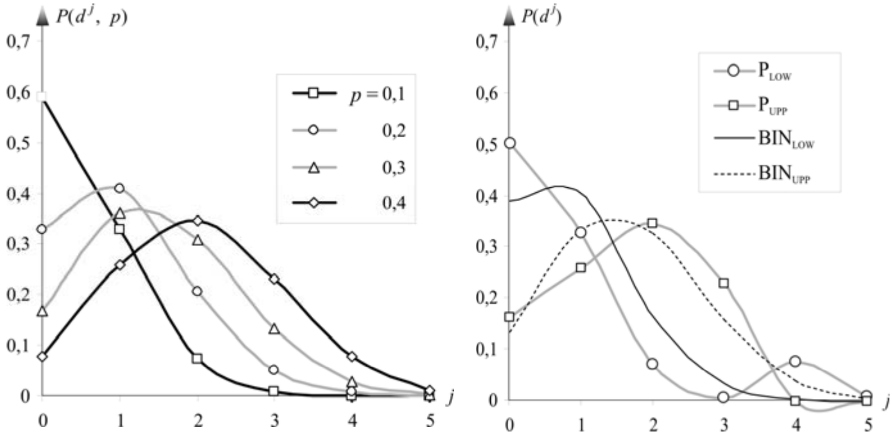



**Fig. 3.29** (a) Probability distributions in Example 3.34 for the 2 extremes and 2 internal points of set $\Psi_p$ = [0.1, 0.4]. (b) Dual extreme distributions ($P_{LOW}$ and $P_{UPP}$) for the permutations (π(0)=3, π(1)=2, π(2)=1, π(3)=0, π(4)=5, π(5)=4) of the indexes, and binomial distributions with the same mean value ($BIN_{LOW}$ and $BIN_{UPP}$)

Therefore for $d^2$ = 1, this extreme value ($p$ = 1/5 = 0.2) is within the interval $\Psi_p$ = [0.1, 0.4] (see Figure 3.29 a): the upper bound of the probability of subset { $d^2$} is not the maximum of the values given by the binomial distribution function defined by two extremes of the convex interval $\Psi_p$.

By checking permutations of the indexes, Eqs. (3.18) (3.19) could be used to derive dual members of the set *EXT* of the extreme distributions. Figure 3.29 b) shows the numerical results for a particular permutation (π(0)=3, π(1)=2, π(2)=1, π(3)=0, π(4)=5, π(5)=4) compared with binomial distributions with the same mean value: the extreme distributions are clearly not binomial distributions. Therefore they do not belong on set Ψ of the probability distributions describing the available information. This conclusion implies that the (coherent) lower probabilities which can be derived by the set Ψ by means of Eq. (3.37)a are not 2-monotone Choquet capacities: the corresponding set EXT contains distributions which do not belong on Ψ (= $\Psi^E$) and hence the lower probabilities are simply 1-monotone set functions.

**Example 3.35** (Example 3 in (Hall and Lawry 2004))**.** A set of log-normal probability distributions is defined on the infinite space $S = \{s \in (0, +\infty)\}$ by imposing that ln $s$ is normal with mean $\mu \in [0.1, 1]$ and standard deviation $\sigma \in [0.1, 0.5]$. Observe that the four extremes of set $\Psi_{\mu\times\sigma} = [0.1, 1]\times[0.1, 0.5]$ define a p-box containing the CDFs of the assumed probability distributions, but the extreme $F_{LOW}$( ln $s$) and $F_{UPP}$( ln $s$) are not normal distributions: for example $F_{LOW}$( ln $s$) =: N (μ = 1, σ = 0.1) for ln s < μ = 1; N (μ = 1, σ = 0.5) for ln $s$ > μ. A finite approximation of the problem was developed in the same paper through a finite partition of space $S$ (or equivalently of space $S_N = \{ \ln s \in (-\infty, +\infty)\}$) given by the 5 singletons $s^j$ (the intervals listed in Table 3.19, second column). The same table gives the upper and lower values of the CDFs on the space of the singletons, and also the upper and lower values of the probability of the singletons, searching for the pair $(\mu,\sigma)$ giving the maximum and minimum respectively. For example $\mu_{UPP}(\{s^3\}) \sim 1$ is given by ($\mu = (0.1+1)/2 = 0.55$, $\sigma = 0.1$); note that it is not an extreme point of $\Psi_{\mu\times\sigma}$. In a similar way the paper gives upper and lower probabilities of each member of the algebra generated by the partition, demonstrating through the Möbius transform $m$ that they are simply 1-monotone set functions.

**Table 3.19** Lower/upper CDFs and probability of the singletons in Example 3.35

| $s^j$ | $s$ | $F_{LOW}(s^j)$ | $F_{UPP}(s^j)$ | $\mu_{LOW}(\{s^j\})$ | $\mu_{UPP}(\{s^j\})$ |
|---|---|---|---|---|---|
| $s^1$ | ( 0, 0.891) | ~0 | 1/3 | ~0 | 1/3 |
| $s^2$ | [0.891, 1.154) | ~0 | 2/3 | ~0 | 0.651 |
| $s^3$ | [1.154, 2.604) | 1/3 | ~1 | 1/3 | ~1 |
| $s^4$ | [2.604, 3.372) | 2/3 | ~1 | ~0 | 0.651 |
| $s^5$ | [3.372, +∞ ) | 1 | 1 | ~0 | 1/3 |

If alternatively the p-box shown in the first 4 columns of Table 3.19 is assumed as the relevant basic information, the above procedure described in § 3.4.2 can be used. Different lower and upper probabilities of the singletons, derived according to Eq. (3.73), are listed in Table 3.20; the focal element and probabilistic assignment of an equivalent non-consonant random set are shown in Figure 3.30a); this random set determines different lower and upper probabilities of each member of the algebra generated by the partition (Choquet and Alternate Choquet capacities of order ∞).

Also observe that the contour function shown in Table 3.20 can be assumed as a possibility distribution, determining the consonant random set shown in Figure 3.30 b), again compatible with the bounds of the p-box. This consonant random set gives upper and lower probabilities not very different from the values directly derived from set $\Psi_{\mu\times\sigma}$: for example $Bel(\{s^3\})$ is exactly equal to 1/3, while $Pla(\{s^2\}) = Pla(\{s^4\})$ increases from 0.651 to 2/3 = 0.667.

This consonant random set is not very different from the nearly-consonant random set obtained in (Hall and Lawry 2004) through a specific procedure proposed by the authors and called Iterative Rescaling Method (IRM). However it is important to underline that the random set obtained by IRM (or its consonant approximation shown in Figure 3.30 b) and the non-consonant random set in Figure 3.30 a), although giving the same upper/lower CDFs, are not comparable: they are not different discrete approximations of the same information, but discrete approximations of different original information (the set of log-normal distributions in the former case, the bounds of the corresponding p-box in the latter).

**Table 3.20** Lower/upper probabilities of the singletons from the p-box in Table 3.19

| $s^j$ | $F_{LOW}(s^j)$ | $F_{UPP}(s^j)$ | $\mu_{LOW}(\{s^j\})$ | $\mu_{UPP}(\{s^j\})$ |
|---|---|---|---|---|
| $s^1$ | ~0 | 1/3 | ~0 | 1/3 |
| $s^2$ | ~0 | 2/3 | max (0,~0-1/3) = 0 | 2/3 - ~0 =0.667 |
| $s^3$ | 1/3 | ~1 | max (0,1/3-2/3) = 0 | ~1 - ~0 = ~1 |
| $s^4$ | 2/3 | ~1 | max (0,2/3- ~1) = 0 | ~1 – 1/3 = 0.667 |
| $s^5$ | 1 | 1 | max (0,1- ~1) = ~0 | 1 – 2/3 = 1/3 |

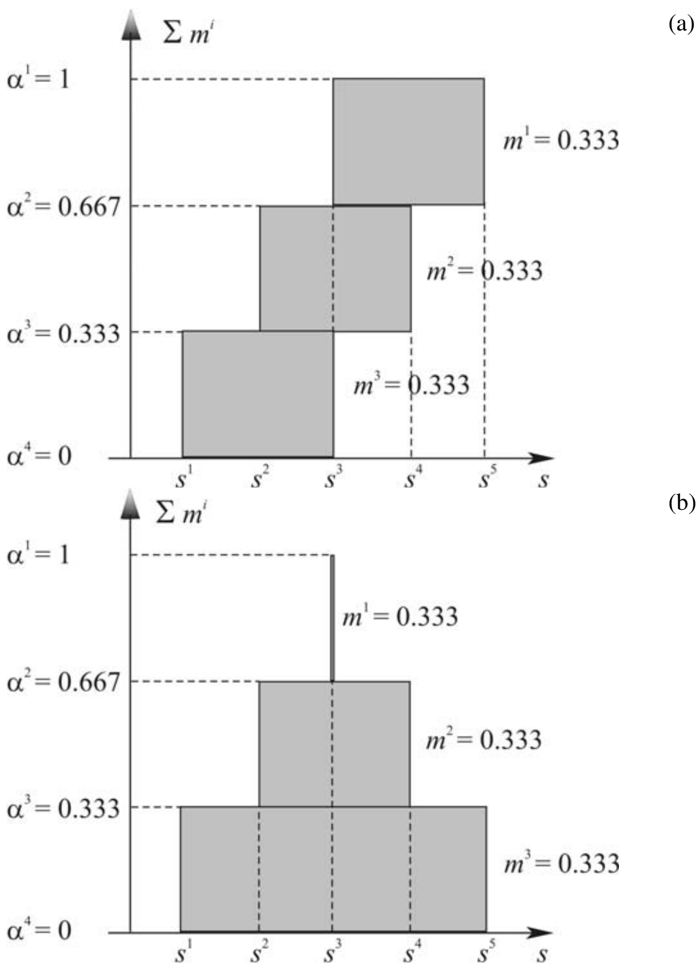



**Fig. 3.30** Random sets in Example 3.35: (a) non-consonant random set by the p-box; (b) consonant random set

## 3.5 Conclusions

The concept of random set has been introduced in this chapter, demonstrating that in many cases, but not all, it can be useful to model the actual uncertainty about variables; in particular, in view of the applications to be discussed in the following chapters, the variables of interest to engineering problems. This model appears to be a powerful generalization of the classical probability theory, but it is a particular case of a more general theory of measures, related to classes of events through monotone non-additive functions.

It has been shown that the theory of random sets contains as particular cases, in addition to the probability measures, both the models of uncertainty based on the classical set theory (convex modelling, interval analysis) and the so-called possibility theory, which, on the other hand, can be considered as a particular interpretation of the general fuzzy sets theory.

In the following chapters the theory of random sets will be developed considering multivariate problems and therefore analyzing the reciprocal interaction (dependence, independence, correlation) between different variables and its influence on the estimation of dependent functions. Within this broader ambit the *extension principle* will be introduced, with reference both to point-valued functions and to more general set-valued functions, and the very important rule of *inclusion* of random sets, which appears to be of particular interest in applications for reliability and risk evaluations of engineering systems.

It will be shown that in many applications, uncertain data, which would require rigorous but complex procedures within the ambits of imprecise probabilities or monotone set measures, can be modelled with good approximation by means of random sets, so obtaining remarkable computational advantages.

## Notes

**N 3-1** Evaluation of the selectors requires precise knowledge of multi-valued mapping, which generates focal elements not necessarily distinct. In computing *Bel*(*T*), *Pla*(*T*) and $\Psi$ , coincident focal elements can be grouped in a unique focal element with probabilistic assignment equal to the sum. However the cardinality of the set of selectors will be decreased by the grouping. In (Miranda, Couso et al. 2002) the case of infinite coincident focal elements over a finite space S is considered; they are generated, through multi-valued mapping, by a probability measure on *X* infinite but "*atomic*": *SCT* is non-convex (it shows fractal properties) and does not coincide with $\Psi$. The coincidence requires, for finite *S*, conditions of continuity for the functions, satisfied if (but not only if) the measurable space on *X* is "*non-atomic*". More complex conditions are required when *S* is infinite (Miranda, Couso et al. 2003).

**N 3-2.** For example observe that, subdividing the first focal element ($\{s^1, s^2\}$,0.5) in two coincident focal elements and probabilistic assignments summing to the first ($\{s^1, s^2\}$,0.4), ($\{s^1, s^2\}$,0.1), the number of selectors increases to 12 (although not all different), in any case on the border, while the 4 extremes, and therefore the set $\Psi$ do not change.

**N 3-3.** Within the ambits of a *behavioral* interpretation of the probability assumed by Walley, the functions *f* can be considered as "*gambles*" and the "*prevision*" $E_{LOW}[f]$ the "*supremum buying price*" for the gamble *f*. Eq. (3.44) must be respected for "*avoiding sure loss*" (see Walley 1991), § 2.4.1 for definition of avoiding sure loss; § 3.8.5 for equivalence to the condition of non-empty set $\Psi$). When condition (3.44) is not respected, accepting a group, each one individually desirable, of gambles produces sure loss.

**N 3-4.** According to Walley "a probability model is incoherent if calculating the implications of the model would lead to its modification". In § 2.5 of his book, Walley gives a formal definition of coherence and also a general condition to uncertain a priori, using the original bounds, coherence of the available information. Of course coherence implies avoiding sure loss, i.e. Eq. (3.44), but not vice versa.

**N 3-5.** For a more extended list see (Walley, 1991; § 2.7.4). A simple example in § 2.7.5 shows that lower probabilities satisfying all listed necessary conditions would be not coherent or directly incur sure loss.

**N 3-6.** This follows from the fact that the infimum for the union of disjoint sets is equal to the sum of the probabilities of the two sets evaluated through a probability distribution in the set Ψ, not necessarily coinciding with the probability distributions giving separately the infima for the probability of the two sets.

**N 3-7.** This follows from the fact that the supremum for the union of disjoint sets is lower or equal to the sum of the probabilities of the two sets evaluated through a probability distribution in the set Ψ, not necessarily coinciding with the probability distributions giving separately the suprema for the probability of the two sets.

**N 3-8.** This condition appears in some cases stronger than the corresponding condition *Pro*(B)>0 when conditioning precise prior probabilities with respect to B. In fact it is more common for an event to have lower probability zero than precise probability zero. In his book, Walley presents an alternative model (coherent sets of *desirable gambles*) equivalent to lower previsions or sets of compatible distributions, but feasible to conditioning with respect to an event or likelihood function with zero lower prevision. See for example simple numerical examples in (Walley 2000).

# Chapter 4
# Random Relations

In this chapter, the notions introduced in Chapter 3 are extended to the case in which the uncertain information is assigned by means of marginal random sets on several different spaces, $S_i$, or by means of a random relation on the Cartesian product $S = \times S_i$. The multifold concept of independence is firstly introduced within the general framework of imprecise probabilities, and then specialized to random relations. A definition for correlation between variables constrained by random sets/relations is proposed.

## 4.1 Random Relations and Marginals

In Chapter 3, we focused our attention on uncertainty affecting one variable. We now want to extend the definitions of random set to multi-dimensional spaces, where several uncertain variables can be described. More precisely, let $S$ be the Cartesian product of sets $S_i$, $i$=1,…,$v$ where the $i$-th uncertain variable takes values on $S_i$. A *random relation* is a random set on $S = S_1 \times ... S_v$, i.e. it is a family of $n$ focal elements, $A^i \subseteq S$, and the basic probabilistic assignment, $m(A^i)$, that satisfies the conditions: $m(\varnothing)=0$; $\Sigma_i\, m(A^i)=1$. In the following, reference will be made to the two-dimensional case, its extension to the multidimensional case being straightforward. Figure 4.1a exemplifies a case in which the random relation is composed of three focal elements.

For any subset $T \subseteq S$, it is possible to evaluate the values of the set-valued functions $Bel(T)$ and $Pla(T)$ and the probability bounds on $T$ by using Eq. (3.3); therefore:

$$P(T) \in [\,Bel(T), Pla(T)] \tag{4.1}$$

For example, in Figure 4.1a: $P(T) \in [\,m(A^2), m(A^1) + m(A^2)]$.

Figure 4.1b illustrates that a focal set, $A^i$, projects onto the $s_1$-axis as the set (interval if $A^i$ is simply connected, multiple intervals if $A^i$ is not simply connected)

$$A_1^i = \{ s_1 \in S_1 \mid (s_1, s_2) \in A^i \text{ for some } s_2 \in S_2 \} \tag{4.2}$$

and onto the $s_2$-axis as the set

$$A_2^i = \{ s_2 \in S_2 \mid (s_1, s_2) \in A^i \text{ for some } s_1 \in S_1 \} \tag{4.3}$$

As in the case of random variables (Eqs. (2.23) and (2.24)), the marginal probability assignment, $m_j$, on the $s_j$-axis is defined by using the marginal (or additivity) rule (if projections of different focal sets coincide)

$$m_j\left(A_j^i\right) = \sum_{A':A'_j = A_j^i} m(A') \tag{4.4}$$

Random sets $\{(A_1^i, m_1^i)\}$ and $\{(A_2^i, m_2^i)\}$ are called *marginal random sets.*

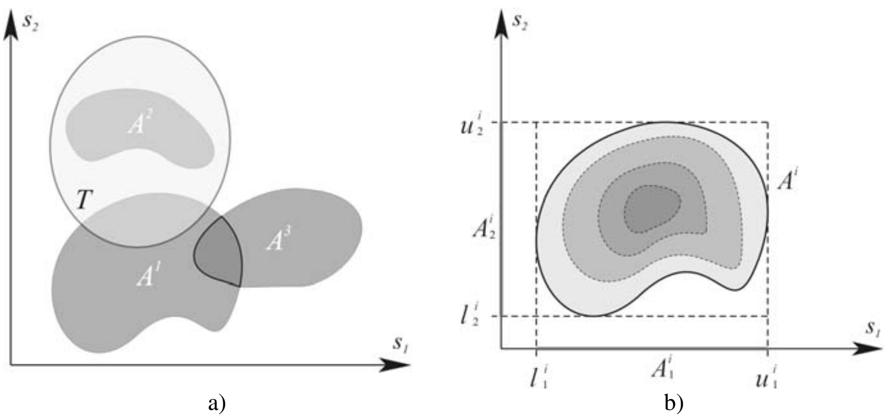



**Fig. 4.1** a) Random relation with three focal elements; b) marginals of focal element $A^i$ and contours of one possible probability distribution $P^i \in \Psi^i$

From Eq. (3.11), recall that a random set can be defined by using convex linear combinations of all probability measures defined over the focal elements (and equal to zero elsewhere). The coefficients of the linear combinations are fixed, and they are equal to the probability assignments ($m^i$ in Eq. (3.11)). Formally, let $P^i$ be a probability measure in the set of probability measures, $\Psi^i$, which are zero outside the focal set $A^i$. Figure 4.1b illustrates the contour of a distribution induced by $P^i \in \Psi^i$. A random relation is the set, $\Psi$, of probability measures, $P_{RS}$, obtained as convex combinations of $P^i$

$$\Psi = \left\{ P_{RS} : P_{RS} = \sum_{i=1}^{n} m\left(A^i\right) P^i \right\} \tag{4.5}$$

Let $\mathcal{F}_k = (\mathcal{A}_k, m_k)$ be a marginal random set, where the focal elements are $\mathcal{A}_k = \left\{ A_k^1, \ldots, A_k^{n_k} \right\}$, and the probability assignment is $m_k = \left\{ m_k^1, \ldots, m_k^{n_k} \right\}$ According to Eq. (4.5), the set of probability measures, $\Psi_k$, associated to the $k$-th marginal is

$$\Psi_k = \left\{ \sum_{i=1}^{n_k} \left( m_k\left(A_k^i\right) \cdot P_k^i \right) \mid P_k^i \in \Psi_k^i \right\}, \tag{4.6}$$

where $\Psi_k^i$ is the set of all probability measures defined over the focal element $A_k^i$ (and equal to zero elsewhere). Let us now investigate the relationship between the elements in $\Psi_1$ and those in $\Psi$ by taking the marginal of $P_{RS} \in \Psi$ onto $S_1$:

$$P_{RS}\left(\cdot \times S_2\right) = \sum_{i=1}^{n} m^i P^i \left(\cdot \times S_2\right) = \sum_{j=1}^{n_1} \sum_{i:A_1^i = A_1^j} m^i P^i \left(\cdot \times S_2\right) \tag{4.7}$$

By equating the last expression in Eq. (4.7) to Eq. (4.6), one has:

$$m_1^j P_1^j\left(\cdot\right) = \sum_{i:A_1^i = A_1^j} m^i P^i \left(\cdot \times S_2\right), \tag{4.8}$$

and by remembering Eq. (4.4), one obtains the final expression for an element of $\Psi_1^i$:

$$P_1^j\left(\cdot\right) = \frac{\sum_{i:A_1^i = A_1^j} m^i P^i \left(\cdot \times S_2\right)}{\sum_{i:A_1^i = A_1^j} m^i} \tag{4.9}$$

Eq. (4.9) can be interpreted as follows. Attach a mass equal to probabilistic assignment $m^i$ to each projection of a joint probability measure $P^i \in \Psi^i$ over a joint focal element $A^i$ whose projection is $A_1^j$. A marginal probability measure in $\Psi_1^j$ is the centroid of this system of masses, which will "resemble" the projection of joint probability measure(s) with the larger probabilistic assignment(s). On the other hand, since $\Psi^i$ contains all joint measures that are zero outside $A^i$, each probability measure in $\Psi_1^j$ can be generated by means of Eq. (4.9).

Similar to the one-dimensional case dealt with in Section 3.2.3.2 (page 35), a *selector* of a random relation $\{(A^i, m^i)\}$ is a random vector, $V = \{(v^i, m^i)\}$ (Section 2.4 page 23), whose values $v^i$ are included in the focal elements $A^i$. Call *SCT* the class of selectors; *marginal selectors* are marginals of $V\in$ *SCT*.

If the focal elements of a random relation $\{(A^i, m^i)\}$ can be ordered in a nested sequence, such that $A^i \subseteq A^{i+1}$, $i = 1, 2, ...n\text{-}1$, the random relation is termed *consonant*, and properties similar to the one-dimensional case hold (see Section 3.2.4). In particular, information given by the random relation is equivalent to the point-valued contour function, i.e. the possibility values, $\pi(s_1, s_2)$, of the singletons $\{(s_1, s_2)\}$, which is the membership function $\mu_F(s_1, s_2) = \pi(s_1, s_2)$ of a fuzzy relation $F$ with (see Eq. (3.24)):

$$\mu_F(s_1, s_2) = Pla(\{s_1, s_2\}) = \sum_{A^i:(s_1,s_2)\in A^i} m(A^i) \tag{4.10}$$

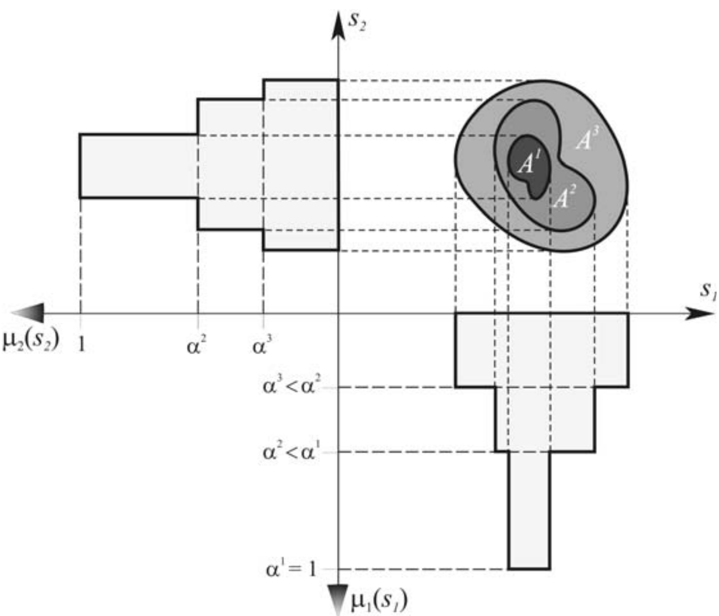



**Fig. 4.2** Consonant random relation with three focal elements and its marginal consonant random sets

The focal elements are the α-cuts of the fuzzy relation for the finite sequence $\alpha_1 = 1$, $\alpha_2 < \alpha_1$ , ..., $\alpha_{n+1} < \alpha_n$ , $\alpha_{n+1} = 0$, with probabilistic assignment $m({}^{\alpha_i}A) = \alpha_i - \alpha_{i+1}$ (Eq. (3.27)). Figure 4.2 illustrates a case in which $n$ =3.

For a consonant random relation, the marginals are consonant random sets that are fuzzy sets, $F_1$ and $F_2$, whose membership functions are simply defined by the following equations ($S_1$ and $S_2$ are finite sets; for infinite sets, the "sup" operator should be substituted for "max"):

$$\mu_1(s_1) = \max_{S_2} \mu(s_1, s_2)\,; \qquad \mu_2(s_2) = \max_{S_1} \mu(s_1, s_2) \tag{4.11}$$

i.e. they are the *projections* (Klir and Yuan 1995) of the fuzzy relation onto the space of the single variables, see Figure 4.2.

When all focal elements $A^i$ are nested Cartesian products, the random relation $\mathcal{F} = \{(A^i, m^i)\}$ is termed *consonant random Cartesian product* or *fuzzy Cartesian product* (Figure 4.3). Section 4.3.5 (page 174) deals with the case in which the marginals are given and the fuzzy Cartesian product is derived.

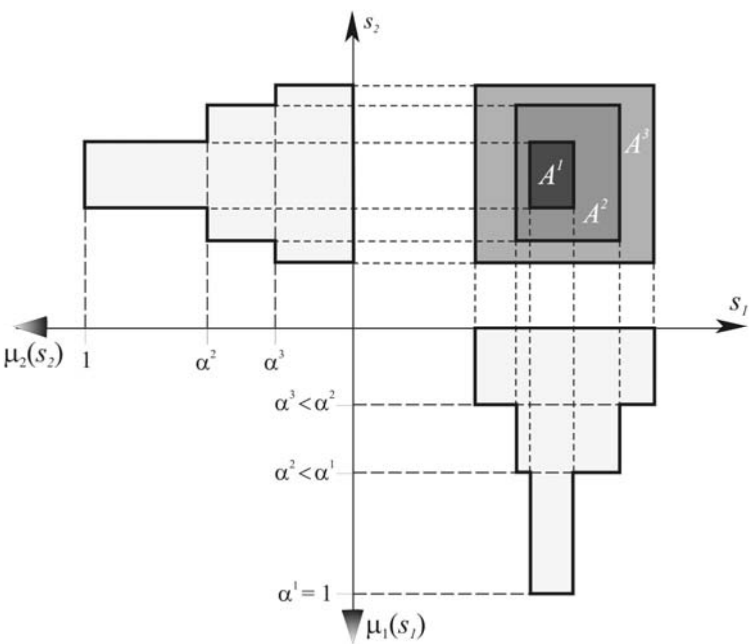



**Fig. 4.3** Consonant random Cartesian product with three focal elements and its marginal consonant random sets

Once a random relation is assigned, its marginal random sets are always uniquely determined by Eqs. (4.2) through (4.4). However, if only marginal random sets are given on $S_1$ and $S_2$, the available information does not uniquely define the information on the joint space $S = S_1 \times S_2$ for two distinct reasons:

a) Unless a rule is known or assumed *a priori*, the marginal focal elements do not uniquely determine the focal elements for the random relation. The Cartesian product is just an example of such a rule: $A^{ij} = A_1^i \times A_2^j$.
b) Per the additivity rule (Eq. (4.4)), a marginal focal element, say $A_1^i$, could be the projection of more than one focal element, say $A_1^i$ and $A_1^j$, among which it is thus necessary to apportion the marginal basic probabilistic assignment $m_1(A_1^i)$.

In the theory of precise probability, it is the second reason that brings about the indeterminateness of the joint probability distribution when only the marginal distributions are given. On the contrary, the first reason does not apply because the marginal focal elements are singletons (say $A_1^i = \left\{ {s_1}^i \right\};\ A_2^j = \left\{ {s_2}^j \right\}$), and thus the focal element is always uniquely determined ($A^{i,j} = \left\{ {s_1}^i, {s_2}^j \right\}$).

In order to understand the implications of combining two marginal random sets on a joint space, it is necessary to make an excursion into the wider context of imprecise probabilities, similar to the approach taken in Section 3.3. Section 4.2 explains that in the theory of imprecise probabilities the concept of independence is not unique. In Section 4.3, the first issue is solved by using the Cartesian product of marginal focal elements, and the concepts of independence in the theory of imprecise probabilities are used to overcome the second source of indeterminateness. In Section 4.4, the hypothesis of Cartesian product will be relaxed in the investigation of correlation.

## 4.2 Stochastic Independence in the Theory of Imprecise Probabilities

Let us recall two notions from the theory of precise probability:

a) Let $P(T^2)>0$. The conditional probability of $T^1$ conditioned on $T^2$ is defined as (Eqs. (2.4)):

$$P(T^1|T^2) := P(T^1 \cap T^2)/\, P(T^2) \tag{4.12}$$

b) Two events, $T^1$ and $T^2$, are said to be *independent* if (Eq. (2.6)):

$$P(T^1 \cap T^2) := P(T^1)\, P(T^2) \tag{4.13}$$

In the case of two marginal variables, let $P_i$ be the probability measure on the σ-algebra of $S_i$. Coherent with Eq. (4.13), the joint probability measure on the joint measurable space $(S, \mathcal{S})$ (Section 2.3) for independent variables is defined as the product measure $P = P_1 \otimes P_2 : \mathcal{C} = \{U_1 \times U_2 : U_i \in \mathcal{S}_i\} \to [0,1]$ given by (Eqs. (2.31)-(2.35)):

$$P_1 \otimes P_2\left(U_1 \times U_2\right) := P_1\left(U_1\right) P_2\left(U_2\right) \tag{4.14}$$

and $P$ can be extended to any subset in the σ-algebra $\mathcal{S}$ generated by $\mathcal{C}$.

Eq. (4.14) and the definition of conditional probability (4.12) establish that the conditional probability measures for independent variables yield the marginal probability measures

$$\begin{aligned} &P\left(\cdot \times S_2 \mid S_1 \times \{s_2\}\right) = P\left(\cdot \times \{s_2\}\right) / P\left(S_1 \times \{s_2\}\right) = \\ &= P_1(\cdot) \cdot P_2\left(\{s_2\}\right) / \left(P_1\left(S_1\right) \cdot P_2\left(\{s_2\}\right)\right) = P_1(\cdot) \quad \forall s_2 : P_2\left(\{s_2\}\right) > 0; \\ &P\left(S_1 \times \cdot \mid \{s_1\} \times S_2\right) = P_2(\cdot) \;\; \forall s_1 : P_1\left(\{s_1\}\right) > 0 \end{aligned} \tag{4.15}$$

This means that: if we learn that the actual value of the second variable is $s_2$, then our knowledge about the probability measure for the first variable does not change. Likewise for the second variable. Let us now consider the extension to the case in which a generic convex set of probability measures, $\Psi_i$, is assigned to the $i$-th variable, i.e. $\Psi_i$ does not have to satisfy Eq. (4.6). Definitions in this section are taken from (Walley 1991; Couso, Moral et al. 1999; Vicig 1999; Ferson, Nelsen et al. 2004).

To exemplify, if sets $S_i$ are finite, as in Section 2.3, let $P_i$ be a probability distribution on $S_i = \{ s_i^j : j = 1,\ldots, n_i\}$, and let $\mathbf{p}_i$ be an $n_i$–column vector whose $j$-th component is $P_i( s_i^j )$. Let $f_i^k : S_i \rightarrow \mathbb{R}$, $k= 1,\ldots,k_i$ be a set of bounded functions on $S_i$ (gambles according to Walley's nomenclature), whose expectations (previsions according to Walley's nomenclature) are (Section 3.3.2 on page 62 and Eq. (3.39)):

$$E\left[ f_i^k \right] = \sum_{j=1}^{n_i} f_i^k\left( s_i^j \right) P_i\left( s_i^j \right) = \left( \mathbf{f}_i^k \right)^{\mathrm{T}} \mathbf{p}_i \tag{4.16}$$

where $\mathbf{f}_i^k$ is an $n_i$–column vector whose $j$-th entry is $f_i^k( s_i^j )$. Set $\Psi_i$ may be assigned as the set of distributions $\mathbf{p}_i$ bounded by hyperplanes (4.16) (Eq. (3.41)):

$$\Psi_i^E = \left\{ \mathbf{p}_i : E_{LOW}\left[ f_i^k \right] \leq \left( \mathbf{f}_i^k \right)^{\mathrm{T}} \mathbf{p}_i \leq E_{UPP}\left[ f_i^k \right] \right\} \tag{4.17}$$

In Eq. (4.17), if no constraint is assigned to the lower or upper expectation bound, then $E_{LOW}\left[ f_i^k \right]$ and $E_{UPP}\left[ f_i^k \right]$ are assigned values equal to $-\infty$ and $+\infty$, respectively.

Alternatively, and in an equivalent way, let $ETX_i$ (of cardinality $\xi_i = |ETX_i|$) indicate the set of extreme distributions (vertices) of $\Psi_i$, $\mathbf{p}_{EXT_i}^{\xi}$, $\xi = 1,\ldots, \xi_i$. $\Psi_i$ is the set of convex combinations of $\mathbf{p}_{EXT_i}^{\xi}$:

$$\Psi_i = \left\{ \mathbf{p}_i : \mathbf{p}_i = \sum_{\xi=1}^{\xi_i} c_i^{\xi} \mathbf{p}_{EXT_i}^{\xi}, \sum_{\xi=1}^{\xi_i} c_i^{\xi} = 1,\ c_i^{\xi} \geq 0 \right\} \tag{4.18}$$

If sets $\Psi_i^E$ and $\Psi_i$ are empty (see Section 3.3.2.1), then the set of joint probability measures is also empty. In the following, it is therefore assumed that sets $\Psi_i^E$ and $\Psi_i$ are not empty. If the bounds in (4.17) are not coherent (Section 3.3.2.2), all optimization problems that follow will yield coherent solutions because the solutions $\mathbf{p}_i$ will be in $\Psi_i^E$, and thus the solutions will satisfy the following coherent bounds (Eq. (3.50)):

$$E_{LOW,c}\left[f_i^k\right]=\min_{\mathbf{p}\in\Psi_i^E} E_{\mathbf{p}}\left[f_i^k\right]; \qquad E_{UPP,c}\left[f_i^k\right]=\max_{\mathbf{p}\in\Psi_i^E} E_{\mathbf{p}}\left[f_i^k\right] \tag{4.19}$$

This will be exemplified in Example 4.5. In order to unclutter the notation, in the following the superscript "$E$" for "natural extension" will be dropped from $\Psi_i^E$.

In finite spaces, let $P$ be a probability of the joint elementary events on $S$, and let $\mathbf{P}$ be an $n_1 \times n_2$ matrix with $(i, j)$-th entry $p^{i,j} := P\left(s_1^i, s_2^j\right)$. Consider now a linear function of the probability mass: $\sum_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j}$ ; the probability of an event (subset) $T \subseteq S$ is obtained by setting $a^{i,j} = 1$ if $\left(s_1^i, s_2^j\right) \in T$ , $a^{i,j} = 0$ otherwise. Likewise, the expectation of a function $g$ on $S$ (prevision of gamble $g$, according to Walley's nomenclature) is obtained by setting $a^{i,j} = g\left(s_1^i, s_2^j\right)$. When expectation bounds are given on the marginals, constraints (4.17) may be expressed in terms of the joint probability $\mathbf{P}$ by using Eq. (2.24), i.e.:

$$\mathbf{p}_1 = \mathbf{P} \cdot \mathbf{1}_{(n_2)}; \quad \mathbf{p}_2 = \mathbf{P}^{\mathrm{T}} \cdot \mathbf{1}_{(n_1)}, \tag{4.20}$$

where $\mathbf{1}_{(n_i)}$ is a column vector of unit components of length $n_i$:

$$E_{LOW}\left[f_i^k\right] \le \left(\mathbf{f}_i^k\right)^{\mathrm{T}} \mathbf{P} \cdot \mathbf{1}_{(n_j)} \le E_{UPP}\left[f_i^k\right];\ k = 1,\ldots,k_i;\ i,j = (1,2),(2,1) \tag{4.21}$$

In the following sections, we will be concerned with the problem of finding the maximum or minimum of a function on the joint distribution (e.g., finding bounds on the probability or on the expectation of an event $T \subseteq S$) subject to constraints, e.g., of the kind in Eq. (4.21). There are two options to find the extreme values of functions on the joint distribution:

1) Using global optimization to find the point(s) at which the minimum or maximum value of the objective function is achieved. This option directly focuses on the optimal solution, which is apparently more efficient if only one maximum or minimum must be calculated.

2) First finding all extreme points for $\Psi$ (the set of joint probability measures/distributions that satisfies the chosen definition of independence), and then restricting the search to the extreme points of $\Psi$. This second option can be used only if the objective function and constraints are linear (e.g., probability of an event or expectation of a function). Theorem 4.1 below ensures that even the maximum and the minimum of conditional probabilities are achieved at the extreme points of $\Psi$.

**Theorem 4.1.** *The minimum and maximum values of a conditional probability in the joint distribution are achieved at the extreme points of the convex set of the joint distributions* $\Psi$.

*Proof:* By inserting the marginal expression for $p_2\left(s_2^j\right)$ (Eq. (2.24)) into the expression for the conditional probability $p_{1|2}\left(s_1^i \mid s_2^j\right)$ (Eq. (2.25)), one obtains:

$$\forall\left(s_1^i, s_2^j\right) \in S: \qquad p_{1|2}\left(s_1^i \mid s_2^j\right) = \frac{p^{i,j}}{\sum_{i=1}^{n_1} p^{i,j}} \tag{4.22}$$

Let $\mathbf{P}_* \in \Psi$ and $\mathbf{P}_{**} \in \Psi$. The conditional probability $p_{1|2}(s_1^i | s_2^j)$ on each joint distribution is $p_{1|2,*}\left(s_1^i \mid s_2^j\right) = p_*^{i,j} / \sum_{i=1}^{n_1} p_*^{i,j}$, and $p_{1|2,**}\left(s_1^i \mid s_2^j\right) = p_{**}^{i,j} / \sum_{i=1}^{n_1} p_{**}^{i,j}$, respectively. Assume $p_{1|2,*}\left(s_1^i \mid s_2^j\right) \geq p_{1|2,**}\left(s_1^i \mid s_2^j\right)$, then any interior point $\mathbf{P}_{(new)}$ between $\mathbf{P}_*$ and $\mathbf{P}_{**}$ may be written as $\lambda \mathbf{P}_* + (1-\lambda)\mathbf{P}_{**}$, $0 \leq \lambda \leq 1$, i.e., $p_{(new)}^{i,j} = \lambda p_*^{i,j} + (1-\lambda) p_{**}^{i,j}$. Consequently, the conditional probability based on the new joint distribution $\mathbf{P}_{(new)}$ is:

$$p_{1|2,new}\left(s_1^i \mid s_2^j\right) = \frac{p_{(new)}^{i,j}}{\sum_{i=1}^{n_1} p_{(new)}^{i,j}} = \frac{\lambda p_*^{i,j} + (1-\lambda) p_{**}^{i,j}}{\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}} \tag{4.23}$$

By subtracting Eq. (4.23) from (4.22), one obtains:

$$\begin{aligned}
& p_{1|2,new}\left(s_1^i \mid s_2^j\right) - p_{1|2,*}\left(s_1^i \mid s_2^j\right) = \frac{\lambda p_*^{i,j} + (1-\lambda) p_{**}^{i,j}}{\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}} - \frac{p_*^{i,j}}{\sum_{i=1}^{n_1} p_*^{i,j}} \\
&= \frac{\left[\lambda p_*^{i,j} + (1-\lambda) p_{**}^{i,j}\right] \sum_{i=1}^{n_1} p_*^{i,j} - \left[\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}\right] p_*^{i,j}}{\left[\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}\right] \sum_{i=1}^{n_1} p_*^{i,j}} \\
&= \frac{(1-\lambda)\left(p_{**}^{i,j} \sum_{i=1}^{n_1} p_*^{i,j} - p_*^{i,j} \sum_{i=1}^{n_1} p_{**}^{i,j}\right)}{\left[\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}\right] \sum_{i=1}^{n_1} p_*^{i,j}} \\
&= \frac{(1-\lambda)\left(p_{**}^{i,j} \sum_{i=1}^{n_1} p_*^{i,j} - p_*^{i,j} \sum_{i=1}^{n_1} p_{**}^{i,j}\right) / \left(\sum_{i=1}^{n_1} p_*^{i,j} \cdot \sum_{i=1}^{n_1} p_{**}^{i,j}\right)}{\left[\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}\right] \sum_{i=1}^{n_1} p_*^{i,j} / \left(\sum_{i=1}^{n_1} p_*^{i,j} \cdot \sum_{i=1}^{n_1} p_{**}^{i,j}\right)} \\
&= \frac{1-\lambda}{\left[\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}\right] / \sum_{i=1}^{n_1} p_{**}^{i,j}} \left(p_{**}^{i,j} / \sum_{i=1}^{n_1} p_{**}^{i,j} - p_*^{i,j} / \sum_{i=1}^{n_1} p_*^{i,j}\right) \\
&= A \cdot \left[p_{1|2,**}\left(s_1^i \mid s_2^j\right) - p_{1|2,*}\left(s_1^i \mid s_2^j\right)\right],
\end{aligned} \tag{4.24}$$

where $A = \dfrac{(1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}}{\lambda \sum_{i=1}^{n_1} p_*^{i,j} + (1-\lambda) \sum_{i=1}^{n_1} p_{**}^{i,j}} \geq 0$.

Since $p_{1|2,*}\left(s_1^i \mid s_2^j\right) \geq p_{1|2,**}\left(s_1^i \mid s_2^j\right)$, $p_{1|2,new}\left(s_1^i \mid s_2^j\right) - p_{1|2,*}\left(s_1^i \mid s_2^j\right) \leq 0$, i.e. $p_{1|2,*}(s_1^i|s_2^j) \geq p_{1|2,new}(s_1^i|s_2^j)$. Likewise, $p_{1|2,new}(s_1^i|s_2^j). \geq p_{1|2,**}(s_1^i|s_2^j)$

In conclusion, given two extreme points on the convex set of joint distribution, $\mathbf{P}_*$ and $\mathbf{P}_{**}$, and $p_{1|2,*}(s_1^i|s_2^j) \geq p_{1|2,**}(s_1^i|s_2^j)$, any interior point $\mathbf{P}_{(new)}$ between them satisfies the inequality $p_{1|2,*}(s_1^i|s_2^j) \geq p_{1|2,\mathrm{new}}(s_1^i|s_2^j) \geq$

$p_{1|2,**}(s_1{}^i|s_2{}^j)$. The minimum and maximum values of conditional probability are thus achieved at the extreme points of the convex set of joint distributions.

Therefore, regardless of the type of independence introduced next, the conditional upper and lower probabilities are reached at an extreme point of $\Psi$. ■

When constraints are linear and the second option is used, the general algorithm for finding the extreme points of $\Psi$ is based on the interpretation of $\Psi$ as a polytope in the $s$-dimensional space of the singleton's probabilities ($s = n_1$ x $n_2$ for joint probabilities) given by the intersection of half-spaces, whose equations are the prevision bounds in Eq. (4.17). The algorithm is as follows (modified after Walley 1991, page 511):

1. In the $s$-dimensional space, $\Psi$ is bounded by $n$ linear inequalities (e.g., Eq. (4.21)) and $s$ nonnegative constraints on singletons $s^i$, $P(s^i) \geq 0$. Consider ($s$-1) constraints at a time in addition to the constraint that the sum of probabilities of the singletons is 1.
2. Write the ($s$-1) inequality constraints as equalities. If the system of equations is singular, then either there is no solution (e.g., upper and lower bounds on the same prevision have been used) or there are infinite solutions (e.g., an entire face of the polytope $\Psi$); in either case, go back to point 1. If the system is not singular, compute the unique solution **P**.
3. If **P** satisfies all the remaining $(n+s+1) - \text{s} = n + 1$ constraints, **P** is an extreme point of the joint distribution set $\Psi$, otherwise it is not.
4. Repeat Step 1 to Step 3 until all combinations of constraints are considered.

### *4.2.1 Unknown Interaction*

If nothing is known about dependence or independence between $s_1$ and $s_2$, and if it is not known which probability measure in $\Psi_1$ and $\Psi_2$ must be combined, then unknown interaction should be used. The set, $\Psi_U$, of probability measures on a sigma algebra of $S$ is the set of all joint probability measures, $P$, that respect the marginal rule (2.23), i.e. whose marginals are in $\Psi_1$ and $\Psi_2$, i.e.:

$$P(\cdot \times S_2) \in \Psi_1 \ \ ; \ P(S_1 \times \cdot) \in \Psi_2 \tag{4.25}$$

In finite spaces, upper and lower bounds for any linear function of the probability masses $\sum_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j}$ (and thus upper and lower probabilities of events or upper and lower expectations) are determined by solving two linear optimization problems in the $p^{i,j}$. Indeed, when expectation bounds are given on the marginals, constraints (4.25) are linear constraints obtained by expressing the marginals in Eq. (4.17) in terms of the joint probability **P** by using Eq. (4.20) as in Eq. (4.21). The complete optimization problem in the $n_1 \times n_2$ components $p^{i,j}$ reads:

$$\begin{aligned}
&\text{minimize } \sum_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j} \quad \left(-\sum_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j}\right)\\
&\textit{subject to}\\
&E_{LOW}\left[f_i^k\right] \le \left(\mathbf{f}_i^k\right)^{\mathrm{T}} \mathbf{P}\cdot\mathbf{1}_{(n_i)} \le E_{UPP}\left[f_i^k\right];\ k=1,\dots,k_i;\ i=1,2\\
&\mathbf{1}_{(n_1)}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{1}_{(n_2)} = 1\\
&p^{i,j} \ge 0;\ i=1,\dots,n_1;\ j=1,\dots,n_2
\end{aligned} \tag{4.26}$$

Besides solving the optimization problem in Eq. (4.26), one may first use the algorithm on page 114 to calculate the extreme points of Ψ defined by the constraints in (4.26), and then calculate the objective function on the extreme points. Example 4.1 exemplifies both ways of finding upper and lower previsions; in particular, Table 4.2 lists all extreme points of Ψ for that specific problem.

When extreme distributions are given on the marginals, constraints (4.25) are linear constraints obtained by expressing the marginals in Eq. (4.18) in terms of the joint probability **P** by using Eq. (4.20), i.e.:

$$\begin{aligned}
\mathbf{p}_1 &= \mathbf{P}\cdot\mathbf{1}_{(n_2)} = \sum_{\xi=1}^{\xi_1} c_1^{\xi}\mathbf{p}_{EXT_1}^{\xi}\\
\mathbf{p}_2 &= \mathbf{P}^{\mathrm{T}}\cdot\mathbf{1}_{(n_1)} = \sum_{\xi=1}^{\xi_2} c_2^{\xi}\mathbf{p}_{EXT_2}^{\xi}\\
&\sum_{\xi=1}^{\xi_i} c_i^{\xi} = 1,\ c_i^{\xi} \ge 0;\ i=1,2
\end{aligned} \tag{4.27}$$

The complete optimization problem in the $n_1 \times n_2$ components $p^{i,j}$ and in the $\xi_1+\xi_2$ components $c_i^{\xi}$ then reads:

$$\text{minimize } \sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j} \quad \left(-\sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j}\right)$$

*subject to*

$$\begin{aligned}
&\mathbf{P}\cdot\mathbf{1}_{(n_2)} - \sum_{\xi=1}^{\xi_1} c_1^{\xi}\mathbf{p}_{EXT_1}^{\xi} = 0\\
&\mathbf{P}^{\mathrm{T}}\cdot\mathbf{1}_{(n_1)} - \sum_{\xi=1}^{\xi_2} c_2^{\xi}\mathbf{p}_{EXT_2}^{\xi} = 0\\
&\sum_{\xi=1}^{\xi_i} c_i^{\xi} = 1,\ c_i^{\xi} \geq 0;\ i = 1,2\\
&p^{i,j} \geq 0;\ i = 1,\ldots,n_1;\ j = 1,\ldots,n_2
\end{aligned} \tag{4.28}$$

In the optimization problem (4.28), it is not necessary to add the constraint $\mathbf{1}_{(n_1)}{}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{1}_{(n_2)} = 1$ because, if the first and third constraints are satisfied, then by pre-multiplying by $\mathbf{1}_{(n_1)}{}^{\mathrm{T}}$ one obtains:

$$\begin{aligned}
\mathbf{1}_{(n_1)}{}^{\mathrm{T}}\mathbf{P}\cdot\mathbf{1}_{(n_2)} &= \sum_{\xi=1}^{\xi_1} c_1^{\xi}\mathbf{1}_{(n_1)}{}^{\mathrm{T}}\mathbf{p}_{EXT_1}^{\xi}\\
&= \sum_{\xi=1}^{\xi_1} c_1^{\xi} = 1
\end{aligned} \tag{4.29}$$

**Example 4.1.** Consider the case in which a two-component resin has to be applied at a construction site to anchor steel bars. Cartridges of resins A and B should be contained in two different 10-cartridge boxes. Unfortunately, the manufacturer mixed up the boxes' contents, and only some of the cartridges were counted when the boxes were opened in the field. Box 1 contains 5 A's, 2 B's, and 3 unknown component cartridges; Box 2 contains 3 A's, 3 B's, and 4 unknown component cartridges. A worker in the field takes one cartridge from Box 1 and then one cartridge from Box 2; we are interested in the joint probability of the selected cartridges. As we are just given the marginal probabilities, all we know is that the joint probability measure must satisfy Eq. (4.25). One cartridge is selected from each of the boxes, but we cannot assume stochastic independence, and it is possible that a correlated joint procedure is used to select the two cartridges. For example, it may be that cartridges are numbered from 1 to 10, and when the worker picks the *i*-th cartridge from either box, he picks the *i*-th cartridge from the other box. In the present case, "independence" is just our lack of information about the interaction between the two selections.

As depicted in Figure 4.4, $S_i = \{A, B\}$, $S = \{(A, A), (A, B), (B, A), (B, B)\}$; the marginal probabilities of elementary events are listed in Table 4.1 (where gray hatches indicate the extreme points of the sets $\Psi_i$, Section 2.2.3.4), and are depicted in Figure 4.5. Under the unknown interaction assumption, $\Psi_U$ contains all joint probabilities of elementary events on $S$ whose marginals are in $\Psi_1$ and $\Psi_2$.

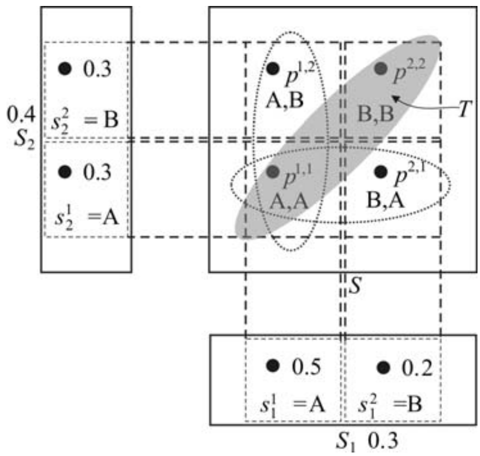
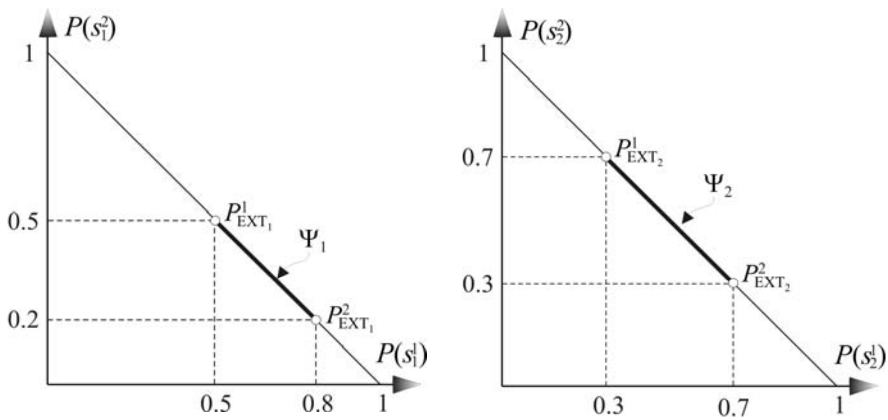



**Fig. 4.4** Example 4.1: marginal distributions and joint events. In marginal distributions, numbers next to each event indicate minimum probability masses, and numbers outside the box indicate free probability mass that gives rise to several possible marginal distributions in $\Psi_i$ (Table 4.1). Dashed lines indicate the joint events whose joint probabilities are constrained by the first two constraints of Eq. (4.31). Hatching indicates event $T$={(A, A), (B, B)}




a) b)

**Fig. 4.5** Example 4.1: marginal distributions and sets $\Psi_1$(a) and $\Psi_2$ (b) (Table 4.1)

**Table 4.1** Example 4.1: Probabilities of elementary events for: (a) Box 1; and (b) Box 2

(a)

| $p_1^j \in \Psi_1$ | A | B |
|---|---|---|
| $p_1^1$ | 0.5 | 0.5 |
| $p_1^2$ | 0.6 | 0.4 |
| $p_1^3$ | 0.7 | 0.3 |
| $p_1^4$ | 0.8 | 0.2 |

(b)

| $p_2^j \in \Psi_2$ | A | B |
|---|---|---|
| $p_2^1$ | 0.3 | 0.7 |
| $p_2^2$ | 0.4 | 0.6 |
| $p_2^3$ | 0.5 | 0.5 |
| $p_2^4$ | 0.6 | 0.4 |
| $p_2^5$ | 0.7 | 0.3 |

The two linear optimization problems (4.28) read as follows:

$$\begin{aligned}\text{minimize } & a^{1,1}p^{1,1}+a^{1,2}p^{1,2}+a^{2,1}p^{2,1}+a^{2,2}p^{2,2}\\ & -\left(a^{1,1}p^{1,1}+a^{1,2}p^{1,2}+a^{2,1}p^{2,1}+a^{2,2}p^{2,2}\right)\end{aligned}$$

*subject to*

$$\begin{aligned}&p^{1,1}+p^{1,2}-0.5c_1^1-0.8c_1^2=0\\&p^{2,1}+p^{2,2}-0.5c_1^1-0.2c_1^2=0\\&p^{1,1}+p^{2,1}-0.3c_2^1-0.7c_2^2=0\\&p^{1,2}+p^{2,2}-0.7c_2^1-0.3c_2^2=0\\&c_1^1+c_1^2=1\\&c_2^1+c_2^2=1\\&c_i^\xi\geq 0\\&p^{i,j}\geq 0\end{aligned}, \tag{4.30}$$

The attentive reader may have noticed that the assigned information on the marginals in fact corresponds to two random sets: {({A}, 0.5), ({B}, 0.2), ({A, B}, 0.3)} on $S_1$, and {({A}, 0.3), ({B}, 0.3), ({A, B}, 0.4)} on $S_2$. From Section 3.2.2, page 67, and Section 3.3.3:

- Since random sets are Choquet capacities of ∞-order, they constrain probabilities of events (i.e. they do not constrain expectations, or previsions), and therefore their sets $\Psi_i$ are bounded by hyper-planes with normals whose components are either 0 or 1.

- When $|S_i|$ is either 2 or 3, assigning bounds to the probability of events yields the same $\Psi_i$ as assigning appropriate bounds to the probability of the singletons, i.e. the hyper-planes bounding $\Psi_i$ are parallel to the coordinate axes.

Since in this example $|S|$=2, constraints in Eq. (4.26) can be written in terms of probabilities of the singletons (of the marginals). In fact, $\left(\mathbf{f}_1^1\right)^{\mathrm{T}}=(1,0)$ and $\left(\mathbf{f}_2^1\right)^{\mathrm{T}}=(0,1)$ so that:

$$
\begin{aligned}
&\text{minimize } a^{1,1}p^{1,1}+a^{1,2}p^{1,2}+a^{2,1}p^{2,1}+a^{2,2}p^{2,2}\\
&\qquad\qquad -\left(a^{1,1}p^{1,1}+a^{1,2}p^{1,2}+a^{2,1}p^{2,1}+a^{2,2}p^{2,2}\right)\\
&\textit{subject to}\\
&0.5\le p^{1,1}+p^{1,2}\le 0.8\\
&0.3\le p^{1,1}+p^{2,1}\le 0.7\\
&p^{1,1}+p^{1,2}+p^{2,1}+p^{2,2}=1\\
&p^{i,j}\ge 0
\end{aligned}
\tag{4.31}
$$

Consider the case in which the resin is not activated because the same resin type is selected: $T$={(A, A), (B, B)} (hatched in Figure 4.4), and $a^{1,1}=a^{2,2}=1$; $a^{1,2}=a^{2,1}=0$. In this simple example, problems (4.31) can be solved “by hand”, and they have multiple solutions. For example, the minimizing solutions can be found by assigning zero to the objective function, which entails that both $p$(A, A)=0 and $p$(B, B)=0. The other constraints are satisfied, for example, for $p$(A, B)=$p$(B, A)=0.5; for $p$(A, B)= 0.6 and $p$(B, A)=0.4; and for $p$(A, B)= 0.7 and $p$(B, A)=0.3. All of these solutions yield $P_{low}(T)=0$. The maximizing solutions that yield $P_{upp}(T)=1$ are more numerous. One set of solutions is obtained by taking $p$(A, B)= $p$(B, A)=0, and then $p$(A, A)=0.5 and $p$(B, B)=0.5; $p$(A, A)=0.6 and $p$(B, B)=0.4; $p$(A, A)=0.7 and $p$(B, B)=0.3.

Under the unknown interaction assumption, the probability that the resin is not activated is vacuous, i.e. it is in the [0, 1] range. The concepts of independence introduced in the following sub-sections will provide narrower probability intervals. Notice that $\Psi_{\mathrm{U}}$ is larger than the set of product probabilities because, for example, $P_{low}(T)=0<0.15=P_{1,low}(\{A\})P_{2,low}(\{A\})=p_1^1(A)\,p_2^1(A)$. The unknown interaction model violates the factorization condition (4.14) because condition (4.14) requires knowledge in addition to that available.

Another option is to find all extreme distributions on the joint space using the algorithm on page 114, and then check all the extreme distributions to find the extreme value of the objective function. All 12 extreme distributions generated by constraints (4.31) are listed in Table 4.2, and the reader can easily check that the results are the same as those obtained by solving problem (4.31).

**Table 4.2** Example 4.1: Extreme joint probability distributions on $S$

| $p^{i}_{EXT} \in \Psi_U$ | $p^{A,A}_{EXT}$ | $p^{A,B}_{EXT}$ | $p^{B,A}_{EXT}$ | $p^{B,B}_{EXT}$ | $p(T)=p(A, A)+p(B,B)$ |
|---|---|---|---|---|---|
| $p^{1}_{EXT}$ | 0.5 | 0 | 0 | 0.5 | 1 |
| $p^{2}_{EXT}$ | 0 | 0.5 | 0.5 | 0 | 0 |
| $p^{3}_{EXT}$ | 0 | 0.7 | 0.3 | 0 | 0 |
| $p^{4}_{EXT}$ | 0.7 | 0 | 0 | 0.3 | 1 |
| $p^{5}_{EXT}$ | 0 | 0.5 | 0.3 | 0.2 | 0.2 |
| $p^{6}_{EXT}$ | 0.3 | 0.2 | 0 | 0.5 | 0.8 |
| $p^{7}_{EXT}$ | 0.3 | 0.5 | 0 | 0.2 | 0.5 |
| $p^{8}_{EXT}$ | 0.1 | 0.7 | 0.2 | 0 | 0.1 |
| $p^{9}_{EXT}$ | 0.5 | 0 | 0.2 | 0.3 | 0.8 |
| $p^{10}_{EXT}$ | 0.2 | 0.3 | 0.5 | 0 | 0.2 |
| $p^{11}_{EXT}$ | 0.7 | 0.1 | 0 | 0.2 | 0.9 |
| $p^{12}_{EXT}$ | 0.5 | 0.3 | 0.2 | 0 | 0.5 |

**Example 4.2.** In order to appreciate the difference between marginals assigned as random sets and marginals assigned as a more general set of probability measures, let us move to a three-dimensional space. In particular, let us now consider the two random sets in Example 3-6 and Example 3-11, i.e.: in $S_1$: $\{(\{s_1{}^1, s_1{}^2\}, 0.1), (\{s_1{}^2\}, 0.6), (\{s_1{}^3\}, 0.3)$, whose set $\Psi_1$ is depicted in Figure 4.6a, and has two vertices identified by vectors: $\mathbf{p}^{1}_{EXT_1} = (0, 0.7, 0.3)^T$ and $\mathbf{p}^{2}_{EXT_1} = (0.1, 0.6, 0.3)^T$. In $S_2$: $\{(\{s_2{}^1, s_2{}^2), 0.1), (\{s_2{}^1, s_2{}^2, s_2{}^3\}, 0.1), (\{s_2{}^2\}, 0.8)\}$, whose set $\Psi_2$ is depicted in Figure 4.6b, and has four vertices identified by vectors $\mathbf{p}^{1}_{EXT_2} = (0, 0.9, 0.1)^T$, $\mathbf{p}^{2}_{EXT_2} = (0, 1, 0)^T$, $\mathbf{p}^{3}_{EXT_2} = (0.1, 0.8, 0.1)^T$, and $\mathbf{p}^{4}_{EXT_2} = (0.2, 0.8, 0)^T$.

Since the marginals are random sets with $|S_i|=3$, in Figure 4.6 sets $\Psi_i$ are bounded by planes parallel to the coordinate planes, i.e. $\Psi_i$ is uniquely defined by bounds on the probabilities of the singletons (Section 4.3.2). As a consequence, $k_i = n_i = 3$, and Table 4.3 gives the upper and lower probabilities calculated as Belief and Plausibility of the singletons, respectively; these values coincide with the bounds that can be calculated with the extreme joint distributions.

Constraints (4.26) are thus equal to the following constraints:

$$
\begin{aligned}
&\textit{Subject to}\\
&0.0 \le p^{1,1} + p^{1,2} + p^{1,3} \le 0.1\\
&0.6 \le p^{2,1} + p^{2,2} + p^{2,3} \le 0.7\\
&0.3 \le p^{3,1} + p^{3,2} + p^{3,3} \le 0.3\\
&0.0 \le p^{1,1} + p^{2,1} + p^{3,1} \le 0.2\\
&0.8 \le p^{1,2} + p^{2,2} + p^{3,2} \le 1.0\\
&0.0 \le p^{1,3} + p^{2,3} + p^{3,3} \le 0.1\\
&p^{1,1} + p^{1,2} + p^{1,3} + p^{2,1} + p^{2,2} + p^{2,3} + p^{3,1} + p^{3,2} + p^{3,3} = 1\\
&p^{i,j} \ge 0
\end{aligned}
\tag{4.32}
$$

On the other hand, the constraints in problems (4.28) read

$$
\begin{aligned}
&\textit{Subject to}\\
&p^{1,1} + p^{1,2} + p^{1,3} - 0.0c_1^1 - 0.1c_1^2 = 0\\
&p^{2,1} + p^{2,2} + p^{2,3} - 0.7c_1^1 - 0.6c_1^2 = 0\\
&p^{3,1} + p^{3,2} + p^{3,3} - 0.3c_1^1 - 0.3c_1^2 = 0\\
&p^{1,1} + p^{2,1} + p^{3,1} - 0.0c_2^1 - 0.0c_2^2 - 0.1c_2^3 - 0.2c_2^4 = 0\\
&p^{1,2} + p^{2,2} + p^{3,2} - 0.9c_2^1 - 1.0c_2^2 - 0.8c_2^3 - 0.8c_2^4 = 0\\
&p^{1,3} + p^{2,3} + p^{3,3} - 0.1c_2^1 - 0.0c_2^2 - 0.1c_2^3 - 0.0c_2^4 = 0\\
&c_1^1 + c_1^2 = 1\\
&c_2^1 + c_2^2 + c_2^3 + c_2^4 = 1\\
&c_i^\xi \ge 0\\
&p^{i,j} \ge 0
\end{aligned}
\tag{4.33}
$$

Let us now consider the event $T = \left\{ \left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right) \right\}$, i.e. $a^{i,j} = 1$ if $i$=$j$, 0 otherwise. The lower and upper probabilities for $T$, which are equal to 0.4 and 0.8, respectively, are obtained at multiple solution points, some of which are given in Table 4.4. When the interior point method is used (Mehrotra 1992), these solutions are not necessarily extreme points in either $\Psi_1$ or $\Psi_2$.

**Table 4.3** Example 4.2: Upper and lower probabilities of the singletons

| Marginal $\boldsymbol{S_i}$ | Singleton $s_i^j$ | $\left(\mathbf{f}_i^j\right)^{\mathrm{T}}$ | $E_{LOW}\left[f_i^j\right]=Bel\left(s_i^j\right)$ | $E_{UPP}\left[f_i^j\right]=Pla\left(s_i^j\right)$ |
|---|---|---|---|---|
| $S_1$ | $s_1^1$ | (1, 0, 0) | 0.0 | 0.1 |
| | $s_1^2$ | (0, 1, 0) | 0.6 | 0.6+0.1=0.7 |
| | $s_1^3$ | (0, 0, 1) | 0.3 | 0.3 |
| $S_2$ | $s_2^1$ | (1, 0, 0) | 0.0 | 0.2 |
| | $s_2^2$ | (0, 1, 0) | 0.8 | 0.8+0.1+0.1=1.0 |
| | $s_2^3$ | (0, 0, 1) | 0.0 | 0.1 |

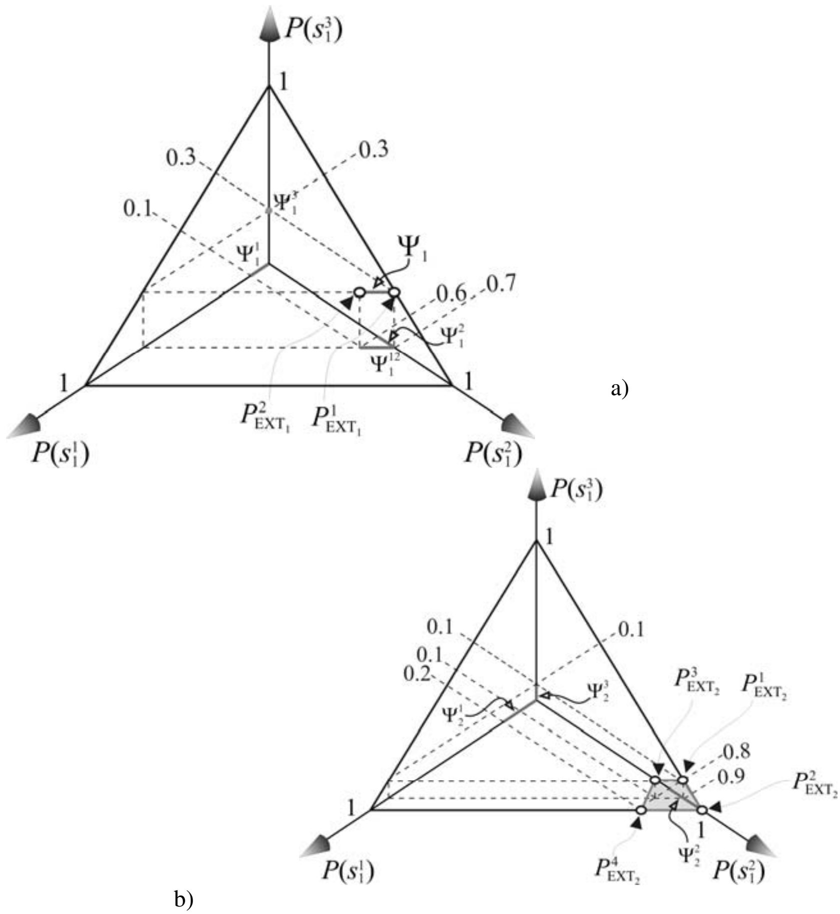



**Fig. 4.6** Example 4.2: a) set $\Psi_1$ and extremes $P_{EXT1}$ b) set $\Psi_2$ and extremes $P_{EXT2}$

**Table 4.4** Example 4.2: Solutions of the linear programming problems for the lower and upper probabilities for $T$. $^1$ = simplex method; $^2$ = interior point method

| Solution for | Joint **P** | $\left(c_1^1, c_1^2\right)$ | Marginal on $S_1$ | $\left(c_2^1, c_2^2, c_2^3, c_2^4\right)$ | Marginal on $S_2$ |
|---|---|---|---|---|---|
| Min$^1$ | $\begin{pmatrix} 0 & 0.1 & 0 \\ 0.2 & 0.4 & 0 \\ 0.3 & 0 & 0 \end{pmatrix}$ | (0, 1) | $\mathbf{p}_{EXT_1}^2 =$ $(0.1, 0.6, 0.3)^T$ | (0,0,0,1) | $\mathbf{p}_{EXT_2}^4 =$ $(0.2, 0.8, 0)^T$ |
| Min$^2$ | $\begin{pmatrix} 0 & 0.1 & 0 \\ 0.1388 & 0.4 & 0.0612 \\ 0 & 0.3 & 0 \end{pmatrix}$ | (0, 1) | $\mathbf{p}_{EXT_1}^2 =$ $(0.1, 0.6, 0.3)^T$ | (0, 0, 0.612, 0.388) | (0.1388, 0.8, 0.0612) |
| Max$^1$ | $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0.2 & 0.1 \end{pmatrix}$ | (1, 0) | $\mathbf{p}_{EXT_1}^1 =$ $(0, 0.7, 0.3)^T$ | (1,0,0,0) | $\mathbf{p}_{EXT_2}^1 =$ $(0, 0.9, 0.1)^T$ |
| Max$^2$ | $\begin{pmatrix} 0.0250 & 0 & 0 \\ 0 & 0.6750 & 0 \\ 0.0391 & 0.1608 & 0.1 \end{pmatrix}$ | (0.75, 0.25) | $(0.025, 0.675, 0.3)^T$ | (0.358, 0, 0.642, 0) | (0.064, 0.836, 0.1) |

**Example 4.3.** Let us now slightly modify Example 4.2. In particular, let us keep the information on $S_2$ the same, but assume that the following information is available on $S_1$: $P(s_1^{\ 1}) + 4/3P(s_1^{\ 2}) \leq 14/15$; $P(s_1^{\ 1}) + 7/5P(s_1^{\ 2}) \geq 47/50$; $P(s_1^{\ 3}) \leq 0.3$. Table 4.5 summarizes the complete information. Since bounds on expectations of general functions are now given on $S_1$, planes bounding $\Psi_1$ are not parallel to any coordinate plane (Figure 4.7), and $\Psi_1$ cannot be generated by bounds on probabilities of events, let alone on the probability of singletons. Constraints (4.26) are thus equal to:

$$
\begin{aligned}
&\textit{Subject to}\\
&1\cdot\left(p^{1,1}+p^{1,2}+p^{1,3}\right)+4/3\cdot\left(p^{2,1}+p^{2,2}+p^{2,3}\right)+0\cdot\left(p^{3,1}+p^{3,2}+p^{3,3}\right)\leq 14/15\\
&47/50\leq 1\cdot\left(p^{1,1}+p^{1,2}+p^{1,3}\right)+7/5\cdot\left(p^{2,1}+p^{2,2}+p^{2,3}\right)+0\cdot\left(p^{3,1}+p^{3,2}+p^{3,3}\right)\\
&p^{3,1}+p^{3,2}+p^{3,3}\leq 0.3\\
&0.0\leq p^{1,1}+p^{2,1}+p^{3,1}\leq 0.2\\
&0.8\leq p^{1,2}+p^{2,2}+p^{3,2}\leq 1.0\\
&0.0\leq p^{1,3}+p^{2,3}+p^{3,3}\leq 0.1\\
&p^{1,1}+p^{1,2}+p^{1,3}+p^{2,1}+p^{2,2}+p^{2,3}+p^{3,1}+p^{3,2}+p^{3,3}=1\\
&p^{i,j}\geq 0
\end{aligned}
\tag{4.34}
$$

As shown in Figure 4.7, the assigned previsions on $S_1$ augment the extreme points of $\Psi_1$ by adding $\mathbf{p}^3_{EXT_1} = (0.8, 0.1, 0.1)^{\mathrm{T}}$ (intersection of the first two previsions) to the two extreme points in Example 4.2. The constraints in problems (4.28) now read as follows:

$$\begin{aligned}
&\textit{Subject to}\\
&p^{1,1} + p^{1,2} + p^{1,3} - 0.0c_1^1 - 0.1c_1^2 - 0.8c_1^3 = 0\\
&p^{2,1} + p^{2,2} + p^{2,3} - 0.7c_1^1 - 0.6c_1^2 - 0.1c_1^3 = 0\\
&p^{3,1} + p^{3,2} + p^{3,3} - 0.3c_1^1 - 0.3c_1^2 - 0.1c_1^3 = 0\\
&p^{1,1} + p^{2,1} + p^{3,1} - 0.0c_2^1 - 0.0c_2^2 - 0.1c_2^3 - 0.2c_2^4 = 0\\
&p^{1,2} + p^{2,2} + p^{3,2} - 0.9c_2^1 - 1.0c_2^2 - 0.8c_2^3 - 0.8c_2^4 = 0\\
&p^{1,3} + p^{2,3} + p^{3,3} - 0.1c_2^1 - 0.0c_2^2 - 0.1c_2^3 - 0.0c_2^4 = 0\\
&c_1^1 + c_1^2 + c_1^3 = 1\\
&c_2^1 + c_2^2 + c_2^3 + c_2^4 = 1\\
&c_i^{\xi} \geq 0\\
&p^{i,j} \geq 0
\end{aligned} \tag{4.35}$$

**Table 4.5** Example 4.3: Upper and lower previsions

| Marginal $S_i$ | $k$ | $\left(\mathbf{f}_i^k\right)^{\mathrm{T}}$ | $E_{LOW}\left[f_i^k\right]$ | $E_{UPP}\left[f_i^k\right]$ |
|---|---|---|---|---|
| $S_1$ | 1 | (1, 4/3, 0) | -∞ | 14/15 |
| | 2 | (1, 7/5, 0) | 47/50 | ∞ |
| | 3 | (0, 0, 1) | 0.3 | 0.3 |
| $S_2$ | 1 | (1, 0, 0) | 0.0 | 0.2 |
| | 2 | (0, 1, 0) | 0.8 | 1.0 |
| | 3 | (0, 0, 1) | 0.0 | 0.1 |

The lower and upper probabilities for the same event $T$ as in Example 4.2 are now equal to 0 and 0.825, respectively; solutions are detailed in Table 4.6. The minimizing solution found by the simplex method is achieved at the same extreme points as in Example 4.2, whereas the maximizing solution is the same with both methods and is not achieved at an extreme point of $\Psi_1$.

**Table 4.6** Example 4.3: Solutions of the linear programming problems for the lower and upper probabilities for $T$. [1] = simplex method; [2] = interior point method

| Solution for | Joint $\mathbf{P}$ | $\left(c_1^1, c_1^2\right)$ | Marginal on $S_1$ | $\left(c_2^1, c_2^2, c_2^3, c_2^4\right)$ | Marginal on $S_2$ |
|---|---|---|---|---|---|
| Min[1] | $\begin{pmatrix} 0 & 0.8 & 0 \\ 0.1 & 0 & 0 \\ 0.1 & 0 & 0 \end{pmatrix}$ | (0, 0, 1) | $\mathbf{p}_{EXT_1}^2 =$ $(0.1, 0.6, 0.3)^T$ | (0, 0, 0, 1) | $\mathbf{p}_{EXT_2}^4 =$ $(0.2, 0.8, 0)^T$ |
| Min[2] | $\begin{pmatrix} 0 & 0.7221 & 0.0009 \\ 0.1021 & 0 & 0.0543 \\ 0.0022 & 0.1184 & 0 \end{pmatrix}$ | (0.05, 0.05, 0.90) | (0.723, 0.157, 0.120) | (0.186, 0.110, 0.366, 0.338) | (0.104, 0.841, 0.055) |
| Max[1,2] | $\begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.625 & 0 \\ 0 & 0.175 & 0.1 \end{pmatrix}$ | (0.875, 0, 0.125) | (0.1, 0.625, 0.275) | (0, 0, 1, 0) | $\mathbf{p}_{EXT_2}^3 =$ $(0.1, 0.8, 0.1)^T$ |

If bounds on the probabilities of the singletons are considered, a much wider set $\Psi_1^*$ is obtained, which has vertices $\mathbf{p}_{EXT_1}^1$, $\mathbf{p}_1^* = (0.2, 0.7, 0.1)^T$, $\mathbf{p}_{EXT_1}^3$, and $\mathbf{p}_1^{**} = (0.6, 0.1, 0.3)^T$ (Figure 4.7). The lower probability for event $T$ remains unchanged, but its upper probability increases to 0.9, and the maximizing solution is:

$$\mathbf{p} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0.1 & 0 \end{pmatrix}, \tag{4.36}$$

which yields the marginals: $\mathbf{p}_1^* = (0.2, 0.7, 0.1)^T$ (vertex of $\Psi_1$) and $\mathbf{p}_{EXT_2}^4$ (same vertex as in Example 4.2).

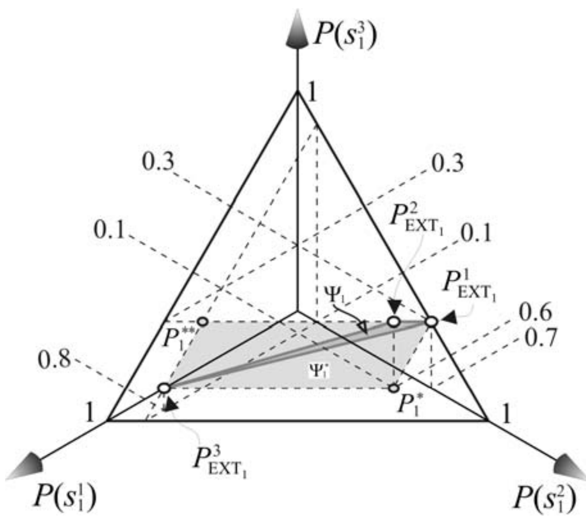



**Fig. 4.7** Example 4.3: sets $\Psi_1$ and $\Psi_1^*$

This example shows that, in a general case of assigned bounds on expectations of the marginals, using bounds on event probabilities would give larger outer bounds on the probability of events or expectations on the joint space.

**Example 4.4.** Let us move to a four-dimensional space, and assume that the following information (random set) be available on $S_1$: $(\{s_1^1, s_1^3\}, 0.2)$, $(\{s_1^2, s_1^3\}, 0.3)$, $(\{s_1^2, s_1^3, s_1^4\}, 0.3)$, and $(\{s_1^3, s_1^4\}, 0.1)$, which is equivalent to $P(s_1^1)\leq 0.2$, $P(s_1^2)\leq 0.7$, $P(s_1^4)\leq 0.5$, $P(s_1^1) + P(s_1^2)\leq 0.9$, $0.2\leq P(s_1^1) + P(s_1^3)$, and $0.4\leq P(s_1^2) + P(s_1^3)$, shown in Figure 4.8 and Table 4.7.

Since $S_1$ is four-dimensional, the information contained in a random set cannot be completely written in terms of bounds on the probability of the singletons, as it could in the three-dimensional spaces introduced in Example 4.2 and Example 4.3 (Section 4.3.2). As a consequence, the boundaries of $\Psi_1$ in Figure 4.8 are not all parallel to the coordinate axes.

The information available on $S_2$ is: $P(s_2^3) + 7/5P(s_2^4) \geq 47/50$; $P(s_2^4) \leq 0.3$ (Figure 4.9 and Table 4.7). Constraints (4.26) are equal to:

$$
\begin{aligned}
&\textit{Subject to}\\
&0.0 \le p^{1,1} + p^{1,2} + p^{1,3} + p^{1,4} \le 0.2\\
&0.0 \le p^{2,1} + p^{2,2} + p^{2,3} + p^{2,4} \le 0.7\\
&0.0 \le p^{4,1} + p^{4,2} + p^{4,3} + p^{4,4} \le 0.5\\
&0.0 \le p^{1,1} + p^{1,2} + p^{1,3} + p^{1,4} + p^{2,1} + p^{2,2} + p^{2,3} + p^{2,4} \le 0.9\\
&0.2 \le p^{1,1} + p^{1,2} + p^{1,3} + p^{1,4} + p^{3,1} + p^{3,2} + p^{3,3} + p^{3,4} \le 1\\
&0.4 \le p^{2,1} + p^{2,2} + p^{2,3} + p^{2,4} + p^{3,1} + p^{3,2} + p^{3,3} + p^{3,4} \le 1\\
&47/50 \le 1\cdot\left(p^{1,3} + p^{2,3} + p^{3,3} + p^{4,3}\right) + 7/5\cdot\left(p^{1,4} + p^{2,4} + p^{3,4} + p^{4,4}\right)\\
&0.0 \le p^{1,4} + p^{2,4} + p^{3,4} + p^{4,4} \le 0.3\\
&p^{1,1} + p^{1,2} + p^{1,3} + p^{1,4} + p^{2,1} + p^{2,2} + p^{2,3} + p^{2,4}\\
&+p^{3,1} + p^{3,2} + p^{3,3} + p^{3,4} + p^{4,1} + p^{4,2} + p^{4,3} + p^{4,4} = 1\\
&p^{i,j} \ge 0
\end{aligned}
\tag{4.37}
$$

As shown in Figure 4.8, constraint $0 \leq P(s_1^1)+ P(s_1^2) \leq 0.9$ does not intersect the convex set $\Psi_1$ , i.e. the given bounds are not coherent. Since the optimization problem is defined by all constraints in Eq. (4.37), the solution on the joint space is coherent because it satisfies the bounds in Eq. (4.19). Constraints (4.37) can be directly applied to find the optimal solution by linear programming or can be used to determine the extreme joint distributions.

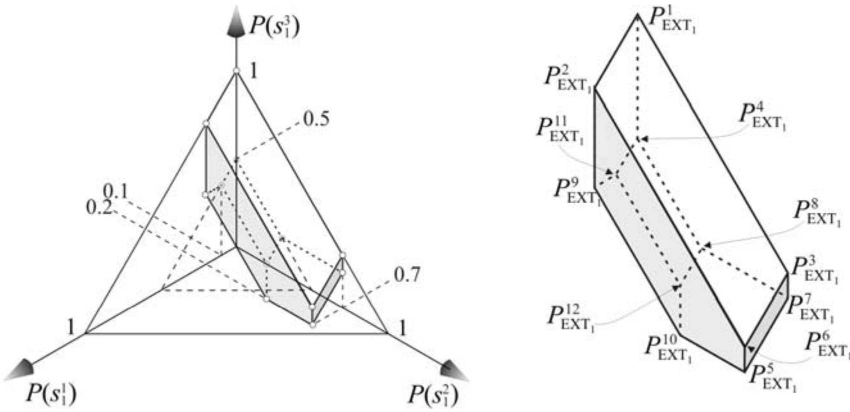
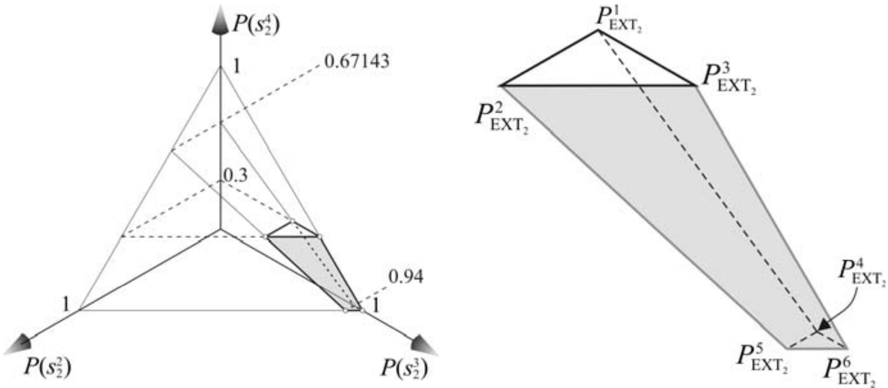



**Fig. 4.8** Example 4.4: Projection of set $\Psi_1$ onto the three-dimensional space $P(s_1{}^1)$, $P(s_1{}^2)$, and $P(s_1{}^3)$. Coordinates of extreme distributions are given in Table 4.7

**Table 4.7** Example 4.4: Extreme distributions for: (a) $\Psi_1$; and (b) $\Psi_2$

| $p^i_{EXT_1} \in \Psi_1$ | $(s_1{}^1, s_1{}^2, s_1{}^3, s_1{}^4)$ |
|---|---|
| $p^1_{EXT_1}$ | (0, 0, 1, 0 ) |
| $p^2_{EXT_1}$ | (0.2, 0, 0.8, 0) |
| $p^3_{EXT_1}$ | (0, 0.7, 0.3, 0) |
| $p^4_{EXT_1}$ | (0, 0, 0.5, 0.5) |
| $p^5_{EXT_1}$ | (0.2, 0.7, 0, 0.1) |
| $p^6_{EXT_1}$ | (0.2, 0.7, 0.1, 0) |
| $p^7_{EXT_1}$ | (0, 0.7, 0.2, 0.1) |
| $p^8_{EXT_1}$ | (0, 0.3, 0.2, 0.5) |
| $p^9_{EXT_1}$ | (0.2, 0, 0.4, 0.4) |
| $p^{10}_{EXT_1}$ | (0.2, 0.4, 0, 0.4) |
| $p^{11}_{EXT_1}$ | (0.1, 0, 0.4, 0.5) |
| $p^{12}_{EXT_1}$ | (0.1, 0.3, 0.1, 0.5) |

| $p^i_{EXT_2} \in \Psi_2$ | $(s_2{}^1, s_2{}^2, s_2{}^3, s_2{}^4)$ |
|---|---|
| $p^1_{EXT_2}$ | (0.18, 0, 0.52, 0.3) |
| $p^2_{EXT_2}$ | (0, 0.18, 0.52, 0.3) |
| $p^3_{EXT_2}$ | (0, 0, 0.7, 0.3) |
| $p^4_{EXT_2}$ | (0.06, 0, 0.94, 0) |
| $p^5_{EXT_2}$ | (0, 0.06, 0.94, 0) |
| $p^6_{EXT_2}$ | (0, 0, 1, 0) |




a) b)

**Fig. 4.9** Example 4.4: set $\Psi_2$. Coordinates of extreme distributions are given in Table 4.7

The constraints in problems (4.28) now read as follows:

$$
\begin{aligned}
&\textit{Subject to}\\
&\mathbf{P} = \begin{pmatrix} p^{1,1} & p^{1,2} & p^{1,3} & p^{1,4} \\ \vdots & & & \vdots \\ p^{4,1} & p^{4,2} & p^{4,3} & p^{4,4} \end{pmatrix}\\
&\mathbf{P}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \left(c_1^1, c_1^2, \ldots, c_1^{12}\right)\begin{pmatrix} P^1_{EXT1} \\ P^2_{EXT1} \\ \vdots \\ P^{12}_{EXT1} \end{pmatrix}\\
&p^{1,1} + p^{2,1} + p^{3,1} + p^{4,1} = 0.18c_2^1 + 0.06c_2^4\\
&p^{1,2} + p^{2,2} + p^{3,2} + p^{4,2} = 0.18c_2^2 + 0.06c_2^5\\
&p^{1,3} + p^{2,3} + p^{3,3} + p^{4,3} = 0.52c_2^1 + 0.52c_2^2 + 0.7c_2^3 + 0.94c_2^4 + 0.94c_2^5 + 1\cdot c_2^6\\
&p^{1,4} + p^{2,4} + p^{3,4} + p^{4,4} = 0.3c_2^1 + 0.3c_2^2 + 0.3c_2^3\\
&c_1^1 + c_1^2 + c_1^3 + c_1^4 + c_1^5 + c_1^6 + c_1^7 + c_1^8 + c_1^9 + c_1^{10} + c_1^{11} + c_1^{12} = 1\\
&c_2^1 + c_2^2 + c_2^3 + c_2^4 + c_2^5 + c_2^6 = 1\\
&c_i^\xi \geq 0\\
&p^{i,j} \geq 0
\end{aligned}
\tag{4.38}
$$

Let us now consider the event $T = \left\{\left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right), \left(s_1^4, s_2^4\right)\right\}$. The lower and upper probabilities for the event $T$ in Example 4.4 are 0 and 1, respectively; solutions are detailed in Table 4.8. The different min/max solutions from different methods show that the optimal solutions are not necessarily achieved at the same point although they achieve the same probability for event $T$, indicating multiple min/max solutions.

**Table 4.8** Example 4.4: Solutions of the linear programming problems for the lower and upper probabilities for $T$. [1] = simplex method; [2] = interior point method; [3] = exhaustive search over all extreme joint distributions

| Solution for | Joint **P** | $\begin{pmatrix} c_1^1, c_1^2, c_1^3, c_1^4, \\ c_1^5, c_1^6, c_1^7, c_1^8, \\ c_1^9, c_1^{10}, c_1^{11}, c_1^{12} \end{pmatrix}$ | Marginal on $S_1$ | $\begin{pmatrix} c_2^1, c_2^2, c_2^3, \\ c_2^4, c_2^5, c_2^6 \end{pmatrix}$ | Marginal on $S_2$ |
|---|---|---|---|---|---|
| Min[1] | $\begin{pmatrix} 0 & 0.1 & 0 & \\ 0.08 & 0 & 0.02 & 0.2 \\ 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0.5 & 0 \end{pmatrix}$ | — | $(0.1, 0.3, 0.1, 0.5)^T$ | — | $(0.08, 0.1, 0.52, 0.3)^T$ |
| Min[2] | $\begin{pmatrix} 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0.16 & 0 \\ 0 & 0.18 & 0 & 0.3 \\ 0 & 0 & 0.16 & 0 \end{pmatrix}$ | $\begin{pmatrix} 0.6, & 0, & 0, & 0, \\ 0, & 0, & 0, & 0, \\ 0, & 0, & 0, & 0, \\ 0, & 0.4, & 0, & 0 \end{pmatrix}$ | $(0.2, 0.16, 0.48, 0.16)^T$ | $\begin{pmatrix} 1, & 0, & 0, \\ 0, & 0, & 0 \end{pmatrix}$ | $(0, 0.18, 0.52, 0.3)^T$ |
| Min[3] | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}$ | — | $(0, 0.7, 0.2, 0.1)^T$ | — | $(0, 0, 0.8, 0.2)^T$ |
| Max[1,3] | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ | — | $(0, 0, 1, 0)^T$ | — | $(0, 0, 1, 0))^T$ |
| Max[2] | $\begin{pmatrix} 0.04 & 0 & 0 & 0 \\ 0 & 0.14 & 0 & 0 \\ 0 & 0 & 0.52 & 0 \\ 0 & 0 & 0 & 0.3 \end{pmatrix}$ | $\begin{pmatrix} 0.2, & 0, & 0, & 0.6, \\ 0, & 0.2, & 0, & 0, \\ 0, & 0, & 0, & 0, \\ 0, & 0, & 0, & 0 \end{pmatrix}$ | $(0.04, 0.14, 0.52, 0.3)^T$ | $\begin{pmatrix} 0.22, & 0.78, & 0, \\ 0, & 0, & 0 \end{pmatrix}$ | $(0.04, 0.14, 0.52, 0.3)^T$ |

## 4.2.2 Epistemic Independence and Irrelevance

Let $S_1$ and $S_2$ be finite sets. In epistemic independence, if we learn that the actual value of $s_2$ is $s_2$*, then the probability measure for $s_1$ is again one of the probability measures in $\Psi_1$ (but in general not always the same for different values $s_2$*), and vice versa. As a consequence, the definition of epistemic independence uses the concept of conditional probability, and the set of joint probability measures, $\Psi_E$, is just the largest set of joint measures that are extensions to Eq. (4.15):

$$\begin{aligned} &P\left(.\times S_2 \mid S_1 \times \{s_2\}\right) \in \Psi_1 \quad \forall s_2 : \exists P_2 \in \Psi_2 : P_2\left(\{s_2\}\right) > 0\,; \\ &P\left(S_1 \times . \mid \{s_1\} \times S_2\right) \in \Psi_2 \quad \forall s_1 : \exists P_1 \in \Psi_1 : P_1\left(\{s_1\}\right) > 0 \end{aligned} \tag{4.39}$$

Epistemic independence is the appropriate model when we are given two marginal sets of probability measures (or the corresponding sets of coherent desirable gambles, or coherent upper or lower previsions, Section 3.3.2), together with a judgment that the experiments are epistemically independent (our uncertainty about either of the two outcomes does not change when we obtain some information about the other outcome). However, we are unwilling to make stronger assumptions, e.g., that there are underlying stochastic mechanisms that are stochastically independent, which could justify the concept of strong independence defined in the next Section 4.2.3.

Let us now derive a useful characterization of epistemic independence. In the equations that follow, expressions will be given in terms of both probability measures and vectors and matrices (Section 2.3). If $T_1 = U_1 \times S_2$, $T_2 = S_1 \times \{s_2\}$, definition (4.12) can be used to derive an expression for the joint measure, $P$, by noticing that $T_1 \cap T_2 = U_1 \times \{s_2\}$ (see also Eq. (2.27):

$$P\left(U_1 \times \{s_2\}\right) = P\left(U_1 \times S_2 \mid S_1 \times \{s_2\}\right) P_2\left(\{s_2\}\right); \mathbf{P} = \mathbf{P}_{1|2}\, Diag(\mathbf{p}_2) \quad (4.40)$$

Likewise (see also Eq. (2.29)):

$$P\left(\{s_1\} \times U_2\right) = P\left(S_1 \times U_2 \mid \{s_1\} \times S_2\right) P_1\left(\{s_1\}\right); \mathbf{P} = Diag(\mathbf{p}_1)\, \mathbf{P}_{2|1}. \quad (4.41)$$

Thus, two variables are epistemically independent if $\forall (s_1, s_2) \in S_1 \times S_2$:

$$\begin{aligned} &\exists P_1^{|s_2} \in \Psi_1 : P\left(U_1 \times \{s_2\}\right) = P_1^{|s_2}\left(U_1\right) P_2\left(\{s_2\}\right) \ AND \\ &\exists P_2^{|s_1} \in \Psi_2 : P\left(\{s_1\} \times U_2\right) = P_2^{|s_1}\left(U_2\right) P_1\left(\{s_1\}\right) \end{aligned} \quad (4.42)$$

In matrix terms: two variables are epistemically independent if each column of $\mathbf{P}_{1|2}$ is a vector in $\Psi_1$ (and this vector may be different for each column) and if each row in $\mathbf{P}_{2|1}$ is the transpose of a vector in $\Psi_2$ (and this vector may be different for each row), i.e. $\forall \mathbf{P}$:

$$\begin{aligned} &\exists n_2 \ \left(\mathbf{p}_1^{(i)} \in \Psi_1\right) : \mathbf{P} = \mathbf{P}_{1|2} Diag\left(\mathbf{p}_2\right), \mathbf{p}_2 \in \Psi_2, \mathbf{P}_{1|2} = \left(\mathbf{p}_1^{(1)}\ \mathbf{p}_1^{(2)} \ldots \mathbf{p}_1^{(n_2)}\right) \\ &\exists n_1 \ \left(\mathbf{p}_2^{(i)} \in \Psi_2\right) : \mathbf{P} = Diag\left(\mathbf{p}_1\right) \mathbf{P}_{2|1} : \mathbf{p}_1 \in \Psi_1, \ \mathbf{P}_{2|1} = \begin{pmatrix} \left(\mathbf{p}_2^{(1)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2^{(2)}\right)^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \left(\mathbf{p}_2^{(n_1)}\right)^{\mathrm{T}} \end{pmatrix} \end{aligned} \quad (4.43)$$

Notice that, if the probabilities are precise, then $\Psi_1$ and $\Psi_2$ contain only one element each (say, $P_i$), and thus for $s_1$:

- $P_1^{|s_2}(U_1) = P(U_1 \times S_2 \mid S_1 \times \{s_2\}) = P_1(U_1)$, and Eq. (4.40) becomes Eq. (4.14).
- The columns of $\mathbf{P}_{1|2}$ are all equal, and Eq. (4.40) becomes Eq. (2.35).

Likewise for $s_2$.

When only the two marginal convex sets $\Psi_1$ and $\Psi_2$ are given, $\Psi_{\mathrm{E}}$ is the convex set:

$$\begin{aligned}
&\Psi_{\mathrm{E}} = \left\{\mathbf{P} \in \Psi_E^{|s_2} \cap \Psi_E^{|s_1}\right\} \\
&\Psi_{\mathrm{E}}^{|s_2} := \left\{\mathbf{P} = \mathbf{P}_{1|2} Diag(\mathbf{p}_2) : \mathbf{p}_2 \in \Psi_2; \mathbf{P}_{1|2} = \left(\underbrace{\mathbf{p}_1^{(1)}\ \mathbf{p}_1^{(2)} \ldots \mathbf{p}_1^{(n_2)}}_{n_2\ \left(\mathbf{p}_1^{(i)} \in \Psi_1\right)}\right)\right\}; \\
&\Psi_{\mathrm{E}}^{|s_1} := \left\{\mathbf{P} = Diag(\mathbf{p}_1)\mathbf{P}_{2|1} : \mathbf{p}_1 \in \Psi_1; \mathbf{P}_{2|1} = \left( n_1\ \left(\mathbf{p}_2^{(i)} \in \Psi_2\right) \begin{cases} \left(\mathbf{p}_2^{(1)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2^{(2)}\right)^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \left(\mathbf{p}_2^{(n_1)}\right)^{\mathrm{T}} \end{cases} \right)\right\}
\end{aligned} \tag{4.44}$$

where $\Psi_{\mathrm{E}}^{|s_2}$ is called the "*irrelevant natural extension"* (Couso, Moral et al. 1999) of the two marginals when the second experiment is *epistemically irrelevant* to the first" (i.e. the set of desirable gambles concerning the first experiment does not change when we learn the outcome of the second experiment). $\Psi_{\mathrm{E}}^{|s_1}$ is called the "irrelevant natural extension of the two marginals when the first experiment is epistemically irrelevant to the second" (i.e. the set of acceptable gambles concerning the second experiment does not change when we learn the outcome of the first experiment); in this case, $s_1$ is selected according to some marginal distribution in $\Psi_1$, and then $s_2$ is selected according to a distribution from $\Psi_2$ that may depend on $s_1$. It is important to notice that $s_2$ may be selected by a different procedure for different values of $s_1$.

Therefore, in imprecise probabilities, irrelevance of one experiment with respect to another is a directional or asymmetric relation. Such a lack of symmetry vanishes:

- Always in precise probability because both $\Psi_1$ and $\Psi_2$ contain only one element each.
- In imprecise probabilities when irrelevance applies in both directions, i.e. when each experiment is epistemically irrelevant to the other experiment: this is the case of epistemic independence.

The calculation of the upper and lower expectations (and thus probabilities) on the joint space $S$ is no longer a linear optimization problem because matrix $\mathbf{P}$ in Eq. (4.44) is the result of the multiplication of two matrices, each containing a marginal's entries. This makes the constraints non-convex because they are of the kind: $p^{i,j} = P_1(s_1{}^i)\,P_2(s_2{}^j)$. Optimization problems that involve non-convex constraints are known to be NP-complete in the strong sense (or NP-hard) (e.g., (Horst, Pardalos et al. 2000)), i.e. there is no fully polynomial-time approximation scheme to solve them. Later in this section (page 142), a different algorithm to avoid this problem will be discussed.

For the case of epistemic independence, let us write down the optimization problem in the (quadratic) constraints when marginals are constrained by upper and lower previsions (Eq. (4.17)):

$$
\begin{aligned}
&\text{minimize } \sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j}\,p^{i,j} \quad \left(-\sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j}\,p^{i,j}\right)\\
&\textit{subject to}\\
&\mathbf{P} = \left(\mathbf{p}_1{}^{(1)}\ \mathbf{p}_1{}^{(2)}\ldots\mathbf{p}_1{}^{(n_2)}\right) Diag\left(\mathbf{p}_2{}^{(n_1+1)}\right)\\
&\mathbf{P} = Diag\left(\mathbf{p}_1{}^{(n_2+1)}\right)\begin{pmatrix} \left(\mathbf{p}_2{}^{(1)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2{}^{(2)}\right)^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \left(\mathbf{p}_2{}^{(n_1)}\right)^{\mathrm{T}} \end{pmatrix}\\
&E_{LOW}\left[f_1^k\right] \le \left(\mathbf{f}_1^k\right)^{\mathrm{T}}\ \mathbf{p}_1{}^{(j)} \le E_{UPP}\left[f_1^k\right];\ k=1,\ldots,k_1;\ j=1,\ldots,n_2+1;\\
&E_{LOW}\left[f_2^k\right] \le \left(\mathbf{f}_2^k\right)^{\mathrm{T}}\ \mathbf{p}_2{}^{(j)} \le E_{UPP}\left[f_2^k\right];\ k=1,\ldots,k_2;\ j=1,\ldots,n_1+1;\\
&\mathbf{1}_{(n_1)}{}^{\mathrm{T}}\cdot\mathbf{p}_1{}^{(j)} = 1\quad j=1,\ldots,n_2+1; \qquad \mathbf{1}_{(n_2)}{}^{\mathrm{T}}\cdot\mathbf{p}_2{}^{(j)} = 1\quad j=1,\ldots,n_1+1\\
&\mathbf{p}_1{}^{(j)} \ge 0\quad j=1,\ldots,n_2+1; \qquad \mathbf{p}_2{}^{(j)} \ge 0\quad j=1,\ldots,n_1+1
\end{aligned}
\tag{4.45}
$$

Optimization problems (4.45) involve $(n_1+1)\times n_2$ + $(n_2+1)\times n_1$ variables $P_i^{(j)}\left(s_i\right)$ and $n_1\times n_2$ variables $p^{i,j}$. The constraints $\mathbf{1}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{1}=1;\ \mathbf{P}\geq 0$ are satisfied automatically because the entries of matrix **P** are products of probability distributions.

When marginals are assigned through their extreme distributions (Eq. (4.18)), the optimization problems become:

$$
\begin{aligned}
&\text{minimize } \sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j}p^{i,j} \quad \left(-\sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j}p^{i,j}\right)\\
&\textit{subject to}\\
&\mathbf{P}=\left(\sum_{\xi=1}^{\xi_1} c_1^{\xi,1}\mathbf{p}_{EXT_1}^{\xi}\cdots\sum_{\xi=1}^{\xi_1} c_1^{\xi,n_2}\mathbf{p}_{EXT_1}^{\xi}\right)Diag\left(\sum_{\xi=1}^{\xi_2} c_2^{\xi,n_1+1}\mathbf{p}_{EXT_2}^{\xi}\right)\\
&\mathbf{P}=Diag\left(\sum_{\xi=1}^{\xi_1} c_1^{\xi,n_2+1}\mathbf{p}_{EXT_1}^{\xi}\right)\begin{pmatrix}\left(\sum_{\xi=1}^{\xi_2} c_2^{\xi,1}\mathbf{p}_{EXT_2}^{\xi}\right)^{\mathrm{T}}\\ \cdot\\ \cdot\\ \cdot\\ \left(\sum_{\xi=1}^{\xi_2} c_2^{\xi,n_1}\mathbf{p}_{EXT_2}^{\xi}\right)^{\mathrm{T}}\end{pmatrix}\\
&\sum_{\xi=1}^{\xi_1} c_1^{\xi,j}=1\ \ j=1,\dots,n_2; \qquad \sum_{\xi=1}^{\xi_2} c_2^{\xi,j}=1\ \ j=1,\dots,n_1\\
&c_1^{\xi,j}\geq 0\ \ j=1,\dots,n_2;\xi=1,\dots,\xi_2 \qquad c_2^{\xi,j}\geq 0\ \ j=1,\dots,n_1;\ \xi=1,\dots,\xi_1
\end{aligned}
\tag{4.46}
$$

where there are $\left(\xi_1\times\left(n_2+1\right)\right)+\left(\xi_2\times\left(n_1+1\right)\right)$ variables $c_i^{\xi,j}$ and $n_1\times n_2$ variables $p^{i,j}$. The constraints $\mathbf{1}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{1}=1;\ \mathbf{P}\geq 0$ are satisfied automatically because the entries of matrix **P** are products of convex sets of distributions.

If expectation bounds are given on one marginal and extreme distributions are assigned on the other marginal, then constraints are properly selected from Eqs. (4.45) and (4.46). In Eqs. (4.45) and (4.46), if the second experiment is epistemically irrelevant to the first, then $n_2 = 0$; if the first experiment is epistemically irrelevant to the second, then $n_1 = 0$.

Condition (4.42) on probability measures can be generalized to infinite sets $S_i$ with σ-algebra $\mathcal{S}_i$ as follows (Fetz and Oberguggenberger 2004). Let $\mathcal{S}$ be the σ-algebra generated by $\mathcal{S}_1\times\mathcal{S}_2$; given $T\in\mathcal{S}$, let $T_{s_1}=\left\{s_2:\left(s_1,s_2\right)\in T\right\}$. $\Psi_{\mathrm{E}}$ is the set of probability measures on $\mathcal{S}$ such that there are families of probability measures $\left\{P_i^{|s_j}\in\Psi_i;s_j\in S_j\right\}$ on $\mathcal{S}_i$ that satisfy the following conditions for all $T\in\mathcal{S}$:

1) The mapping $s_j \mapsto P_i^{|s_j}\left(T_{s_j}\right)$ is $P_j$–measurable: see (Section 2.2).
2) $P(T)=\int_{S_1} P_2^{|s_1}\left(T_{s_1}\right)dP_1\left(s_1\right)=\int_{S_2} P_1^{|s_2}\left(T_{s_2}\right)dP_2\left(s_2\right)$

**Example 4.5.** Consider again the situation and knowledge available in Example 4.1, where a two-component resin has to be applied at a construction site to anchor steel bars. Suppose now that three boxes are delivered to the construction site. The first box has the same content as in Example 4.1. Our knowledge about the other two boxes is the same as the second box in Example 4.1, but the four cartriges of unknown type may be different in the second and third box. This time, the worker in the field picks his first cartridge from box 1:

- If it is Type A, then he picks the second cartridge from box 2.
- Otherwise, he picks the second cartridge from box 3.

We want to write down the optimization problems for finding upper and lower expectations on the joint space and then calculate the upper and lower probabilities for the case in which the resin is not activated because the same resin type is selected, i.e. event $T$={(A, A), (B, B)}. Finally, calculate the conditional upper and lower probabilities that the first resin is Type A given the type of the second resin, and contrast this with the conditional upper and lower probabilities that the second resin is Type A given the type of the first resin.

In this example, the first experiment is epistemically irrelevant to the second experiment because:

1) The set of acceptable gambles concerning the second experiment does not change when we learn the outcome of the first experiment.
2) $s_1$ is selected according to some marginal distribution in $\Psi_1$, and then $s_2$ is selected according to a distribution from $\Psi_2$ that depends on $s_1$.
3) $s_2$ is selected by a different procedure for different values of $s_1$

As a consequence, $n_2 = 0$ and constraints in Eq. (4.45) become:

$$\begin{aligned}
&\textit{Subject to}\\
&\mathbf{P}=Diag\left(\mathbf{p}_1^{(1)}\right)\begin{pmatrix}\left(\mathbf{p}_2^{(1)}\right)^{\mathrm{T}}\\ \left(\mathbf{p}_2^{(2)}\right)^{\mathrm{T}}\end{pmatrix}\\
&0.5\le(1,0)^{\mathrm{T}}\ \mathbf{p}_1^{(1)}\le 0.8\\
&0.3\le(1,0)^{\mathrm{T}}\ \mathbf{p}_2^{(1)}\le 0.7\\
&0.3\le(1,0)^{\mathrm{T}}\ \mathbf{p}_2^{(2)}\le 0.7\\
&\mathbf{1}^{\mathrm{T}}\cdot\mathbf{p}_1^{(1)}=1;\ \mathbf{1}^{\mathrm{T}}\cdot\mathbf{p}_2^{(1)}=1;\ \mathbf{1}^{\mathrm{T}}\cdot\mathbf{p}_2^{(2)}=1\\
&\mathbf{p}_1^{(1)}\ge 0;\ \mathbf{p}_2^{(1)}\ge 0;\ \mathbf{p}_2^{(2)}\ge 0
\end{aligned}\tag{4.47}$$

Based on the extreme distributions given in Table 4.1, Eq. (4.46) gives the following constraints:

*Subject to*

$$\mathbf{P} = Diag\left(c_1^{1,1}\begin{pmatrix}0.5\\0.5\end{pmatrix} + c_1^{2,1}\begin{pmatrix}0.8\\0.2\end{pmatrix}\right)\begin{pmatrix}\left(c_2^{1,1}\begin{pmatrix}0.3\\0.7\end{pmatrix} + c_2^{2,1}\begin{pmatrix}0.7\\0.3\end{pmatrix}\right)^{\mathrm{T}}\\ \left(c_2^{1,2}\begin{pmatrix}0.3\\0.7\end{pmatrix} + c_2^{2,2}\begin{pmatrix}0.7\\0.3\end{pmatrix}\right)^{\mathrm{T}}\end{pmatrix} \tag{4.48}$$

$$c_1^{1,1} + c_1^{2,1} = 1;\ \ c_2^{1,1} + c_2^{2,1} = 1;\ \ c_2^{1,2} + c_2^{2,2} = 1$$

$$c_1^{1,1} \ge 0;\ c_1^{2,1} \ge 0;\ c_2^{1,1} \ge 0;\ c_2^{2,1} \ge 0;\ \ c_2^{1,2} \ge 0;\ c_2^{2,2} \ge 0$$

i.e.:

*Subject to*

$$\mathbf{P} = \begin{pmatrix}\left(0.5c_1^{1,1} + 0.8c_1^{2,1}\right)\left(0.3c_2^{1,1} + 0.7c_2^{2,1}\right) & \left(0.5c_1^{1,1} + 0.8c_1^{2,1}\right)\left(0.7c_2^{1,1} + 0.3c_2^{2,1}\right)\\ \left(0.5c_1^{1,1} + 0.2c_1^{2,1}\right)\left(0.3c_2^{1,2} + 0.7c_2^{2,2}\right) & \left(0.5c_1^{1,1} + 0.2c_1^{2,1}\right)\left(0.7c_2^{1,2} + 0.3c_2^{2,2}\right)\end{pmatrix} \tag{4.49}$$

$$c_1^{1,1} + c_1^{2,1} = 1;\ \ c_2^{1,1} + c_2^{2,1} = 1;\ \ c_2^{1,2} + c_2^{2,2} = 1$$

$$c_1^{1,1} \ge 0;\ c_1^{2,1} \ge 0;\ c_2^{1,1} \ge 0;\ c_2^{2,1} \ge 0;\ \ c_2^{1,2} \ge 0;\ c_2^{2,2} \ge 0$$

The upper and lower probabilities for the event $T$={(A, A), (B, B)} are equal to 3/10 and 7/10, respectively. The solutions summarized in Table 4.9 indicate that the minimizing $\mathbf{p}_2^{(1)} \neq \mathbf{p}_2^{(2)}$, i.e. they do not satisfy stochastic independence (Eq. (2.25)). Likewise for the maximizing solution. Notice that both the marginal on $S_1$ and the marginal on $S_2$ must be calculated by using the marginal rule (Eq. (2.24)).

We started from, and based our solution on, the observation that the first experiment is epistemically irrelevant to the second experiment. As a consequence, the conditional upper and lower probabilities that the second resin is Type A given the type of the first resin are equal to the marginal ones, i.e. 0.7 and 0.3, respectively. In order to check that epistemic irrelevance is a directional, asymmetric property, let us calculate the conditional upper and lower probabilities that the first resin is Type A given the type of the second resin. According to Eq. (2.25), the (non-linear) function to minimize and maximize is $p^{1,1}/(p^{1,1}+p^{2,1})$, where, by the marginal rule (Eq. (2.24)), $p^{1,1}+p^{2,1}$ is the first component of the marginal on $S_2$.

**Table 4.9** Example 4.5: Solutions of the optimization problems for the lower and upper probabilities for $T$

| Solution for | Joint $\mathbf{P}$ | $\left(c_1^{1,1}, c_1^{2,1}\right)$ $(\mathbf{p}_1^{(1)})^{\mathrm{T}}$ | Marginal on $S_1$ | $\left(c_2^{1,1}, c_2^{2,1}\right)$ $(\mathbf{p}_2^{(1)})^{\mathrm{T}}$ | $\left(c_2^{1,2}, c_2^{2,2}\right)$ $(\mathbf{p}_2^{(2)})^{\mathrm{T}}$ | Marginal on $S_2$ |
|---|---|---|---|---|---|---|
| Min | $\begin{pmatrix}39 & 91\\49 & 21\end{pmatrix}/200$ | (1/2, 1/2)<br>(13, 7)/20 | $(13/20, 7/20)^{\mathrm{T}}$ | (1, 0)<br>(0.3, 0.7) | (0, 1)<br>(0.7, 0.3) | $(0.44, 0.56)^{\mathrm{T}}$ |
| Max | $\begin{pmatrix}91 & 39\\21 & 49\end{pmatrix}/200$ | (1/2, 1/2)<br>(13, 7)/20 | $(13/20, 7/20)^{\mathrm{T}}$ | (0, 1)<br>(0.7, 0.3) | (1, 0)<br>(0.3, 0.7) | $(0.56, 0.44)^{\mathrm{T}}$ |

**Table 4.10** Example 4.5: Solutions of the optimization problems for the conditional upper and lower probabilities that the first resin is Type A given the type of the second resin

| Solution for | Joint $\mathbf{P}$ | $\left(c_1^{1,1}, c_1^{2,1}\right)$ $(\mathbf{p}_1^{(1)})^{\mathrm{T}}$ | Marginal on $S_1$ | $\left(c_2^{1,1}, c_2^{2,1}\right)$ $(\mathbf{p}_2^{(1)})^{\mathrm{T}}$ | $\left(c_2^{1,2}, c_2^{2,2}\right)$ $(\mathbf{p}_2^{(2)})^{\mathrm{T}}$ | Marginal on $S_2$ |
|---|---|---|---|---|---|---|
| Min | $\begin{pmatrix} 3 & 7 \\ 7 & 3 \end{pmatrix} / 20$ | (1, 0) (0.5, 0.5) | $(0.5, 0.5)^{\mathrm{T}}$ | (1, 0) (0.7, 0.3) | (0, 1) (0.3, 0.7) | $(0.5, 0.5)^{\mathrm{T}}$ |
| Max | $\begin{pmatrix} 28 & 12 \\ 3 & 7 \end{pmatrix} / 50$ | (0, 1) (0.8, 0.2) | $(0.8, 0.2)^{\mathrm{T}}$ | (0, 1) (0.3, 0.7) | (1, 0) (0.7, 0.3) | $(31, 19)/50^{\mathrm{T}}$ |

The constraints are still given by Eq. (4.47) or (4.49). The conditional upper and lower probabilities are equal to 28/31=0.903 and 0.3, which are larger bounds than the marginal bounds, i.e. 0.8 and 0.5. This means that the second experiment is epistemically relevant to the first one. The results summarized in Table 4.10 indicate that the minimizing and maximizing solutions again violate stochastic independence (Eq. (2.25)) because $\mathbf{p}_2^{(1)} \neq \mathbf{p}_2^{(2)}$.

**Example 4.6.** Consider again the situation and knowledge available in Example 4.1, where a two-component resin has to be applied at a construction site to anchor steel bars. In addition to the knowledge available in Example 4.1, all we now know about the stochastic mechanism for picking the two cartridges, i.e. the joint probability measure $P$ is that: (a) whatever the resin type of the first cartridge, the conditional probability that the second cartridge is A lies between 0.3 and 0.7; and (b) whatever the resin type of the second cartridge, the conditional probability that the first cartridge is A lies between 0.5 and 0.8.

We want to write down the optimization problems for finding upper and lower expectations on the joint space and then calculate the upper and lower probabilities for the case in which the resin is not activated because the same resin type is selected, i.e. event $T$={(A, A), (B, B)}. Finally, we calculate the conditional upper and lower probabilities that the first resin is Type A given the type of the second resin, and contrast to the conditional upper and lower probabilities that the second resin is Type A given the type of the first resin.

This is a case of epistemic independence, where each experiment is epistemically irrelevant to the other. As a consequence, this example is the symmetric counterpart of Example 4.5 above, where only the first experiment was epistemically irrelevant to the other. Constraints in Eqs. (4.45) and (4.46) become:

*Subject to*

$$\mathbf{P} = \left(\mathbf{p}_1^{(1)}\ \mathbf{p}_1^{(2)}\right) Diag\left(\mathbf{p}_2^{(3)}\right)$$

$$\mathbf{P} = Diag\left(\mathbf{p}_1^{(3)}\right) \begin{pmatrix} \left(\mathbf{p}_2^{(1)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2^{(2)}\right)^{\mathrm{T}} \end{pmatrix} \tag{4.50}$$

$$0.5 \le (1,0)^{\mathrm{T}}\ \mathbf{p}_1^{(i)} \le 0.8;\ i=1,2,3; \qquad 0.3 \le (1,0)^{\mathrm{T}}\ \mathbf{p}_2^{(i)} \le 0.7;\ i=1,2,3$$

$$\mathbf{1}^{\mathrm{T}} \cdot \mathbf{p}_1^{(i)} = 1;\ \mathbf{1}^{\mathrm{T}} \cdot \mathbf{p}_2^{(i)} = 1;\ i=1,2,3; \qquad \mathbf{p}_1^{(i)} \ge 0;\ \mathbf{p}_2^{(i)} \ge 0;\ i=1,2,3$$

Based on the extreme distributions given in Table 4.1, Eq. (4.46) gives the following constraints:

$$
\begin{aligned}
&\textit{Subject to}\\
&\mathbf{P}=\left(c_1^{1,1}\begin{pmatrix}0.5\\0.5\end{pmatrix}+c_1^{2,1}\begin{pmatrix}0.8\\0.2\end{pmatrix}\quad c_1^{1,2}\begin{pmatrix}0.5\\0.5\end{pmatrix}+c_1^{2,2}\begin{pmatrix}0.8\\0.2\end{pmatrix}\right)Diag\left(c_2^{1,3}\begin{pmatrix}0.3\\0.7\end{pmatrix}+c_2^{2,3}\begin{pmatrix}0.7\\0.3\end{pmatrix}\right)\\
&\mathbf{P}=Diag\left(c_1^{1,3}\begin{pmatrix}0.5\\0.5\end{pmatrix}+c_1^{2,3}\begin{pmatrix}0.8\\0.2\end{pmatrix}\right)\begin{pmatrix}\left(c_2^{1,1}\begin{pmatrix}0.3\\0.7\end{pmatrix}+c_2^{2,1}\begin{pmatrix}0.7\\0.3\end{pmatrix}\right)^{\mathrm{T}}\\ \left(c_2^{1,2}\begin{pmatrix}0.3\\0.7\end{pmatrix}+c_2^{2,2}\begin{pmatrix}0.7\\0.3\end{pmatrix}\right)^{\mathrm{T}}\end{pmatrix}\\
&c_1^{1,1}+c_1^{2,1}=1;\ c_1^{1,2}+c_1^{2,2}=1;\ c_1^{1,3}+c_1^{2,3}=1;\\
&c_2^{1,1}+c_2^{2,1}=1;\ c_2^{1,2}+c_2^{2,2}=1;\ c_2^{1,3}+c_2^{2,3}=1;\\
&c_1^{1,1}\ge 0;\ c_1^{2,1}\ge 0;\ c_1^{1,2}\ge 0;\ c_1^{2,2}\ge 0;\ c_1^{1,3}\ge 0;\ c_1^{2,3}\ge 0;\\
&c_2^{1,1}\ge 0;\ c_2^{2,1}\ge 0;\ c_2^{1,2}\ge 0;\ c_2^{2,2}\ge 0;\ c_2^{1,3}\ge 0;\ c_2^{2,3}\ge 0
\end{aligned}
\tag{4.51}
$$

i.e.:

$$
\begin{aligned}
&\textit{Subject to}\\
&\mathbf{P}=\begin{pmatrix}\left(0.5c_1^{1,1}+0.8c_1^{2,1}\right)\left(0.3c_2^{1,3}+0.7c_2^{2,3}\right) & \left(0.5c_1^{1,2}+0.8c_1^{2,2}\right)\left(0.7c_2^{1,3}+0.3c_2^{2,3}\right)\\ \left(0.5c_1^{1,1}+0.2c_1^{2,1}\right)\left(0.3c_2^{1,3}+0.7c_2^{2,3}\right) & \left(0.5c_1^{1,2}+0.2c_1^{2,2}\right)\left(0.7c_2^{1,3}+0.3c_2^{2,3}\right)\end{pmatrix}\\
&\mathbf{P}=\begin{pmatrix}\left(0.5c_1^{1,3}+0.8c_1^{2,3}\right)\left(0.3c_2^{1,1}+0.7c_2^{2,1}\right) & \left(0.5c_1^{1,3}+0.8c_1^{2,3}\right)\left(0.7c_2^{1,1}+0.3c_2^{2,1}\right)\\ \left(0.5c_1^{1,3}+0.2c_1^{2,3}\right)\left(0.3c_2^{1,2}+0.7c_2^{2,2}\right) & \left(0.5c_1^{1,3}+0.2c_1^{2,3}\right)\left(0.7c_2^{1,2}+0.3c_2^{2,2}\right)\end{pmatrix}\\
&c_1^{1,1}+c_1^{2,1}=1;\ c_1^{1,2}+c_1^{2,2}=1;\ c_1^{1,3}+c_1^{2,3}=1;\\
&c_2^{1,1}+c_2^{2,1}=1;\ c_2^{1,2}+c_2^{2,2}=1;\ c_2^{1,3}+c_2^{2,3}=1;\\
&c_1^{1,1}\ge 0;\ c_1^{2,1}\ge 0;\ c_1^{1,2}\ge 0;\ c_1^{2,2}\ge 0;\ c_1^{1,3}\ge 0;\ c_1^{2,3}\ge 0;\\
&c_2^{1,1}\ge 0;\ c_2^{2,1}\ge 0;\ c_2^{1,2}\ge 0;\ c_2^{2,2}\ge 0;\ c_2^{1,3}\ge 0;\ c_2^{2,3}\ge 0
\end{aligned}
\tag{4.52}
$$

The upper and lower probabilities for the event $T$={(A, A), (B, B)} are equal to 19/59 and 40/59, respectively. These bounds are tighter than those in Example 4.5 because now additional contraints on **p** have been added, reflecting the fact that the second experiment is epistemically irrelevant to the first one. This differs from the results in Example 3-1 (unknown interaction) where learning the resin type of either cartridge made our probabilities (of the resin type of the other cartridge) vacuous.

The solutions summarized in Table 4.11 indicate that the minimizing $\mathbf{p}_1^{(i)}$ all differ from one another and that the $\mathbf{p}_2^{(i)}$ all differ from one another as well, i.e. they do not satisfy stochastic independence (Eq. (2.25)). Likewise for the maximizing solution.

**Table 4.11** Example 4.6: Solutions of the optimization problems for the lower and upper probabilities for $T$

| Solution for | Joint $\mathbf{P}$ | $\left(c_1^{1,1}, c_1^{2,1}\right)$ $(\mathbf{p}_1^{(1)})^T$ | $\left(c_1^{1,2}, c_1^{2,2}\right)$ $(\mathbf{p}_1^{(2)})^T$ | $\left(c_1^{1,3}, c_1^{2,3}\right)$ $(\mathbf{p}_1^{(3)})^T$ | Marginal on $S_1$ |
|---|---|---|---|---|---|
| Min | $\begin{pmatrix} 12 & 28 \\ 12 & 7 \end{pmatrix}/59$ | (1, 0)<br>(0.5, 0.5) | (0, 1)<br>(0.8, 0.2) | (24, 35)/59<br>(40, 19)/59 | $(40/59, 19/59)^T$ |
| Max | $\begin{pmatrix} 28 & 12 \\ 7 & 12 \end{pmatrix}/59$ | (0, 1)<br>(0.8, 0.2) | (1, 0)<br>(0.5, 0.5) | (24, 35)/59<br>(40, 19)/59 | $(40/59, 19/59)^T$ |

| Solution for | $\left(c_2^{1,1}, c_2^{2,1}\right)$ $(\mathbf{p}_2^{(1)})^T$ | $\left(c_2^{1,2}, c_2^{2,2}\right)$ $(\mathbf{p}_2^{(2)})^T$ | $\left(c_2^{1,3}, c_2^{2,3}\right)$ $(\mathbf{p}_2^{(3)})^T$ | Marginal on $S_2$ |
|---|---|---|---|---|
| Min | (173, 63)/236<br>(24, 35)/59 | (1, 0)<br>(0.3, 0.7) | (13, 63)/76<br>(12, 7)19 | $(24/59, 35/24)^T$ |
| Max | (63, 173)/263<br>(35, 24)/59 | (0, 1)<br>(0.3, 0.7) | (63, 13)/79<br>(7, 12)/19 | $(35/59, 24/59)^T$ |

Indeed in epistemic independence, if we learn that the actual value of $s_2$ is $s_2$*, then the probability measure for $s_1$ is again one of the probability measures in $\Psi_1$, but in general not always the same for different values $s_2$*; and vice versa. Strong independence (dealt with in the next section) imposes that the probability measures be the same.

As in the case of epistemic irrelevance in Example 4.5:

- Both the marginal on $S_1$ and the marginal on $S_2$ must be calculated by using the marginal rule (Eq. (2.24))
- It may happen that the marginal on $S_1$ ($S_2$) is different from $\mathbf{p}_1^{(i)}$ ($\mathbf{p}_2^{(i)}$).

In contrast to Example 4.5, each experiment is epistemically irrelevant to the other experiment. As a consequence, the conditional upper and lower probabilities that the first (second) resin is Type A given the type of the second (first) resin are equal to the marginal ones, i.e. 0.8 and 0.5 (0.7 and 0.3), respectively. The conditional upper and lower probabilities that the first (second) resin is Type A given the type of the second (first) resin are obtained by minimizing and maximizing $p^{1,1}/(p^{1,1}+p^{2,1})$ ($p^{1,1}/(p^{1,1}+p^{1,2})$). The constraints are still given by Eq. (4.52). The results are summarized in Table 4.12 for the conditional upper and lower probabilities that the first resin is Type A given the type of the second resin. They indicate that the minimizing and maximizing solutions again violate stochastic independence (Eq. (2.35)) because minimizing $\mathbf{p}_1^{(i)}$ are all different from one another and that the $\mathbf{p}_2^{(i)}$ also all differ from one another. Likewise for the maximizing solution.

**Table 4.12** Example 4.6: Solutions of the optimization problems for the lower and upper probabilities that the first resin is Type A given the type of the second resin

| Solution for | Joint $\mathbf{P}$ | $\left(c_1^{1,1}, c_1^{2,1}\right)$ $(\mathbf{p}_1^{(1)})^T$ | $\left(c_1^{1,2}, c_1^{2,2}\right)$ $(\mathbf{p}_1^{(2)})^T$ | $\left(c_1^{1,3}, c_1^{2,3}\right)$ $(\mathbf{p}_1^{(3)})^T$ | Marginal on $S_1$ |
|---|---|---|---|---|---|
| Min | $\begin{pmatrix} 5777 & 2943 \\ 5777 & 2703 \end{pmatrix}/17200$ | (1, 0)<br>(0.5, 0.5) | (2623, 200) /2823<br>(981, 901) /1882 | (42, 1)/43<br>(109, 106) /215 | (8720, 8480)$^T$/17200 |
| Max | $\begin{pmatrix} 428 & 212 \\ 107 & 107 \end{pmatrix}/854$ | (0, 1)<br>(0.8, 0.2) | (144, 175) /319<br>(212, 107) /319 | (72, 35) /427<br>(320, 107) /427 | (640, 214)$^T$/854 |

| Solution for | $\left(c_2^{1,1}, c_2^{2,1}\right)$ $(\mathbf{p}_2^{(1)})^T$ | $\left(c_2^{1,2}, c_2^{2,2}\right)$ $(\mathbf{p}_2^{(2)})^T$ | $\left(c_2^{1,3}, c_2^{2,3}\right)$ $(\mathbf{p}_2^{(3)})^T$ | Marginal on $S_2$ |
|---|---|---|---|---|
| Min | (243, 3197)/ 3440<br>(5777, 2823) /8600 | (3, 29)/32<br>(27, 53) /80 | (3, 61)/64<br>(109, 51) /160 | (11554, 5646)$^T$/17200 |
| Max | (157, 697/854<br>(535, 319) /854 | (5, 59) /64<br>(107, 53) /160 | (0.5, 0.5)<br>(0.5, 0.5) | (535, 319)$^T$/854 |

**Example 4.7.** Consider again the situation and knowledge available in the four-dimensional case of Example 4.4. Our knowledge about $S_1$ and $S_2$ is the same as in Example 4.4, but assume now the experiment of picking an element in S2 is epistemically irrelevant to the experiment of picking an element in $S_1$. We want to write down the optimization problems for finding upper and lower expectations on the joint space and then calculate the upper and lower probabilities for the event $T = \left\{\left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right), \left(s_1^4, s_2^4\right)\right\}$. Finally, we calculate the conditional upper and lower probabilities $P(s_2^1 | s_1^1)$, and contrast to the conditional upper and lower probabilities $P(s_1^1 | s_2^1)$.

Constraints in Eq. (4.45) become:

$$\begin{aligned}
&\textit{Subject to}\\
&\mathbf{P} = Diag\left(\mathbf{p}_1^{(1)}\right)\begin{pmatrix} \left(\mathbf{p}_2^{(1)}\right)^T \\ \left(\mathbf{p}_2^{(2)}\right)^T \\ \left(\mathbf{p}_2^{(3)}\right)^T \\ \left(\mathbf{p}_2^{(4)}\right)^T \end{pmatrix}\\
&(1,0,0,0)\ \mathbf{p}_1^{(1)} \le 0.2;\ (0,1,0,0)\ \mathbf{p}_1^{(1)} \le 0.7;\ (0,0,0,1)\ \mathbf{p}_1^{(1)} \le 0.5\\
&(1,1,0,0)\ \mathbf{p}_1^{(1)} \le 0.9;\ 0.2 \le (1,0,1,0)\ \mathbf{p}_1^{(1)};\ 0.4 \le (0,1,1,0)\ \mathbf{p}_1^{(1)}\\
&0 \le (0,0,0,1)\ \mathbf{p}_2^{(1)} \le 0.3;\ 0 \le (0,0,0,1)\ \mathbf{p}_2^{(2)} \le 0.3\\
&0 \le (0,0,0,1)\ \mathbf{p}_2^{(3)} \le 0.3;\ 0 \le (0,0,0,1)\ \mathbf{p}_2^{(4)} \le 0.3\\
&47/50 \le (0,0,1,7/5)\ \mathbf{p}_2^{(1)};\ 47/50 \le (0,0,1,7/5)\ \mathbf{p}_2^{(2)}\\
&47/50 \le (0,0,1,7/5)\ \mathbf{p}_2^{(3)};\ 47/50 \le (0,0,1,7/5)\ \mathbf{p}_2^{(4)}\\
&\mathbf{1}^T \cdot \mathbf{p}_1^{(1)} = 1;\ \mathbf{1}^T \cdot \mathbf{p}_2^{(1)} = 1; \mathbf{1}^T \cdot \mathbf{p}_2^{(2)} = 1;\ ;\ \mathbf{1}^T \cdot \mathbf{p}_2^{(3)} = 1; \mathbf{1}^T \cdot \mathbf{p}_2^{(4)} = 1\\
&\mathbf{p}_1^{(1)} \ge 0; \mathbf{p}_2^{(1)} \ge 0;\ \mathbf{p}_2^{(2)} \ge 0\ ; \mathbf{p}_2^{(3)} \ge 0;\ \mathbf{p}_2^{(4)} \ge 0
\end{aligned} \tag{4.53}$$

The upper and lower probabilities for the event

$$T = \left\{ \left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right), \left(s_1^4, s_2^4\right) \right\}$$

are equal to 0 and 1, respectively. The solutions summarized in Table 4.13 indicate that, for the minimizing solution, $\mathbf{p}_2^{(1)} \neq \mathbf{p}_2^{(2)} \neq \mathbf{p}_2^{(3)} \neq \mathbf{p}_2^{(4)}$, i.e. stochastic independence is not satisfied (Eq. 2.35). Likewise for the maximizing solution. Notice that both the marginal on $S_1$ and the marginal on $S_2$ must be calculated by using the marginal rule (Eq. (2.24)).

**Table 4.13** Example 4.7: Solutions of the nonlinear programming problems for the lower and upper probabilities for $T$

| Solution for | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min | $\begin{pmatrix} 0 & 0.108 & 0.312 & 0.18 \\ 0.033 & 0 & 0.26 & 0.107 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ | $(0.6, 0.4, 0, 0)^{\mathrm{T}}$ | $\begin{pmatrix} 0.033 \\ 0.108 \\ 0.572 \\ 0.287 \end{pmatrix}^{\mathrm{T}}$ |
| Max | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ | $(0, 0, 1, 0)^{\mathrm{T}}$ | $(0, 0, 1, 0))^{\mathrm{T}}$ |

The conditional upper and lower probabilities $P(s_2^1 | s_1^1)$ are equal to the marginal ones, i.e. 0.18 and 0, respectively, while the conditional upper and lower probabilities $P(s_1^1 | s_2^1)$ are equal to 1 and 0, which are larger bounds than the marginal bounds, i.e. 0.2 and 0. This means that the first experiment is epistemically relevant to the second one. The results summarized in Table 4.14 indicate that the minimizing and maximizing solutions again violate stochastic independence (Eq. (2.35)) because $\mathbf{p}_2^{(1)} \neq \mathbf{p}_2^{(2)} \neq \mathbf{p}_2^{(3)} \neq \mathbf{p}_2^{(4)}$.

**Table 4.14** Example 4.7: Solutions of the optimization problems for the conditional upper and lower probabilities $p\left(s_1^1 | s_2^1\right)$

| Solution for | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.085 & 0 & 0.373 & 0.124 \\ 0.024 & 0.005 & 0.127 & 0.044 \\ 0.024 & 0.015 & 0.114 & 0.066 \end{pmatrix}$ | $(0, 0.581, 0.2, 0.219)^{\mathrm{T}}$ | $\begin{pmatrix} 0.133 \\ 0.020 \\ 0.613 \\ 0.234 \end{pmatrix}^{\mathrm{T}}$ |
| Max | $\begin{pmatrix} 0.036 & 0 & 0.104 & 0.06 \\ 0 & 0.077 & 0.398 & 0.106 \\ 0 & 0 & 0 & 0 \\ 0 & 0.038 & 0.120 & 0.061 \end{pmatrix}$ | $(0.2, 0.581, 0, 0.219)^{\mathrm{T}}$ | $\begin{pmatrix} 0.036 \\ 0.115 \\ 0.622 \\ 0.227 \end{pmatrix}^{\mathrm{T}}$ |

**Example 4.8.** Consider again the situation and knowledge available in Example 4.4. In addition to the knowledge available in Example 4.4, all we now know about the stochastic mechanism between $S_1$ and $S_2$, indicating epistemic independence. We want to write down the optimization problems for finding upper and lower expectations on the joint space and then calculate the upper and lower probabilities for the event $T = \left\{ \left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right), \left(s_1^4, s_2^4\right) \right\}$. Finally, we calculate the conditional upper and lower probabilities $P(s_2^1 | s_1^1)$ and $P(s_1^1 | s_2^1)$.

Since this is a case of epistemic independence, where each experiment is epistemically irrelevant to the other, this example is the symmetric counterpart to Example 4.7 above, where only the first experiment was epistemically irrelevant to the other. Constraints in Eqs. (4.45) and (4.46) become:

*Subject to*

$$\mathbf{P} = Diag\left(\mathbf{p}_1^{(1)}\right) \begin{pmatrix} \left(\mathbf{p}_2^{(1)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2^{(2)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2^{(3)}\right)^{\mathrm{T}} \\ \left(\mathbf{p}_2^{(4)}\right)^{\mathrm{T}} \end{pmatrix}; \ \mathbf{P} = \left(\mathbf{p}_2^{(2)} \quad \mathbf{p}_2^{(3)} \quad \mathbf{p}_2^{(4)} \quad \mathbf{p}_2^{(5)}\right) Diag\left(\mathbf{p}_2^{(5)}\right) \tag{4.54}$$

$$\left(1,0,0,0\right)\ \mathbf{p}_1^{(i)} \le 0.2;\ \left(0,1,0,0\right)\ \mathbf{p}_1^{(i)} \le 0.7;\ \left(0,0,0,1\right)\ \mathbf{p}_1^{(i)} \le 0.5$$
$$\left(1,1,0,0\right)\ \mathbf{p}_1^{(i)} \le 0.9;\ 0.2 \le \left(1,0,1,0\right)\ \mathbf{p}_1^{(i)};\ 0.4 \le \left(0,1,1,0\right)\ \mathbf{p}_1^{(i)};\ i = 1,\ldots 5$$
$$0 \le \left(0,0,0,1\right)\ \mathbf{p}_2^{(j)} \le 0.3;\ 47/50 \le \left(0,0,1,7/5\right)\ \mathbf{p}_2^{(j)};\ \mathrm{j} = 1,\ldots 5$$
$$\mathbf{1}^{\mathrm{T}} \cdot \mathbf{p}_1^{(i)} = 1;\ \mathbf{1}^{\mathrm{T}} \cdot \mathbf{p}_2^{(j)} = 1$$
$$\mathbf{p}_1^{(i)} \ge 0;\ \mathbf{p}_2^{(j)} \ge 0$$

**Table 4.15** Example 4.7: Solutions of the nonlinear programming problems for the lower and upper probabilities for $T$

| Solution for | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min | $\begin{pmatrix} 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0 \end{pmatrix}$ | $(0.2, 0.4, 0, 0.4)^{\mathrm{T}}$ | $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}^{\mathrm{T}}$ |
| Max | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ | $(0, 0, 1, 0)^{\mathrm{T}}$ | $(0, 0, 1, 0))^{\mathrm{T}}$ |

In this example of epistemic independence, the upper and lower probabilities for the event $T = \left\{\left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right), \left(s_1^4, s_2^4\right)\right\}$ are equal to 0 and 1, respectively. The solutions summarized in Table 4.13 still indicate that the joint distribution does not satisfy stochastic independence (Eq. (2.35)). Likewise for the maximizing solution.

The conditional upper and lower probabilities $P(s_2^{\,1} | \, s_1^{\,1})$ are equal to the marginal ones, i.e. 0.18 and 0, respectively, and the conditional upper and lower probabilities $P(s_1^{\,1} | \, s_2^{\,1})$ are equal to 0.2 and 0, which means that the first experiment is epistemically relevant to the second one and vice versa. It should be noted that the minimizing and maximizing solutions in the case of epistemic independence again violate stochastic independence (Eq. (2.35)) because $\mathbf{p}_\mathrm{i}^{(1)} \neq \mathbf{p}_\mathrm{i}^{(2)} \neq \mathbf{p}_\mathrm{i}^{(3)} \neq \mathbf{p}_\mathrm{i}^{(4)} \neq \mathbf{p}_\mathrm{i}^{(5)}$, $i = 1, 2$.

Since the optimization problems contained in the previous examples are NP-hard, regardless of the numerical method applied, finding the optimal solution is time-consuming even when the problems are 4-dimensional. When marginals' expectations are bounded (Eq. (4.17)), these non-linear optimization problems may be turned into linear ones by starting from the joint distribution matrix $\mathbf{P}$.

Given $\mathbf{P} = Diag(\mathbf{p}_1)\, \mathbf{P}_{2|1}$ (Eq. (4.17), the marginal on $S_1$, $\mathbf{P} \cdot \mathbf{1}_{(n_2)}$, is

$$\begin{aligned} \mathbf{P} \cdot \mathbf{1}_{(n_2)} &= Diag\left(\mathbf{p}_1\right) \mathbf{P}_{2|1} \cdot \mathbf{1}_{(n_2)} \\ &= Diag\left(\mathbf{p}_1\right) \cdot \left(\mathbf{P}_{2|1} \mathbf{1}_{(n_2)}\right) = Diag\left(\mathbf{p}_1\right) \cdot \mathbf{1}_{(n_2)} = \mathbf{p}_1 \end{aligned} \tag{4.55}$$

i.e., the marginal on $S_1$ is the same as $\mathbf{p}_1$. Consequently, by substituting $\mathbf{P} \cdot \mathbf{1}_{(n_2)}$ for $\mathbf{p}_1$ one obtains:

$$\mathbf{P} = Diag\left(\mathbf{p}_1\right) \cdot \mathbf{P}_{2|1} = Diag\left(\mathbf{P} \cdot \mathbf{1}_{(n_2)}\right) \cdot \mathbf{P}_{2|1} \tag{4.56}$$

$$\text{i.e., } \mathbf{P}^{j,\cdot} = \left(\sum_{m=1}^{n_2} p^{j,m}\right) \left(\mathbf{p}_2^{(j)}\right)^{\mathrm{T}};\; j = 1, \ldots, n_1 \tag{4.57}$$

where $\mathbf{P}^{j,\cdot}$ is the $j^{\text{th}}$ row of matrix $\mathbf{P}$.

Therefore, the constraints on $\mathbf{p}_1$, i.e. $E_{LOW}\left[f_1^k\right] \leq \left(\mathbf{f}_1^k\right)^{\mathrm{T}} \mathbf{p}_1 \leq E_{UPP}\left[f_1^k\right]$, can be rewritten as $E_{LOW}\left[f_1^k\right] \leq \left(\mathbf{f}_1^k\right)^{\mathrm{T}} \mathbf{P} \cdot \mathbf{1}_{(n_2)} \leq E_{UPP}\left[f_1^k\right]$. As for constraints on $\mathbf{p}_2$, i.e.:

$$E_{LOW}\left[f_2^k\right] \leq \left(\mathbf{f}_2^k\right)^{\mathrm{T}} \mathbf{p}_2^{(j)} \leq E_{UPP}\left[f_2^k\right] \tag{4.58}$$

notice that $\mathbf{p}_2^{(j)} \cdot \left( \sum_{m=1}^{n_2} p^{j,m} \right) = \left( \mathbf{P}^{j,\cdot} \right)^{\mathrm{T}}$. Therefore, by multiplying Eq. (4.58) by $\sum_{m=1}^{n_2} p^{j,m}$, constraints on $\mathbf{p}_2$ are rewritten as:

$$E_{LOW}\left[ f_2^k \right] \cdot \left( \sum_{m=1}^{n_2} p^{j,m} \right) \le \mathbf{P}^{j,\cdot} \cdot \mathbf{f}_2^k \le E_{UPP}\left[ f_2^k \right] \cdot \left( \sum_{m=1}^{n_2} p^{j,m} \right) \qquad (4.59)$$

To summarize, given the epistemic irrelevance of the first experiment with respect to the second experiment, $\mathbf{P} = Diag(\mathbf{p}_1)\,\mathbf{P}_{2|1}$, the optimization problem may be written in the linear form:

$$\begin{aligned}
&\text{minimize } \sum_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j} \quad \left( -\sum_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j} \right) \\
&\textit{subject to} \\
&\quad E_{LOW}\left[ f_1^k \right] \le \left( \mathbf{f}_1^k \right)^{\mathrm{T}} \ \mathbf{P} \cdot \mathbf{1}_{(n_2)} \le E_{UPP}\left[ f_1^k \right];\ k = 1,\ldots,k_1; \\
&\quad E_{LOW}\left[ f_2^k \right] \cdot \left( \sum_{m=1}^{n_2} p^{j,m} \right) \le \mathbf{P}^{j,\cdot} \cdot \mathbf{f}_2^k \le E_{UPP}\left[ f_2^k \right] \cdot \left( \sum_{m=1}^{n_2} p^{j,m} \right); \\
&\qquad\qquad\qquad\qquad k = 1,\ldots,k_2;\ j = 1,\ldots,n_1; \\
&\quad \mathbf{1}^{\mathrm{T}} \cdot \mathbf{P} \cdot \mathbf{1} = 1;\ \ \mathbf{P} \ge 0
\end{aligned} \qquad (4.60)$$

where $\mathbf{P}^{j,\cdot}$ is the $j^{\text{th}}$ row of matrix $\mathbf{P}$.

Constraints in Eq. (4.60) are equivalent to the constraints on $\Psi_1$ (Eq. (4.17)). If the set of extreme distributions on the marginals (Eq. (4.18)) is given, constraints remain quadratic even if they are rewritten as in Eq. (4.60):

$$\begin{aligned}
&\mathbf{P} \cdot \mathbf{1}_{(n_2)} = \sum_{\xi=1}^{\xi_1} c_1^{\xi} \mathbf{p}_{EXT_1}^{\xi}; \\
&\mathbf{P}^{j,\cdot} = \left( \sum_{\xi=1}^{\xi_2} c_2^{\xi,j} \mathbf{p}_{EXT_2}^{\xi} \right) \cdot \left( \sum_{m=1}^{n_2} p^{j,m} \right) \\
&\sum_{\xi=1}^{\xi_1} c_1^{\xi} = 1,\ c_1^{\xi} \ge 0; \\
&\sum_{\xi=1}^{\xi_2} c_2^{\xi,j} = 1,\ c_2^{\xi,j} \ge 0,\ j = 1,\ldots,n_1;
\end{aligned} \qquad (4.61)$$

where $\mathbf{P}^{j,\cdot}$ is the $j^{\text{th}}$ row of matrix $\mathbf{P}$.

Therefore, constraints in Eq. (4.61) have no computational advantage over constraints in Eq. (4.46).

Likewise, when the second experiment is epistemically irrelevant to the first, $\mathbf{P} = \mathbf{P}_{1|2}Diag(\mathbf{p}_2)$ (i.e.), and $\mathbf{1}_{(n_1)}^{\mathrm{T}}\mathbf{P} = \mathbf{p}_2$, i.e., the marginal on $S_2$ is equal to $\mathbf{p}_2$. The linear optimization problem is

$$\begin{aligned}
&Minimize \sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j} \quad \left(-\sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j}\right) \\
&Subject\ to \\
&\quad E_{LOW}\left[f_1^k\right]\cdot\left(\sum_{m=1}^{n_1} p^{m,j}\right) \le \left(\mathbf{f}_1^k\right)^{\mathrm{T}} \mathbf{P}^{\cdot,j} \le E_{UPP}\left[f_1^k\right]\cdot\left(\sum_{m=1}^{n_2} p^{m,j}\right); \\
&\qquad\qquad k=1,\ldots,k_1;\ j=1,\ldots,n_2; \\
&\quad E_{LOW}\left[f_2^k\right] \le \mathbf{1}_{(n_1)}^{\mathrm{T}} \mathbf{P}\cdot\mathbf{f}_2^k \le E_{UPP}\left[f_2^k\right];\ k=1,\ldots,k_2; \\
&\quad \mathbf{1}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{1}=1;\ \ \mathbf{P}\ge 0
\end{aligned} \tag{4.62}$$

where $\mathbf{P}^{\cdot,j}$ is the $j^{\text{th}}$ column of matrix $\mathbf{P}$.

For the case of epistemic independence, let us write down the optimization problem in the (linear) constraints when marginals are constrained by upper and lower previsions (Eq. (4.18)):

$$\begin{aligned}
&Minimize \sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j} \quad \left(-\sum\nolimits_{i=1;j=1}^{i=n_1;j=n_2} a^{i,j} p^{i,j}\right) \\
&Subject\ to \\
&\quad E_{LOW}\left[f_1^k\right] \le \left(\mathbf{f}_1^k\right)^{\mathrm{T}} \mathbf{P}\cdot\mathbf{1}_{(n_2)} \le E_{UPP}\left[f_1^k\right];\ k=1,\ldots,k_1; \\
&\quad E_{LOW}\left[f_2^k\right] \le \left(\mathbf{1}_{(n_1)}\right)^{\mathrm{T}} \mathbf{P}\cdot\mathbf{f}_2^k \le E_{UPP}\left[f_2^k\right];\ k=1,\ldots,k_2; \\
&\quad E_{LOW}\left[f_1^k\right]\cdot\left(\sum_{m=1}^{n_1} p^{m,j}\right) \le \left(\mathbf{f}_1^k\right)^{\mathrm{T}} \mathbf{P}^{\cdot,j} \le E_{UPP}\left[f_1^k\right]\cdot\left(\sum_{m=1}^{n_2} p^{m,j}\right); \\
&\qquad\qquad k=1,\ldots,k_1;\ j=1,\ldots,n_2; \\
&\quad E_{LOW}\left[f_2^k\right]\cdot\left(\sum_{m=1}^{n_2} p^{j,m}\right) \le \mathbf{P}^{j,\cdot}\cdot\mathbf{f}_2^k \le E_{UPP}\left[f_2^k\right]\cdot\left(\sum_{m=1}^{n_2} p^{j,m}\right); \\
&\qquad\qquad k=1,\ldots,k_2;\ j=1,\ldots,n_1; \\
&\quad \mathbf{1}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{1}=1;\ \ \mathbf{P}\ge 0
\end{aligned} \tag{4.63}$$

When marginals are bounded by extreme distributions (Eq. (4.46)), the constraints are rewritten as:

$$\begin{aligned}
&\mathbf{P}\cdot\mathbf{1}_{(n_2)}=\sum_{\xi=1}^{\xi_1}c_1^{\xi}\mathbf{p}_{EXT_1}^{\xi};\\
&\mathbf{P}^{\mathrm{T}}\cdot\mathbf{1}_{(n_1)}=\sum_{\xi=1}^{\xi_2}c_2^{\xi}\mathbf{p}_{EXT_2}^{\xi};\\
&\mathbf{P}^{\cdot,j}=\left(\sum_{\xi=1}^{\xi_1}c_1^{\xi,j}\mathbf{p}_{EXT_1}^{\xi}\right)\cdot\left(\sum_{m=1}^{n_1}p^{m,j}\right)\\
&\mathbf{P}^{j,\cdot}=\left(\sum_{\xi=1}^{\xi_2}c_2^{\xi,j}\mathbf{p}_{EXT_2}^{\xi}\right)\cdot\left(\sum_{m=1}^{n_2}p^{j,m}\right)\\
&\sum_{\xi=1}^{\xi_1}c_1^{\xi}=1,\ c_1^{\xi}\geq 0;\\
&\sum_{\xi=1}^{\xi_2}c_2^{\xi,j}=1,\ c_2^{\xi,j}\geq 0,\ j=1,\ldots,n_1;\\
&\sum_{\xi=1}^{\xi_1}c_1^{\xi,j}=1,\ c_1^{\xi,j}\geq 0,\ j=1,\ldots,n_2;\\
&\sum_{\xi=1}^{\xi_2}c_2^{\xi}=1,\ c_2^{\xi}\geq 0
\end{aligned}\tag{4.64}$$

Similar to the case of epistemic irrelevance, these quadratic constraints make the epistemic independence problem non-linear, and there is no computational advantage with respect to Eq. (4.46)

Both $\mathbf{P} = Diag(\mathbf{p}_1)\,\mathbf{P}_{2|1}$ and $\mathbf{P} = \mathbf{P}_{1|2}Diag(\mathbf{p}_2)$ ensure that the constraints on the marginal distributions are satisfied automatically. In the case of $\mathbf{P} = Diag(\mathbf{p}_1)\,\mathbf{P}_{2|1}$, we already know that the marginal on $S_1$, $\mathbf{P}\cdot\mathbf{1}_{(n_2)}$, is equal to $\mathbf{p}_1$; now let us check the marginal on $S_2$, i.e. $\mathbf{1}_{(n_1)}^{\mathrm{T}}\mathbf{P}$.

$$\mathbf{1}_{(n_1)}^{T}\mathbf{P}=\mathbf{1}_{(n_1)}^{\mathrm{T}}Diag(\mathbf{p}_1)\mathbf{P}_{2|1}=(\mathbf{p}_1)^{\mathrm{T}}\mathbf{P}_{2|1}=\left(p_1(s_1)\ldots p_1\left(s_{n_1}\right)\right)\begin{pmatrix}\left(\mathbf{p}_2^{(1)}\right)^{\mathrm{T}}\\ \vdots\\ \left(\mathbf{p}_2^{(n_1)}\right)^{\mathrm{T}}\end{pmatrix}\tag{4.65}$$

This is a linear combination of vectors $\mathbf{p}_2^{(i)}$ with coefficients $p_1(s_i)$. Since $0\leq p_1(s_i)\leq 1,\ i=1,\ldots,n_1$, and $\sum_{i=1}^{n_1}p_1(s_i)=1$, the marginal on $S_2$ $\mathbf{1}_{(n_1)}^{\mathrm{T}}\mathbf{P}$ is a convex combination of elements $\mathbf{p}_2\in\Psi_2$. Since $\Psi_2$ is convex, the marginal on $S_2$ is in $\Psi_2$. Therefore, the constraints on marginal probabilities on joint space will be automatically satisfied for the epistemic irrelevance problem.

Now, let us go back to the definition of epistemic irrelevance to introduce an effective algorithm for calculating the extreme joint distributions, which is addressed by the following two theorems:

**Theorem 4.2.** *Let the extreme points of the convex sets of marginal probability distributions on* $S_1$ *and* $S_2$ *be* $\mathbf{p}^{\xi}_{EXT_1}$, $\xi = 1,...\xi_1$, *and* $\mathbf{p}^{\xi}_{EXT_2}$, $\xi = 1,...\xi_2$, *respectively. If the first experiment is epistemically irrelevant to the second, the set of extreme points of the joint distributions,* $\Psi_E^{|s_1}$, *is*:

$$EXT = \left\{ \mathbf{P} = Diag\left(\mathbf{p}_1\right)\cdot \mathbf{P}_{2|1} : \mathbf{p}_1 \in EXT_1, \mathbf{P}_{2|1}^{(i)} \in EXT_2; i = 1,...n_1 \right\}, \text{ i.e.}$$

$$EXT = \left\{ \mathbf{P}_{EXT} = Diag\left(\mathbf{p}^{m}_{EXT_1}\right) \begin{pmatrix} \left(\mathbf{p}^{\eta_1}_{EXT_2}\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}^{\eta_{n_1}}_{EXT_2}\right)^{\mathrm{T}} \end{pmatrix} : \mathbf{p}^{m}_{EXT_1} \in EXT_1, \mathbf{p}^{\eta_i}_{EXT_2} \in EXT_2;\ i = 1,...n_1 \right\} \tag{4.66}$$

*Proof* : Any $\mathbf{p}_1 \in \Psi_1$ and $\mathbf{p}_2 \in \Psi_2$ can be written as a linear combination of extreme points in $\Psi_1$ and $\Psi_2$, respectively:

$$\mathbf{p}_1 = \left(\mathbf{p}^{1}_{EXT_1} \cdots \mathbf{p}^{\xi}_{EXT_1} \cdots \mathbf{p}^{\xi_1}_{EXT_1}\right)\left(\lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1}\right)^{\mathrm{T}} \tag{4.67}$$

$$\begin{aligned} &\mathbf{P}_{2|1} = \begin{bmatrix} \lambda_2^{1,1} & \ldots & \lambda_2^{\xi_2,1} \\ \vdots & \lambda_2^{\xi,i} & \vdots \\ \lambda_2^{1,n_1} & \cdots & \lambda_2^{\xi_2,n_1} \end{bmatrix} \begin{pmatrix} \left(\mathbf{p}^{1}_{EXT_2}\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}^{\xi_2}_{EXT_2}\right)^{\mathrm{T}} \end{pmatrix} \\ &0 \le \lambda_1^{\xi} \le 1,\ \xi = 1,...,\xi_1;\ \sum_{\xi=1}^{\xi_1} \lambda_1^{\xi} = 1 \\ &0 \le \lambda_2^{\xi,i} \le 1,\ \xi = 1,...,\xi_2,\ i = 1,...,n_1;\ \sum_{\xi=1}^{\xi_2} \lambda_2^{\xi,i} = 1 \end{aligned} \tag{4.68}$$

By inserting Eqs. (4.67) and (4.68) into $\mathbf{P} = Diag(\mathbf{p}_1) \cdot \mathbf{P}_{2|1}$, one obtains:

$$\mathbf{P} = Diag\left( \left( \mathbf{p}_{EXT_1}^1 \ldots \mathbf{p}_{EXT_1}^{\xi} \ldots \mathbf{p}_{EXT_1}^{\xi_1} \right) \begin{pmatrix} \lambda_1^1 \\ \vdots \\ \lambda_1^{\xi} \\ \vdots \\ \lambda_1^{\xi_1} \end{pmatrix} \right) \begin{bmatrix} \lambda_2^{1,1} & \ldots & \lambda_2^{\xi_2,1} \\ \vdots & \lambda_2^{\xi,i} & \vdots \\ \lambda_2^{1,n_1} & \ldots & \lambda_2^{\xi_2,n_1} \end{bmatrix} \begin{pmatrix} \left(\mathbf{p}_{EXT_2}^1\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}_{EXT_2}^{\xi}\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}_{EXT_2}^{\xi_2}\right)^{\mathrm{T}} \end{pmatrix} \tag{4.69}$$

Extreme points of $\mathbf{P}$ are achieved if and only if $\lambda_1^{\xi} = \begin{cases} 1, & \xi = m \\ 0, & \xi \neq m \end{cases}$ and $\lambda_2^{\xi,i} = \begin{cases} 1, & \xi = \eta_i \\ 0, & \xi \neq \eta_i \end{cases}$, $m = 1, \ldots, \xi_1, \eta_i = 1, \ldots, \xi_2$.

Therefore, $\mathbf{P}_{EXT} = Diag\left(\mathbf{p}_{EXT_1}^m\right) \begin{pmatrix} \left(\mathbf{p}_{EXT_2}^{\eta_1}\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}_{EXT_2}^{\eta_{n_1}}\right)^{\mathrm{T}} \end{pmatrix}$ ■

**Theorem 4.3.** *If the second experiment is epistemically irrelevant to the first, the set of extreme points of the joint distributions is*

$$EXT = \left\{ \mathbf{P} = \mathbf{P}_{1|2} \cdot Diag(\mathbf{p}_2) : \mathbf{P}_{1|2}^{(i)} \in EXT_1, \mathbf{p}_2 \in EXT_2; i = 1, \ldots n_2 \right\}, \text{ i.e.} \tag{4.70}$$

$$EXT = \left\{ \begin{array}{c} \mathrm{P}_{EXT} = \left( \mathbf{p}_{EXT_1}^{\eta_1} \quad \cdots \quad \mathbf{p}_{EXT_1}^{\eta_{n_2}} \right) Diag\left(\mathbf{p}_{EXT_2}^m\right) : \\ \mathbf{p}_{EXT_1}^{\eta_i} \in EXT_1, \mathbf{p}_{EXT_2}^m \in EXT_2; i = 1, \ldots n_2 \end{array} \right\} \tag{4.71}$$

Theorem 4.2 enables us to efficiently find the extreme joint distributions given the extreme distributions on the marginals, and the upper limit for the number of extreme joint distribution is $\xi_1 \times \xi_2^{\ n_1}$. Likewise, when the second experiment is epistemically irrelevant to the first, Theorem 4.3 yields an upper limit equal to $\xi_1^{\ n_2} \times \xi_2$. However, the algorithms in Theorem 4.2 and Theorem 4.3 cannot be used in the case of epistemic independence because the convex set of joint distributions is the intersection of the two convex sets for the epistemic irrelevance cases, i.e. $\Psi_E = \Psi_E^{|S_1} \cap \Psi_E^{|S_2}$ ; as illustrated in Figure 4.10, the extreme points of $\Psi_E$ may not be the extreme points of $\Psi_E^{|S_1}$ and $\Psi_E^{|S_2}$ . The only way to determine the extreme points for $\Psi_E$ is to use the linear constraints in Eq.(4.63).

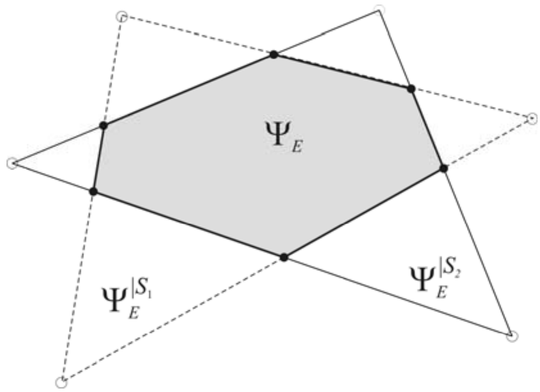



**Fig. 4.10** Set of extreme points for the case of epistemic independence

Now let us rework Example 4.5 and Example 4.6 with the new linear algorithm (Eqs.(4.60), (4.62) and (4.63)), in which constraints are written in terms of the joint distribution. Since the problem is now linear, it may be solved with two different methods. One is a linear optimization problem, and the other consists of first finding all extreme joint distributions and then checking which extreme distribution maximizes or minimizes the objective function. We want to first calculate the probability for the event $T = (\{A, A\}, \{B, B\})$, and then determine the upper and lower conditional probabilities that the first resin is Type A given the type of the second resin, defined by $p^{1,1} / \left(p^{1,1} + p^{2,1}\right)$.

**Example 4.9.** Let us redo Example 4.5 by rewriting the constraints as:

$$\begin{aligned}
&\textit{Subject to}\\
&0.5 \le p^{1,1} + p^{1,2} \le 0.8\\
&0.3\left(p^{1,1} + p^{1,2}\right) \le p^{1,1} \le 0.7\left(p^{1,1} + p^{1,2}\right)\\
&0.3\left(p^{2,1} + p^{2,2}\right) \le p^{2,1} \le 0.7\left(p^{2,1} + p^{2,2}\right)\\
&\mathbf{1}^{\mathrm{T}} \cdot \mathbf{P} \cdot \mathbf{1} = 1;\\
&\mathbf{P} \ge 0
\end{aligned} \tag{4.72}$$

Table 4.16 gives the results obtained by using linear optimization. Both the upper and lower probabilities of $T$ and the conditional upper and lower probabilities (Table 4.17) achieve the same upper and lower values as in our previous calculations in Example 4.5 but the computational effort is now greatly reduced.

**Table 4.16** Example 4.5: lower and upper probabilities for $T$ by the simplex method

| Solution for | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min $P$ (T) = 0.3 | $\begin{pmatrix} 15 & 35 \\ 35 & 15 \end{pmatrix}/100$ | $(1/2, 1/2)^T$ | $(1/2, 1/2)^T$ |
| Max $P$ (T) = 0.7 | $\begin{pmatrix} 35 & 15 \\ 15 & 35 \end{pmatrix}/100$ | $(1/2, 1/2)^T$ | $(1/2, 1/2)^T$ |

**Table 4.17** Example 4.5: lower and upper probabilities for conditional probability $P_{1\|2}\left(S_1 = A \mid S_2 = A\right)$ by the simplex method

| Solution for | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min $P_{1\|2}\left(S_1 = A \mid S_2 = A\right) = 0.3$ | $\begin{pmatrix} 15 & 35 \\ 35 & 15 \end{pmatrix}/100$ | $(1/2, 1/2)^T$ | $(1/2, 1/2)^T$ |
| Max $P_{1\|2}\left(S_1 = A \mid S_2 = A\right) = 0.903$ | $\begin{pmatrix} 0.56 & 0.24 \\ 0.06 & 0.14 \end{pmatrix}$ | $(0.8, 0.2)^T$ | $(0.62, 0.38)^T$ |

Let us now redo Example 4.5 by first calculating all joint extreme distributions using Theorem 4.2. The extreme points on $\mathbf{p}_1$ are $(0.5, 0.5)^T$ and $(0.8, 0.2)^T$, and the extreme points on $\mathbf{p}_2$ are $(0.3, 0.7)^T$ and $(0.7, 0.3)^T$. Table 4.18 gives the 8 extreme points of $\Psi_E^{|S_1}$ calculated by using Theorem 4.2. As in all previous calculations, the lower and upper probabilities for $T$ are 0.3 and 0.7, respectively, while the conditional probability that the first resin is Type A given the type of the second resin ranges between 0.3 and 0.903. Notice that the upper and lower probability of $T$ might be achieved at different extreme points of $\Psi_E^{|S_1}$.

**Table 4.18** Example 4.5 Extreme joint distributions

| $i$ | $P_{EXT_i}^{1,1}$ | $P_{EXT_i}^{1,2}$ | $P_{EXT_i}^{2,1}$ | $P_{EXT_i}^{2,2}$ | $P(T)$ | $P_{1\|2}\left(S_1 = A \mid S_2 = A\right)$ |
|---|---|---|---|---|---|---|
| 1 | 0.15 | 0.35 | 0.15 | 0.35 | 0.5 | 0.500 |
| 2 | 0.15 | 0.35 | 0.35 | 0.15 | 0.3 | 0.300 |
| 3 | 0.35 | 0.15 | 0.15 | 0.35 | 0.7 | 0.700 |
| 4 | 0.35 | 0.15 | 0.35 | 0.15 | 0.5 | 0.500 |
| 5 | 0.24 | 0.56 | 0.06 | 0.14 | 0.38 | 0.800 |
| 6 | 0.24 | 0.56 | 0.14 | 0.06 | 0.3 | 0.632 |
| 7 | 0.56 | 0.24 | 0.06 | 0.14 | 0.7 | 0.903 |
| 8 | 0.56 | 0.24 | 0.14 | 0.06 | 0.62 | 0.800 |

Finally, let us rework Example 4.6 by rewriting the constraints in a linear form:

$$\begin{aligned}
&\textit{Subject to}\\
&0.5 \le P^{1,1} + P^{1,2} \le 0.8\\
&0.3 \le P^{1,1} + P^{2,1} \le 0.7\\
&0.3\left(P^{1,1} + P^{1,2}\right) \le P^{1,1} \le 0.7\left(P^{1,1} + P^{1,2}\right)\\
&0.3\left(P^{2,1} + P^{2,2}\right) \le P^{2,1} \le 0.7\left(P^{2,1} + P^{2,2}\right)\\
&0.5\left(P^{1,1} + P^{2,1}\right) \le P^{1,1} \le 0.8\left(P^{1,1} + P^{2,1}\right)\\
&0.5\left(P^{1,2} + P^{2,2}\right) \le P^{2,1} \le 0.8\left(P^{1,2} + P^{2,2}\right)\\
&\mathbf{1}^{\mathrm{T}} \cdot \mathbf{P} \cdot \mathbf{1} = 1;\\
&\mathbf{P} \ge 0
\end{aligned} \tag{4.73}$$

The results obtained by using linear optimization are given in Table 4.19 and Table 4.20. Since epistemic independence is assumed, extreme points of joint probability distributions are calculated by using the algorithm on page 114 and are listed in Table 4.21. Although the upper and lower probabilities are the same as in Example 4.6, they are achieved at different extreme points.

**Table 4.19** Example 4.6 Lower and upper probabilities for $T$ by the simplex method

| **Solution for** | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min<br>$P$ (T) = 0.322 | $\begin{pmatrix} 0.203 & 0.475 \\ 0.203 & 0.119 \end{pmatrix}$ | $(0.678, 0.322)^{\mathrm{T}}$ | $(0.407, 0.593)^{\mathrm{T}}$ |
| Max<br>$P$ (T) = 0.678 | $\begin{pmatrix} 0.475 & 0.203 \\ 0.119 & 0.203 \end{pmatrix}$ | $(0.678, 0.322)^{\mathrm{T}}$ | $(0.593, 0.407)^{\mathrm{T}}$ |

**Table 4.20** Example 4.6 Lower and upper probabilities for conditional probability $P_{1|2}\left(S_1 = A \mid S_2 = A\right)$ by the simplex method

| Solution for | Joint **P** | Marginal on $S_1$ | Marginal on $S_2$ |
|---|---|---|---|
| Min<br>$P_{1\|2}\left(s_1 = A \mid s_2 = A\right) = 0.5$ | $\begin{pmatrix} 0.15 & 0.35 \\ 0.15 & 0.35 \end{pmatrix}$ | $(0.5, 0.5)^{\mathrm{T}}$ | $(0.3, 0.7)^{\mathrm{T}}$ |
| Max<br>$P_{1\|2}\left(s_1 = A \mid s_2 = A\right) = 0.8$ | $\begin{pmatrix} 0.24 & 0.56 \\ 0.06 & 0.14 \end{pmatrix}$ | $(0.8, 0.2)^{\mathrm{T}}$ | $(0.62, 0.38)^{\mathrm{T}}$ |

**Table 4.21** Example 4.6 Lower and upper probabilities for $T$ and lower and upper conditional probability $P_{1|2}\left(S_1 = A \mid S_2 = A\right)$ by checking all extreme joint distributions

| $i$ | $P^{1,1}_{EXT_i}$ | $P^{1,2}_{EXT_i}$ | $P^{2,1}_{EXT_i}$ | $P^{2,2}_{EXT_i}$ | $P(T)$ | $P_{1\|2}\left(S_1 = A \mid S_2 = A\right)$ |
|---|---|---|---|---|---|---|
| 1 | 0.15 | 0.35 | 0.15 | 0.35 | 0.5 | 0.500 |
| 2 | 0.24 | 0.56 | 0.06 | 0.14 | 0.38 | 0.800 |
| 3 | 0.35 | 0.15 | 0.35 | 0.15 | 0.5 | 0.500 |
| 4 | 0.56 | 0.24 | 0.14 | 0.06 | 0.62 | 0.800 |
| 5 | 0.475 | 0.203 | 0.119 | 0.203 | 0.678 | 0.800 |
| 6 | 0.414 | 0.241 | 0.103 | 0.241 | 0.655 | 0.800 |
| 7 | 0.203 | 0.475 | 0.203 | 0.119 | 0.322 | 0.500 |
| 8 | 0.241 | 0.414 | 0.241 | 0.103 | 0.344 | 0.500 |

### *4.2.3 Strong Independence*

In *strong independence* (or *type*-1 *extension*), the set of probability measures, $\Psi_S$, is composed of all product measures, i.e.

$$\Psi_S = \left\{ P = P_1 \otimes P_2 : P_1 \in \Psi_1, P_2 \in \Psi_2 \right\} \quad ; \tag{4.74}$$

By comparing Eq. (4.42) with Eq. (4.74), one realizes that strong independence is obtained by imposing that *any* $P_1^{|s_2} \in \Psi_1$ and *any* $P_2^{|s_1} \in \Psi_2$ and these measures are the same for *all* values $s_2$ and $s_1$, respectively. i.e. by adding the following conditions to epistemic independence:

$$\forall s_2 \in S_2 : P_1^{|s_2} = P_1 \quad \text{and} \quad \forall s_1 \in S_1 : P_2^{|s_1} = P_2 \tag{4.75}$$

Since we added a constraint to epistemic independence, the previous independence properties are kept. In particular, learning the outcome of one experiment does not change our uncertainty about the outcome of the other experiment, in accordance with the intuitive notion of independence. Strong independence is an appropriate model when the following assumptions are satisfied:

- The outcomes for $s_2$ and $s_1$ result from random experiments, each governed by a unique (but unknown) probability measure, or distribution.
- We know that the probability measures, or distributions, belong to sets $\Psi_1$ and $\Psi_2$, respectively.

- The random experiments are stochastically independent (Eqs. (4.13) (4.14)).
- We do not know of any relationship between the two marginal probability measures (or distributions) that would enable us to rule out some of the possible combinations of marginal measures (or distributions).

Theorem 4.4 addresses the calculation of the extreme joint distributions and measures of the set of joint distributions, $\Psi_S$. Sets of joint distributions under assumptions of unknown interaction, epistemic irrelevance and epistemic independence are convex. However, when strong independence is assumed, $\Psi_S$ is not a convex set, as explained in the following Theorem 4.5. Although $\Psi_S$ is not convex, maxima and minima of linear functions are always achieved at extreme points of $\Psi_S$.

**Theorem 4.4.** *Under the assumption of strong independence, the set of extreme joint distributions (measures), EXT, is the set of product distributions (measures), each taken from the extreme distributions (measures) of the marginals, ETX$_i$:*

$EXT = \{ P = P_1 \otimes P_2 : P_1 \in EXT_1, P_2 \in EXT_2 \}$, i.e. per Eq. (**2.35**):

$$EXT = \left\{ \mathbf{P}^{\xi,\eta}_{EXT} = \mathbf{p}^{\xi}_{EXT_1} \left( \mathbf{p}^{\eta}_{EXT_2} \right)^{\mathrm{T}} : \mathbf{p}^{\xi}_{EXT_1} \in EXT_1, \mathbf{p}^{\eta}_{EXT_2} \in EXT_2 \right\} \tag{4.76}$$

*Proof*: Let the extreme points of the convex set of probability distributions on $S_1$ and $S_2$ be $\mathbf{p}^{\xi}_{EXT_1}$, $\xi = 1,\ldots\xi_1$, and $\mathbf{p}^{\xi}_{EXT_2}$, $\xi = 1,\ldots\xi_2$, respectively. Any $\mathbf{p}_1$ and $\mathbf{p}_2$ can be written as a linear combination of extreme points:

$$\begin{aligned}
&\mathbf{p}_1 = \left( \mathbf{p}^{1}_{EXT_1} \ldots \mathbf{p}^{\xi}_{EXT_1} \ldots \mathbf{p}^{\xi_1}_{EXT_1} \right) \left( \lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1} \right)^{\mathrm{T}}; \\
&\mathbf{p}_2 = \left( \mathbf{p}^{1}_{EXT_2} \ldots \mathbf{p}^{\xi}_{EXT_2} \ldots \mathbf{p}^{\xi_2}_{EXT_2} \right) \left( \lambda_2^1 \ldots \lambda_2^{\xi} \ldots \lambda_2^{\xi_2} \right)^{\mathrm{T}} \\
&0 \le \lambda_1^{\xi} \le 1,\ \xi = 1,\ldots,\xi_1;\ \sum_{\xi=1}^{\xi_1} \lambda_1^{\xi} = 1 \\
&0 \le \lambda_2^{\xi} \le 1,\ \xi = 1,\ldots,\xi_2;\ \sum_{\xi=1}^{\xi_2} \lambda_2^{\xi} = 1
\end{aligned} \tag{4.77}$$

Since $\mathbf{P} = \mathbf{p}_1 \otimes \mathbf{p}_2$, any joint probability distribution may be written as follows:

$$\mathbf{P} = \left(\mathbf{p}_{EXT_1}^{1} \cdots \mathbf{p}_{EXT_1}^{\xi} \cdots \mathbf{p}_{EXT_1}^{\xi_1}\right) \begin{pmatrix} \lambda_1^1 \\ \vdots \\ \lambda_1^{\xi} \\ \vdots \\ \lambda_1^{\xi_1} \end{pmatrix} \begin{pmatrix} \lambda_2^1 & \ldots & \lambda_2^{\xi} & \ldots & \lambda_2^{\xi_2} \end{pmatrix} \begin{pmatrix} \left(\mathbf{p}_{EXT_2}^{1}\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}_{EXT_2}^{\xi}\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{p}_{EXT_2}^{\xi_2}\right)^{\mathrm{T}} \end{pmatrix} \tag{4.78}$$

Extreme points of $\mathbf{P}$ are achieved if and only if:

$$\lambda_1^{\xi} = \begin{cases} 1, & \xi = m_1 \\ 0, & \xi \neq m_1 \end{cases} \text{ and } \lambda_2^{\xi} = \begin{cases} 1, & \xi = m_2 \\ 0, & \xi \neq m_2 \end{cases}; \; m_1 = 1,\ldots,\xi_1; \; m_2 = 1,\ldots,\xi_2 \tag{4.79}$$

Therefore, $EXT = \left\{ \mathrm{P}_{EXT}^{\xi,\eta} = \mathbf{p}_{EXT_1}^{\xi} \left(\mathbf{p}_{EXT_2}^{\eta}\right)^{\mathrm{T}} : \mathbf{p}_{EXT_1}^{\xi} \in EXT_1, \mathbf{p}_{EXT_2}^{\eta} \in EXT_2 \right\}$. ■

**Theorem 4.5.** *Under strong independence, the set of joint distributions (measures),* $\Psi_{\mathrm{S}}$*, is not convex.*

*Proof* : Consider a counterexample with $\xi_1 = 3$, $\xi_2 = 2$ in Eq. (4.77) and let us proceed by contradiction by assuming that $\Psi_{\mathrm{S}}$ is a convex set.

Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two joint distributions in $\Psi_{\mathrm{S}}$ such that $\mathbf{P}_1$ is generated by taking $\left(\lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1}\right) = (1,0,0)$, $\left(\lambda_2^1 \ldots \lambda_2^{\xi} \ldots \lambda_2^{\xi_2}\right) = (0,1)$, and thus $\left(\lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1}\right)^{\mathrm{T}} \left(\lambda_2^1 \ldots \lambda_2^{\xi} \ldots \lambda_2^{\xi_2}\right) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, and $\mathbf{P}_2$ is generated by taking $\left(\lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1}\right) = (0,0,1)$, $\left(\lambda_2^1 \ldots \lambda_2^{\xi} \ldots \lambda_2^{\xi_2}\right) = (1,0)$, and thus $\left(\lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1}\right)^{\mathrm{T}} \left(\lambda_2^1 \ldots \lambda_2^{\xi} \ldots \lambda_2^{\xi_2}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$.

The mid point, $\mathbf{P}_m$, between $\mathbf{P}_1$ and $\mathbf{P}_2$ is $1/2(\mathbf{P}_1+\mathbf{P}_2)$. Consequently, $\left(\lambda_1^1 \ldots \lambda_1^{\xi} \ldots \lambda_1^{\xi_1}\right)^{\mathrm{T}}\left(\lambda_2^1 \ldots \lambda_2^{\xi} \ldots \lambda_2^{\xi_2}\right)$ for $\mathbf{P}_m$ is equal to $\begin{pmatrix} 0 & 1/2 \\ 0 & 0 \\ 1/2 & 0 \end{pmatrix}$, which could be written in the form $\left(\lambda_1^1, \lambda_1^2, \lambda_1^3\right)^{\mathrm{T}}\left(\lambda_2^1, \lambda_2^2\right)$ based on the assumption that $\mathbf{P}_m$ is in the convex set $\Psi_{\mathrm{S}}$.

Thus, $\left(\lambda_1^1, \lambda_1^2, \lambda_1^3\right)^{\mathrm{T}}\left(\lambda_2^1, \lambda_2^2\right)= \begin{pmatrix} \lambda_1^1\lambda_2^1, \lambda_1^1\lambda_2^2 \\ \lambda_1^2\lambda_2^1, \lambda_1^2\lambda_2^2 \\ \lambda_1^3\lambda_2^1, \lambda_1^3\lambda_2^2 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 0 & 0 \\ 1/2 & 0 \end{pmatrix}$, subject to $\lambda_i^j \geq 0$, $\sum_j \lambda_i^j = 1$. Since $\lambda_1^1\lambda_2^2$ =1/2 and $\lambda_1^3\lambda_2^1$=1/2, $\lambda_1^1\lambda_2^2\lambda_1^3\lambda_2^1$=1/4.

However, we also have $\lambda_1^1\lambda_2^1$=0 and $\lambda_1^3\lambda_2^2$ =0, so $\lambda_1^1\lambda_2^1\lambda_1^3\lambda_2^2 = \lambda_1^1\lambda_2^2\lambda_1^3\lambda_2^1$ =0, which contradicts the previous result $\lambda_1^1\lambda_2^2\lambda_1^3\lambda_2^1$ =1/4. Therefore, there are no $\lambda_i^j$ that satisfy all requirements, i.e., $\mathbf{P}_m \neq \mathbf{P}_1 \otimes \mathbf{P}_2$. Therefore, $\Psi_{\mathrm{S}}$ is not convex. ■

There are several options for carrying out calculations on the joint space:

1. If extreme distributions are available for the marginals (Eq. (4.18)), first calculate all $\xi_1 \times \xi_2$ extreme joint distributions $\mathbf{p}_{EXT}$ as indicated in Eq. (4.76). If the objective function is linear (e.g., prevision or probability), then the solution is simply found by evaluating the objective function on all vertices $\mathbf{p}_{EXT}$. i. e:

   *Subject to*:

   $$\mathbf{P} \in EXT = \left\{\mathbf{p}_{EXT}^{\xi,\eta}\right\} \tag{4.80}$$

   This problem is much easier to solve than the non-convex and NP-hard optimization problems that follow.
2. If extreme distributions are available for the marginals (Eq. (4.18)), use directly constraints in Eqs. (4.14) and (4.74), thus obtaining quadratic constraints in the $n_1 \times n_2$ components $p^{i,j}$, $n_1 + n_2$ components $p_i^k$ and in the $\xi_1+\xi_2$ components $c_i^{\xi}$:

$$\begin{aligned}
&\textit{Subject to}:\\
&\mathbf{P}-\mathbf{p}_1\left(\mathbf{p}_2\right)^{\mathrm{T}}=0\\
&\mathbf{p}_i-\sum_{\xi=1}^{\xi_i} c_i^{\xi}\mathbf{p}_{EXT_i}^{\xi}=0;\ i=1,2\\
&\sum_{\xi=1}^{\xi_i} c_i^{\xi}-1=0,\ c_i^{\xi}\geq 0;\ i=1,2
\end{aligned}\tag{4.81}$$

3. If expectation (prevision) bounds are given on the marginals (Eq. (4.17)), use directly constraints in Eqs. (4.74) and (4.76), thus obtaining quadratic constraints in the $n_1 \times n_2$ components $p^{i,j}$ and $n_1 + n_2$ components $P_i(s_i)$:

$$\begin{aligned}
&\textit{Subject to}:\\
&\mathbf{P}-\mathbf{p}_1\left(\mathbf{p}_2\right)^{\mathrm{T}}=0\\
&E_{LOW}\left[f_i^k\right]\leq\left(\mathbf{f}_i^k\right)^{\mathrm{T}}\ \mathbf{p}_i\leq E_{UPP}\left[f_i^k\right];\ k=1,...,k_i;\ i=1,2
\end{aligned}\tag{4.82}$$

**Example 4.10.** Consider again the situation and knowledge available in Example 4.1, but now suppose that a cartridge is picked from each box in a stochastically independent way. We want to write down the entire set of joint distributions in $\Psi_S$ and then calculate the upper and lower probabilities for the case in which the resin is not activated because the same resin type is selected, i.e. event $T$={(A, A), (B, B)}.

The possible distributions for Type A and Type B resins in each box are given in Table 4.1. Since the two picks are stochastically independent, the probability of picking, say, two Type B cartridges is the product of the two relative frequencies listed in the third columns of Table 4.1a and Table 4.1b, which can take any of 4·5=20 possible values ranging from 0.2·0.3 = 0.06 to 0.5·0.7 = 0.35. The interval [0.06, 0.35] represents the convex hull of the possible probabilities of picking two Type B cartridges, and it represents our uncertainty about this event.

The physical meaning of this example makes $\Psi_S$ finite. $\Psi_S$ has 4·5=20 probability measures (given in Table 4.22), of which 4 are extreme points $p_{EXT}$ (at least one $P(s_2, s_1)$ attains a maximum or a minimum at one of the extreme points). The 4 extreme points are the products of the gray-hatched probabilities in Table 4.1, and are hatched gray in Table 4.22.

**Table 4.22** Example 4.10: joint probabilities of elementary events (gray hatch denotes extreme points $\mathbf{p}_{EXT}$)

| ($P$(A, A), $P$(A, B), $P$(B, A), $P$(B, B)) | ($P$(A, A), $P$(A, B), $P$(B, A), $P$(B, B)) |
|---|---|
| (0.15, 0.35, 0.15, 0.35) | (0.21, 0.49, 0.09, 0.21) |
| (0.2, 0.3, 0.2, 0.3) | (0.28, 0.42, 0.12, 0.18) |
| (0.25, 0.25, 0.25, 0.25) | (0.35, 0.35, 0.15, 0.15) |
| (0.3, 0.2, 0.3, 0.2) | (0.42, 0.28, 0.18, 0.12) |
| (0.35, 0.15, 0.35, 0.15) | (0.49, 0.21, 0.21, 0.09) |
| (0.18, 0.42, 0.12, 0.28) | (0.24, 0.56, 0.06, 0.14) |
| (0.24, 0.36, 0.16, 0.24) | (0.32, 0.48, 0.08, 0.12) |
| (0.3, 0.3, 0.2, 0.2) | (0.4, 0.4, 0.1, 0.1) |
| (0.36, 0.24, 0.24, 0.16) | (0.48, 0.32, 0.12, 0.08) |
| (0.42, 0.18, 0.28, 0.12) | (0.56, 0.24, 0.14, 0.06) |

Consider again the case in which the resin is not activated because the same resin type is selected: $T$={(A, A), (B, B)}. The solution is to be searched through the extreme distributions $\mathbf{p}_{EXT}$. The lower probability of this event is given by the seventh row, second column in Table 4.22, i.e. 0.24+0.14 = 0.38; the upper probability is given by the last row in Table 4.22, i.e. 0.56+0.06 = 0.62. The upper and lower conditional probabilities that the first resin is Type A given the type of the second resin, defined by $p^{1,1}/\left(p^{1,1}+p^{2,1}\right)$, are equal to 0.8 and 0.5, and are achieved at the two extreme distributions in the second column, and in the first column of Table 4.22, respectively.

**Example 4.11.** Consider again the situation and knowledge available in Example 4.4, but now suppose that experiments are conducted in a stochastically independent way. We want to write down the entire set of joint distribution in $\Psi_S$ and then calculate the upper and lower probabilities for the event $T=\left\{\left(s_1^1,s_2^1\right),\left(s_1^2,s_2^2\right),\left(s_1^3,s_2^3\right),\left(s_1^4,s_2^4\right)\right\}$.

The extreme distributions in $S_1$ and $S_2$ are listed in Table 4.7. The product of marginal extreme distributions generates an extreme joint distribution in $\Psi_S$. When the probability distribution in $\Psi_1$ is $p_{EXT1}^1$ = (0.2, 0.7, 0, 0.1), and the distribution in $\Psi_2$ is $p_{EXT2}^6$ = (0, 0, 1, 0), the joint probability distribution is

$$\mathbf{P}=\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.2 & 0.7 & 0 & 0.1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and the probability of event $T$ achieves its minimum value, i.e., $P(T) = 0$. When $p^5_{EXT1} = (0, 0, 1, 0)$ in $\Psi_1$ and $p^6_{EXT2} = (0, 0, 1, 0)$ in $\Psi_2$, the joint probability distribution is

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and the probability of event $T$ achieves its maximum value, i.e., $P(T) = 1$.

### *4.2.4 Relationships between the Four Types of Independence*

The constraints that define the three sets of probability measures are summarized as follows:

- Unknown interaction: $\Psi_{\text{U}}$: (4.25)
- Epistemic irrelevance: $\Psi_{\text{E}}^{|s_i}$ : (4.25) + (one of (4.42))
- Epistemic independence: $\Psi_{\text{E}}$: (4.25) + (4.42)
- Strong independence: $\Psi_{\text{S}}$: (4.25) + (4.42) + (4.75)

Since constraints are consecutively added, the sets of probability measures are nested, i.e. $\Psi_{\text{S}} \subseteq \Psi_{\text{E}} \subseteq \Psi_{\text{E}}^{|s_i} \subseteq \Psi_{\text{U}}$. As a consequence, the upper and lower probability bounds are also nested:

$$P_{\text{U},LOW} \leq P_{\text{E},LOW}^{|s_i} \leq P_{\text{E},LOW} \leq P_{\text{S},LOW} \leq P_{\text{S},UPP} \leq P_{\text{E},UPP} \leq P_{\text{E},UPP}^{|s_i} \leq P_{\text{U},UPP} \qquad (4.83)$$

This is exemplified by the probability of set $T$={(A, A), (B, B)} in Example 4.1, Example 4.5, Example 4.6, and Example 4.10:

- Unknown interaction: $P_{\text{U},LOW} = 0.00;\ P_{\text{U},UPP} = 1.00$
- Epistemic irrelevance: $P_{\text{E},LOW}^{|s_1} = 0.30;\ P_{\text{E},UPP}^{|s_1} = 0.70$
- Epistemic independence: $P_{\text{E},LOW} = 19/59 = 0.322;\ P_{\text{E},UPP} = 40/59 = 0.678$
- Strong independence: $P_{\text{S},LOW} = 0.38;\ P_{\text{S},UPP} = 0.62$

## 4.3 Independence When Marginals Are Random Sets

When two marginal random sets, $\mathcal{F}_1=\{(A_1, m_1)\}$ and $\mathcal{F}_2=\{(A_2, m_2)\}$, are assigned on $S_1$ and $S_2$, respectively, two lines of thought can be pursued in order to combine the information on the Cartesian product $S=S_1 \times S_2$:

1) Consider each marginal random set as a convex set of probability distributions/measures, $\Psi_i$ defined either by
   a. The Belief (lower probability bound) of each subset of $S_i$ (Section 3.2.2); or, alternatively, the Plausibility (upper probability bound) of each subset of $S_i$. In this case, $\Psi_i$ is defined as in Eq. (4.17).
   b. Its extreme distributions (Section 3.2.3.4); in this case, $\Psi_i$ is defined as in Eq. (4.18).

   According to this line of thought, one would proceed as described in the previous Section 4.2 within the theory of imprecise probabilities and without any attempt to obtain a random relation on $S$. This approach will not be expanded on further because it has already been dealt with.
2) Consider a random relation defined (similarly to Eq. (4.5)) as the set of probability distributions/measures:

$$\Psi = \left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left( m\left( A_1^i \times A_2^j \right) \cdot P^{i,j} \right) \mid P^{i,j} \in \Psi^{i,j} \right\} \tag{4.84}$$

   where $\Psi^{i,j}$ is the set of *all* probability measures defined over the focal element $A_1^i \times A_2^j$ (and equal to zero elsewhere). The only unknown in Eq. (4.84) is thus the basic probability assignment $m\left( A_1^i \times A_2^j \right)$. Since this is a precise probability, the problem of assigning $m$ is similar to the reconstruction of a joint probability from precise marginal distributions. Rows 2 and 3 in Table 4.23 indicate two possible choices, namely stochastic independence and unknown interaction. The question now arises as to whether and how the definitions of independence given in Section 4.2 can be recast in the form (4.84). Sections 4.3.1 and 4.3.2 address this question.
3) If one sets constraints on distribution sets $\Psi^{i,j}$ in Eq. (4.84), then the resulting set $\Psi$ is no longer a random relation. In Sections 4.3.3 and 4.3.4, additional constraints will be introduced in order to recover epistemic independence and strong independence, respectively[N 4-1].

In Eq. (4.84), the focal elements for the random relation are taken as the Cartesian products $A_1^i \times A_2^j$ because any other subset $A \subset A_1^i \times A_2^j$ would indicate some type of dependence between the two variables, $s_1$ and $s_2$, and is therefore ruled out in this section on independence. Subsets $A \subset A_1^i \times A_2^j$ will be introduced in Section 4.4, which deals with correlation.

**Table 4.23** Types of independence when the marginals are random sets

| **Name of independence for the joint set of probability measures** | **Indep. for *m*** | **Indep. for $P^{i,j}$** | **Unique *m*?** | **Generate Random Relation?** |
|---|---|---|---|---|
| Random set independence | Stochastic Indep. | Unknown Interaction | Yes | Yes, only one |
| Unknown Interaction | Unknown Interaction | Unknown Interaction | No | Yes, infinite ones |
| Epistemic independence | Stochastic Indep. | Epistemic Indep. | Yes | No |
| Strong Independence | Stochastic Indep. | Stochastic Indep. | Yes | No |

Finally, Section 4.3.5 covers the special case of a fuzzy Cartesian product, in which only some Cartesian products $A_1^i \times A_2^j$ are used as opposed to *all* Cartesian products in Eq. (4.84). The basic definitions introduced in this section are mainly taken from references ((Fetz 2001; Fetz 2003; Fetz 2004; Fetz and Oberguggenberger 2004)).

### *4.3.1 Random Set Independence*

Given the choices in Table 4.23 (second row), there is random set independence when the set of joint probability measures is:

$$\Psi_{\mathrm{R,R}} = \left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left( m_1\left(A_1^i\right) m_2\left(A_2^j\right) \cdot P^{i,j} \right) \mid P^{i,j} \in \Psi^{i,j} \right\} \tag{4.85}$$

where $\Psi^{i,j}$ is the set of all probability measures defined over the focal element $A_1^i \times A_2^j$ (and zero elsewhere). $\Psi_{\mathrm{R,R}}$ corresponds to the set of distributions/measures obtained by assigning mass $m_1\left(A_1^i\right) m_2\left(A_2^j\right)$ arbitrarily to elements of $A_1^i \times A_2^j$, and is thus the unique random relation whose focal elements are all Cartesian products $A_1^i \times A_2^j$ and whose basic probability assignment is $m(A^i \times A^j) := m_1(A_1^i)m_2(A_2^j)$. Upper and lower probabilities on a set, $T$, are then calculated using Eqs. (3.3) and (4.1). Alternatively, since $\Psi^{i,\,j}$ contains all probability measures, it also contains the Dirac measure (Section 2.1), and thus upper and lower probabilities are obtained by appropriately collocating the Dirac measures in their focal elements; this is illustrated in Example 4.12 below.

Random set independence is an appropriate model for outcomes $s_1$ and $s_2$ under the following assumptions:

1) There are two random experiments with possibility spaces $X_1$ and $X_2$, where known probability of elementary events $m_1$* and $m_2$* are assigned, respectively.
2) Each space, $X_i$, is related to $S_i$ through a multi-valued mapping $G_i$, meaning that if $x_i$ is the outcome of random experiment $i$, then we only learn that the true state of $S_i$ belongs to the subset $G_i(x_i) \subseteq S_i$. (Section 2.1.1).
3) The probability distribution $m_i$* on $X_i$ induces the probability assignment $m_i(G_i(x_i)) = m_i^*(x_i)$ on $S_i$.
4) The probabilities of elementary events $m_1$* and $m_2$* are stochastically independent.
5) Nothing is known about the interaction between the two mechanisms for selecting outcomes $s_1$ and $s_2$ from the sets $G_1(x_1)$ and $G_2(x_2)$.

**Example 4.12.** Let us consider the two boxes of resin as in Example 4.1 (page 116). In the current example, if a cartridge of unknown type is picked from one box, it is replaced with another cartridge by a completely unknown procedure. If two cartridges of unknown types are picked, then there can be arbitrary correlation between the types of the replacement cartridges. In addition, the operations of picking cartridges from the two boxes are stochastically independent. The marginal probability assignments are:

a) First box: $m_1(\{A\})=0.5$, $m_1(\{B\})=0.2$, $m_1(\{A, B\})=0.3$.
b) Second box: $m_2(\{A\})=0.3$, $m_2(\{B\})=0.3$, $m_2(\{A, B\})=0.4$.

Figure 4.11 illustrates the stochastic independent random relation together with the two marginal random sets, whereas Table 4.24 gives the composition of the focal elements. The probability of the event $T$ that the two cartridges are of the same type, $T= \{(A, A), (B, B)\}$, is comprised between the Belief and Plausibility of $T$ (Eq. 2.3).

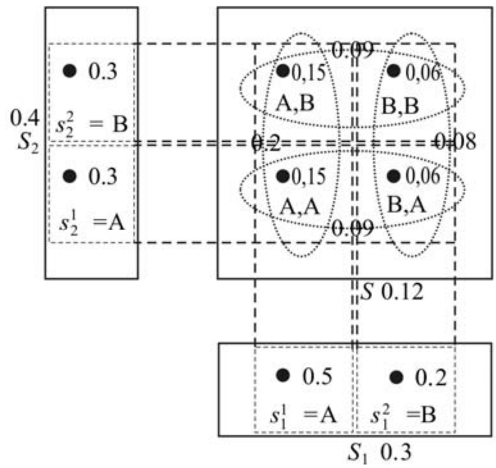



**Fig. 4.11** Example 4.12: marginal random sets and stochastic independent random relation. Numbers indicate probability assignment; dashed lines envelop focal elements for the stochastic independent random relation

**Table 4.24** Example 4.12: Stochastic independent random relation

| Focal element no. | Focal element | Probability assignment |
|---|---|---|
| 1 | {( A, A)} | $m^{1,1} = 0.15$ |
| 3 | {( A, B)} | $m^{1,2} = 0.15$ |
| 5 | {( A, A), (A, B)} | $m^{1,3} = 0.2$ |
| 2 | {( B, A)} | $m^{2,1} = 0.06$ |
| 4 | {( B, B)} | $m^{2,2} = 0.06$ |
| 6 | {( B, A), (B, B)} | $m^{2,3} = 0.08$ |
| 7 | {( A, A), (B, A)} | $m^{3,1} = 0.09$ |
| 8 | {( A, B), (B, B)} | $m^{3,2} = 0.09$ |
| 9 | {( A, A), (B, B), (A, B), (B, A)} | $m^{3,3} = 0.12$ |

$$
\begin{aligned}
& Pla(T) = \sum_i m^i \mid A^i \cap T \neq \varnothing = m^{1,1} + m^{2,2} + m^{1,3} + m^{2,3} + m^{3,1} + m^{3,2} + m^{3,3} = \\
& 0.15 + 0.06 + 0.2 + 0.08 + 0.09 + 0.09 + 0.12 = 0.79 = 1 - Bel(T^{\mathrm{C}}) = \\
& = 1 - (m^2 + m^3) = 1 - (0.15 + 0.06) = 0.79 \\
& Bel(T) = \sum_i m^i \mid A^i \subseteq T = m^{1,1} + m^{2,2} = 0.15 + 0.06 = 0.21 \\
& 0.21 \ \le \ P(T) \ \le \ 0.79
\end{aligned}
\tag{4.86}
$$

Alternatively, the upper probability is achieved by collocating the Dirac measures on (A, A) for focal elements that contain (A, A) but not (B, B), namely, 1, 5 and 7; on (B, B) for focal elements that contain (B, B) but not (A, A), namely 4, 6 and 8; and on either (A, A) or (B, B) for focal element 9. This yields the joint distributions:

$$
{}_1\mathbf{p} = \begin{pmatrix} 0.44 & 0.15 \\ 0.06 & 0.35 \end{pmatrix} \qquad {}_2\mathbf{p} = \begin{pmatrix} 0.56 & 0.15 \\ 0.06 & 0.23 \end{pmatrix}
$$

Likewise, the lower probability is achieved by collocating the Dirac measure on ($A$, $B$) for focal elements 5 and 8, ($B$, $A$) for focal elements 6 and 7, and either ($A$, $B$) or ($B$, $A$) for focal element 9. This yields the probabilities of elementary events ($AA$, $AB$, $BA$, $BB$): (0.15, 0.44, 0.35, 0.06) or (0.15, 0.56, 0.23, 0.06).

$$ {}_3\mathbf{p} = \begin{pmatrix} 0.15 & 0.44 \\ 0.35 & 0.06 \end{pmatrix} \qquad {}_4\mathbf{p} = \begin{pmatrix} 0.15 & 0.56 \\ 0.23 & 0.06 \end{pmatrix} $$

Notice that the marginals (Eq. (2.24)) of ${}_1\mathbf{p}$ through ${}_4\mathbf{p}$ are in the sets of distributions defined by the marginal random sets a) and b) above. The four joint probabilities distributions ${}_1\mathbf{p}$ through ${}_4\mathbf{p}$ are also extreme points of $\Psi$; a complete calculation by means of the permutation algorithm, Sections 3.2.3.4 and 3.2.3.5, shows that $\Psi$ has 16 extreme points (8 of which have multiplicity equal to 2).

Distributions ${}_1\mathbf{p}$ through ${}_4\mathbf{p}$ yield the upper and lower probabilities that the second cartridge is Type A, given that the first cartridge is Type A, which can also be calculated using Eqs. (3.35) (Let $D$ = {(A, A), (B,A)} and $E$ = {(A,A), (A,B)}):

$$ Bel\left(D/E\right) = \frac{Bel(D\cap E)}{Bel(D\cap E)+Pla(D^{\mathrm{C}}\cap E)} = \frac{Bel(\{(\mathrm{A,A})\})}{Bel(\{(\mathrm{A,A})\})+Pla(\{(\mathrm{A,B})\})} $$

$$ = \frac{0.15}{0.15+(0.15+0.2+0.09+0.12)} = \frac{0.15}{0.71} = 0.211 $$

$$ = \min_{P\in\Psi} \frac{P\left(D\cap E\right)}{P\left(E\right)} = \frac{{}_4p(\mathrm{A,A})}{{}_4p(\mathrm{A,A})+{}_4p(\mathrm{A,B})} = \frac{0.15}{0.15+0.56} = 0.211 $$

$$ Pla\left(D/E\right) = \frac{Pla(D\cap E)}{Pla(D\cap E)+Bel(D^{\mathrm{C}}\cap E)} = \frac{Pla(\{(\mathrm{A,A})\})}{Pla(\{(\mathrm{A,A})\})+Bel(\{(\mathrm{A,B})\})} $$

$$ = \frac{0.56}{0.56+0.15} = \frac{0.56}{0.71} = 0.789 $$

$$ = \max_{P\in\Psi} \frac{P\left(D\cap E\right)}{P\left(E\right)} = \frac{{}_2p(\mathrm{A,A})}{{}_2p(\mathrm{A,A})+{}_2p(\mathrm{A,B})} = \frac{0.56}{0.56+0.15} = 0.789 $$

Notice that these probabilities may also be obtained as Belief and Plausibility, respectively, of the conditional event $T|E$, where $E$ = {(A,A), (A,B)} (the first cartridge is A) because the posterior event {(B, B)} is impossible. When compared with the upper and lower probabilites of $T$ ([0.21, 0.79]) the posterior probability interval is slightly reduced. However, when compared with the interval [0.3, 0.7] of marginal probabilities that the cartridge picked from the second box is of type A, the posterior probability interval is larger. Once again, we find that learning the type of cartridge changes our uncertainty about the type of the other cartridge, contrary to the intuitive notion of independence.

### 4.3.2 Unknown Interaction

Given the choices in Table 4.23 (third row), there is unknown interaction when the set of joint probability measures is:

$$\Psi_{\mathrm{R,U}} = \left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left( m\left(A_1^i \times A_2^j\right) \cdot P^{i,j} \right) \mid P^{i,j} \in \Psi^{i,j} \right\} \tag{4.87}$$

where $m\left(A_1^i \times A_2^j\right)$ satisfies the marginals rules (4.4), and $\Psi^{i,j}$ is the set of all probability measures defined over the focal element $A_1^i \times A_2^j$ (and zero elsewhere). Fetz (Fetz 2001) showed that this is also the set of probability measures obtained by applying the definition of unknown interaction (Section 4.2.1) to the marginal random sets, i.e. $\Psi_{\mathrm{U}} = \Psi_{\mathrm{R,U}}$.

Set $\Psi_{\mathrm{R,U}}$ defines an infinite number of random relations, namely all random relations, such that $m\left(A_1^i \times A_2^j\right)$ satisfies the marginal rules (4.4). Since $\Psi^{i,j}$ contains all probability measures, it also contains the Dirac measure (Section 2.1), and thus upper and lower probabilities are obtained by appropriately collocating the Dirac measures $\delta^{i,j}$ in their focal elements, as was done in Section 4.3.1, and by optimizing the probability assignments. More precisely, the upper probability on a set, $T$ is:

$$\begin{aligned} P_{upp}(T) &= \max\left\{P(T) : P \in \Psi_{R,U}\right\} = \\ &= \max\left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left( m\left(A_1^i \times A_2^j\right) \cdot \delta^{i,j} \right) \mid \delta^{i,j} \in \Psi^{i,j} \right\}, \end{aligned} \tag{4.88}$$

where the maximum is achieved by collocating $\delta^{i,j}$ in $T \cap (A_1^i \times A_2^j)$, i.e.

$$P_{upp}(T) = \max \sum_{i,j: T \cap \left(A_1^i \times A_2^j\right) \neq \varnothing} m\left(A_1^i \times A_2^j\right) \tag{4.89}$$

The function to maximize is just the Plausibility of $T$. Therefore, by letting $m^{i,j} := m\left(A_1^i \times A_2^j\right)$, the probability assignment that yields the upper probability on a set, $T$, is found by solving the linear optimization problem:

$$\begin{aligned}
&\text{maximize} \quad \sum_{i,j:T\cap\left(A_1^i\times A_2^j\right)\neq\varnothing} m^{i,j} \\
&\textit{subject to}: \\
&m_1\left(A_1^i\right)=\sum_{j=1}^{n_2} m^{i,j} \qquad i=1,\dots,n_1 \\
&m_2\left(A_2^j\right)=\sum_{i=1}^{n_1} m^{i,j} \qquad j=1,\dots,n_2 \\
&m^{i,j}\geq 0
\end{aligned} \tag{4.90}$$

This problem can be easily programmed in the standard form:

$$\begin{aligned}
&\text{minimize} -\mathbf{c}^{\mathrm{T}}\mathbf{x} \\
&\textit{subject to:} \\
&\mathbf{Mx}=\mathbf{b}\,, \quad \mathbf{x}\geq\mathbf{0}
\end{aligned} \tag{4.91}$$

where:

$$\begin{aligned}
&c_h=\begin{cases}1 & \text{if } \; T\cap\left(A_1^i\times A_2^j\right)\neq\varnothing \\ 0 & \text{otherwise}\end{cases} \\
&x_h=m^{i,j} \\
&h=j+(i-1)\cdot n_2 \quad \text{for } \; i=1,\dots,n_1;\, j=1,\dots,n_2 \\
&M_{k,h}=\begin{cases}1 & \text{if} \quad k=1,\dots,n_1;\; h=(i-1)\cdot n_2+1,\dots,i\cdot n_2 \\ 1 & \text{if} \quad k=n_1+1,\dots,n_1+n_2;\, h=i-n_1,\dots,(i-n_1)+n_1(n_2-1) \\ 0 & \text{otherwise}\end{cases} \\
&b_k=\begin{cases}m_1\left(A_1^i\right) & \text{if} \quad k=1,\dots,n_1 \\ m_2\left(A_2^j\right) & \text{if} \quad k=n_1+1,\dots,n_1+n_2\end{cases} \\
&k=1,\dots,n_1,n_1+1,\dots,n_1+n_2
\end{aligned} \tag{4.92}$$

The lower probability on a set $T$ is obtained by collocating $\delta^{i,j}$ in $\left(A_1^i\times A_2^j\right)\cap T^{\mathrm{C}}$ whenever ( $A_1^i\times A_2^j$ )$\not\subset T$ and ( $A_1^i\times A_2^j$ )$\cap T\neq\varnothing$, i.e.

$$P_{low}\left(T\right)=\min\sum_{i,j:\left(A_1^i\times A_2^j\right)\subseteq T} m\left(A_1^i\times A_2^j\right) \tag{4.93}$$

The function to minimize is just the Belief of $T$. Therefore, the probability assignment that yields the lower probability on a set, $T$, is found by solving the linear optimization problem:

$$\begin{aligned}
&\text{minimize} \sum_{i,j:\left(A_1^i \times A_2^j\right)\subseteq T} m^{i,j} \\
&\textit{subject to}: \\
&m_1\left(A_1^i\right)=\sum_{j=1}^{n_2} m^{i,j} \qquad i=1,...,n_1 \\
&m_2\left(A_2^j\right)=\sum_{i=1}^{n_1} m^{i,j} \qquad j=1,...,n_2 \\
&m^{i,j}\geq 0
\end{aligned} \tag{4.94}$$

This problem can be easily programmed in the standard form:

$$\begin{aligned}
&\text{minimize} -\mathbf{c}^{\mathrm{T}}\mathbf{x} \\
&\textit{subject to:} \\
&\mathbf{Mx}=\mathbf{b} \quad , \quad \mathbf{x}\geq \mathbf{0}
\end{aligned} \tag{4.95}$$

where:

$$\begin{aligned}
&c_h=\begin{cases} 1 \ \text{ if } \ \left(A_1^i\times A_2^j\right)\subseteq T \\ 0 \ \text{ otherwise} \end{cases} \\
&x_h=m^{i,j} \\
&h=j+(i-1)\cdot n_2 \quad \text{for } \ i=1,...,n_1;\, j=1,...,n_2
\end{aligned} \tag{4.96}$$

Unknown interaction is an appropriate model for outcomes $s_1$ and $s_2$ under the following assumptions:

1) There are two random experiments with possibility spaces $X_1$ and $X_2$, where known probability of elementary events $m_1$* and $m_2$* are assigned, respectively.
2) Each space, $X_i$, is related to $S_i$ through a multi-valued mapping $G_i$, meaning that if $x_i$ is the outcome of random experiment $i$, then we only learn that the true state of $S_i$ belongs to the subset $G_i(x_i) \subseteq S_i$. (Section 2.1).
3) The probability distribution $m_i$* on $X_i$ induces the probability assignment $m_i(G_i(x_i)) = m_i^*(x_i)$ on $S_i$.

4) Nothing is known about the interaction between the two mechanisms for selecting outcomes $x_1$ and $x_2$ from the sets $X_1$ and $X_2$, respectively, i.e., the probabilities of elementary events $m_1$* and $m_2$* are linked by unknown interaction.
5) Nothing is known about the interaction between the two mechanisms for selecting outcomes $s_1$ and $s_2$ from the sets $G_1(x_1)$ and $G_2(x_2)$, respectively.

**Example 4.13.** Let us consider the two boxes of resin as in Example 4.12 (page 160). In the current example, however, nothing is known about the relationship between picking a cartridge from the first box and picking one from the second box. The focal elements for the set of random relations along with the numbering of the probability assignment are still as in Table 4.24. For $T$= {($A$, $A$), ($B$, $B$)}, problem (4.90) is written as:

$$\text{maximize} \sum_{i,j:T\cap\left(A_1^i\times A_2^j\right)\neq\varnothing} m^{i,j} = m^{1,1}+m^{1,3}+m^{2,2}+m^{2,3}+m^{3,1}+m^{3,2}+m^{3,3}$$

$$\begin{aligned}
&\textit{subject to:}\\
&m^{1,1}+m^{1,2}+m^{1,3}=m_1\left(A_1^1\right)=0.5\\
&m^{2,1}+m^{2,2}+m^{2,3}=m_1\left(A_1^2\right)=0.2\\
&m^{3,1}+m^{3,2}+m^{3,3}=m_1\left(A_1^3\right)=0.3\\
&m^{1,1}+m^{2,1}+m^{3,1}=m_2\left(A_2^1\right)=0.3\\
&m^{1,2}+m^{2,2}+m^{3,2}=m_2\left(A_2^3\right)=0.3\\
&m^{1,3}+m^{2,3}+m^{3,3}=m_2\left(A_2^3\right)=0.4
\end{aligned} \tag{4.97}$$

Vector **c** and matrix **M** take the values:

$$\mathbf{c}=(1,\,0,\,1,\,0,\,1,\,1,\,1,\,1,\,1)^{\mathrm{T}}$$
$$\mathbf{M}=\begin{pmatrix}1&1&1&0&0&0&0&0&0\\0&0&0&1&1&1&0&0&0\\0&0&0&0&0&0&1&1&1\\1&0&0&1&0&0&1&0&0\\0&1&0&0&1&0&0&1&0\\0&0&1&0&0&1&0&0&1\end{pmatrix} \tag{4.98}$$

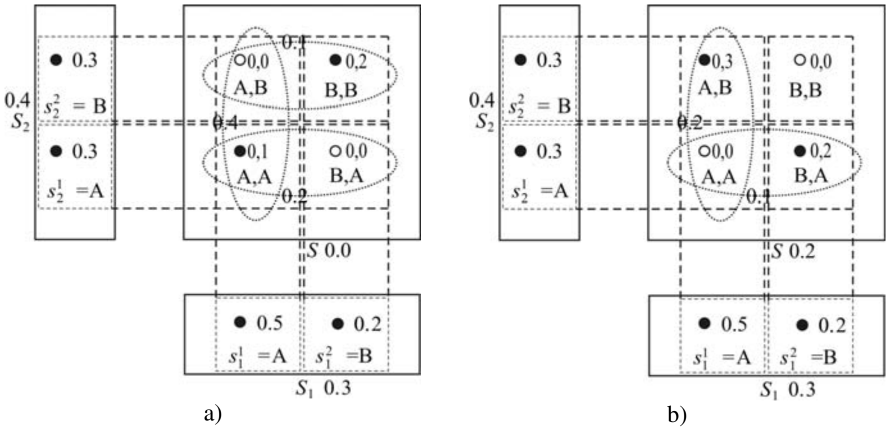



**Fig. 4.12** Example 4.13: marginal random sets and random relations that: a) maximize the Plausibility; b) minimize the Belief. Numbers indicate values of probability assignment; dashed lines envelop focal elements for random relation

By using the simplex method, the maximizing solution is found to be equal to (0.1, 0.0, 0.4, 0.0, 0.2, 0.0, 0.2, 0.1, 0.0); the solution that minimizes the Belief is (0.0, 0.3, 0.2, 0.2, 0.0, 0.0, 0.1, 0.0, 0.2), as shown in Figure 4.12. The Plausibility and the Belief of $T$ calculated with the maximizing and minimizing solutions, respectively, are as follows:

$$\begin{aligned} Pla(T) &= m^{1,1} + m^{1,3} + m^{2,2} + m^{2,3} + m^{3,1} + m^{3,2} + m^{3,3} \\ &= 0.1 + 0.4 + 0.2 + 0.0 + 0.2 + 0.1 + 0.0 = 1 \\ Bel(T) &= m^{1,1} + m^{2,2} = 0.0 + 0.0 = 0 \end{aligned} \tag{4.99}$$

These random relations also yield the upper and lower probabilities that the second cartridge is A, given that the first cartridge is A (event $E$ = {(A,A), (A,B)}), which can be calculated using Eqs. (3.32):

$$\begin{aligned} &Bel\big(\{(\text{A,A}),(\text{B,A})\}/\{(\text{A,A}),(\text{A,B})\}\big) = \frac{Bel(\{(\text{A,A})\})}{Bel(\{(\text{A,A})\}) + Pla(\{(\text{A,B})\})} = \\ &= \frac{0.0}{0.0+0.5} = 0 \\ &Pla\big(\{(\text{A,A}),(\text{B,A})\}/\{(\text{A,A}),(\text{A,B})\}\big) = \frac{Pla(\{(\text{A,A})\})}{Pla(\{(\text{A,A})\}) + Bel(\{(\text{A,B})\})} = \\ &= \frac{0.7}{0.7+0.0} = 1 \end{aligned} \tag{4.100}$$

This interval is larger than [0.3, 0.7], the interval of marginal probabilities that the cartridge picked from the second box is of type A. Once again, we find that learning the type of cartridge changes our uncertainty about the type of the other cartridge, contrary to the intuitive notion of independence. On the other hand, if event $E$ occurs, the upper and lower posterior probabilities for the event $T$={(A,A), (B,B)}, are calculated as follows:

$$Bel(T \mid E) = \frac{Bel(T \cap E)}{Bel(T \cap E) + Pla(T^c \cap E)} = \frac{Bel(\{(\mathrm{A,A})\})}{Bel(\{(\mathrm{A,A})\}) + Pla(\{(\mathrm{A,B})\})} =$$
$$= \frac{0}{0 + (0.3 + 0.2 + 0.2)} = \frac{0}{0.7} = 0 \tag{4.101}$$
$$Pla(T \mid E) = \frac{Pla(\{(\mathrm{A,A})\})}{Pla(\{(\mathrm{A,A})\}) + Bel(\{(\mathrm{A,B})\})} = \frac{0.7}{0.7 + 0} = \frac{0.7}{0.7} = 1$$

This interval is the same as the prior interval because the information obtained does not reduce uncertainty, which remains the maximum possible. Since the posterior event {(B, B)} is impossible (because prior information is that the first cartridge is A), the calculated interval is the same as in Eq. (4.100).

**Example 4.14.** Let us consider the two random sets introduced in Example 4.2 (page 120). Table 4.25 gives the focal elements for the set of random relations along with the probability assignment. For $T = \left\{\left(s_1^1, s_2^1\right), \left(s_1^2, s_2^2\right), \left(s_1^3, s_2^3\right)\right\}$, problems (4.90) are written as:

$$\text{maximize} \sum_{i,j: T \cap \left(A_1^i \times A_2^j\right) \neq \varnothing} m^{i,j} = m^{1,1} + m^{1,2} + m^{1,3} + m^{2,1} + m^{2,2} + m^{2,3} + m^{3,2}$$

*subject to*:

$$\begin{aligned}
&m^{1,1} + m^{1,2} + m^{1,3} = m_1\left(A_1^1\right) = 0.1 \\
&m^{2,1} + m^{2,2} + m^{2,3} = m_1\left(A_1^2\right) = 0.6 \\
&m^{3,1} + m^{3,2} + m^{3,3} = m_1\left(A_1^3\right) = 0.3 \\
&m^{1,1} + m^{2,1} + m^{3,1} = m_2\left(A_2^1\right) = 0.1 \\
&m^{1,2} + m^{2,2} + m^{3,2} = m_2\left(A_2^3\right) = 0.1 \\
&m^{1,3} + m^{2,3} + m^{3,3} = m_2\left(A_2^3\right) = 0.8
\end{aligned} \tag{4.102}$$

Vector **c** and matrix **M** take the values:

$$\mathbf{c} = (1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 0,\ 1,\ 0)^{\mathrm{T}}$$
$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \tag{4.103}$$

Notice that matrix **M** is the same as in Eq.(4.98) because matrix **M** only depends on the number of focal elements in each marginal (Eq. (4.96)), and both Example 4.13 and Example 4.14 have 3 focal elements per marginal.

By using the simplex method, the maximizing solution is found to be equal to (0.0, 0.0, 0.1, 0.0, 0.0, 0.6, 0.1, 0.1, 0.1). The minimum *Bel*(*T*) is achieved by setting **c**= $(0, 0, 0, 0, 0, 1, 0, 0, 0)^T$, and yields the joint basic probability assignment (0.0, 0.0, 0.1, 0.1, 0.1, 0.4, 0.0, 0.0, 0.3). The Plausibility and the Belief of *T* calculated with the maximizing and minimizing solutions, respectively, are as follows:

$$\begin{aligned} Pla(T) &= m^{1,1} + m^{1,2} + m^{1,3} + m^{2,1} + m^{2,2} + m^{2,3} + m^{3,2} \\ &= 0.0 + 0.0 + 0.1 + 0.0 + 0.0 + 0.6 + 0.1 = 0.8 \\ Bel(T) &= m^{2,3} = 0.4 \end{aligned} \tag{4.104}$$

and coincide with the values of the upper and lower probabilities, respectively, found in Example 4.2.

On the other hand, the minimization of *Pla*(*T*) (i.e. minimization of Eq. (4.102)) yields the joint basic probability assignment (0.0, 0.1, 0.0, 0.0, 0.0, 0.6, 0.1, 0.0, 0.2), which gives $Bel(T) = m^{2,3} = 0.6 > 0.4$. Therefore, it is necessary to minimize *Bel*(*T*) as explained in Eqs. (4.94) and (4.95), instead of first minimizing the Plausibility and then calculating the Belief as suggested in the literature (e.g., (Fetz 2001)).

**Table 4.25** Example 4.14: Unknown interaction random relations

| Focal element no. | Focal element | Probability assignment |
|---|---|---|
| 1 | $\{s_1^1, s_1^2\} \times \{s_2^1, s_2^2\}$ | $m^{1,1}$ |
| 2 | $\{s_1^1, s_1^2\} \times \{s_2^1, s_2^2, s_2^3\}$ | $m^{1,2}$ |
| 3 | $\{s_1^1, s_1^2\} \times \{s_2^2\}$ | $m^{1,3}$ |
| 4 | $\{s_1^2\} \times \{s_2^1, s_2^2\}$ | $m^{2,1}$ |
| 5 | $\{s_1^2\} \times \{s_2^1, s_2^2, s_2^3\}$ | $m^{2,2}$ |
| 6 | $\{s_1^2\} \times \{s_2^2\}$ | $m^{2,3}$ |
| 7 | $\{s_1^3\} \times \{s_2^1, s_2^2\}$ | $m^{3,1}$ |
| 8 | $\{s_1^3\} \times \{s_2^1, s_2^2, s_2^3\}$ | $m^{3,2}$ |
| 9 | $\{s_1^3\} \times \{s_2^2\}$ | $m^{3,3}$ |

### 4.3.3 Epistemic Independence

First of all, let us show that the definition of epistemic independence (Eq. (4.42)) applied to two marginal random sets yields the set of probability distributions/measures identified in Table 4.23 (fourth row), i.e. $\Psi_E = \Psi_{R,E}$. Indeed, let the set of probabilities for the $k$-th marginal, $\Psi_k$, be as in Eq. (4.6). For all $s_1 \in S_1$, the second condition (4.42) is satisfied with marginal $P_1 \in \Psi_1$, and arbitrary probability measure $P_2^{|s_1} \in \Psi_2$, which can be decomposed using Eq. (4.6):

$$P_1 = \sum_{i=1}^{n_1}\left(m_1\left(A_1^i\right)\cdot P_1^i\right);\; P_2^{|s_1} = \sum_{i=1}^{n_2}\left(m_2\left(A_2^i\right)\cdot P_2^{i|s_1}\right). \tag{4.105}$$

By inserting Eq. (4.105) into the second Equation (4.42), one obtains the sought joint probability for any $(s_1, s_2) \in S_1 \times S_2$:

$$\begin{aligned}
&P(s_1,s_2) = P_1(\{s_1\})\,P_2^{|s_1}(\{s_2\}) = \\
&= \sum_{i=1}^{n_1}\left(m_1\left(A_1^i\right)\cdot P_1^i(\{s_1\})\right)\sum_{i=1}^{n_2}\left(m_2\left(A_2^i\right)\cdot P_2^{i|s_1}(\{s_2\})\right) = \\
&= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} m_1\left(A_1^i\right)\cdot m_2\left(A_2^j\right)\cdot P_1^i(\{s_1\})\,P_2^{j|s_1}(\{s_2\})
\end{aligned} \tag{4.106}$$

One may repeat the same reasoning for the first condition in Eq. (4.42), to obtain:

$$P(s_1,s_2) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} m_1\left(A_1^i\right)\cdot m_2\left(A_2^j\right)\cdot P_1^{i|s_2}(\{s_1\})\,P_2^j(\{s_2\}) \tag{4.107}$$

By comparing Eq. (4.106) with Eq. (4.84), it is evident that the probability assignment and the focal set probabilities are, respectively:

$$\begin{aligned}
&m\left(A^{i,j}\right) = m_1\left(A_1^i\right)\cdot m_2\left(A_2^j\right); \\
&\forall (s_1,s_2) \in A_1^i \times A_2^j,\; \exists P_2^{j|s_1} \in \Psi_2^j \;\; AND \;\; P_1^{i|s_2} \in \Psi_1^i : \\
&\qquad P^{i,j} = P_1^i \cdot P_2^{j|s_1} = P_1^{i|s_2} P_2^j
\end{aligned} \tag{4.108}$$

which are exactly the conditions given in Table 4.23 (fourth row).

The set of joint probability distributions/measures is:

$$\Psi_{\mathrm{R,E}} = \left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \begin{array}{c} \left( m_1\left(A_1^i\right) m_2\left(A_2^j\right) P^{i,j} \right): \\ P^{i,j} = P_1^i \cdot P_2^{j|s_1} = P_1^{i|s_2} P_2^j \end{array} \right\} \tag{4.109}$$

$P^{i,j}$ is a probability distribution on $A^{i,j}$ because it is the product of marginal probabilities on $A_1^i$ and $A_2^j$, respectively. However, it is a special probability distribution because not all joint probability distributions on $A^{i,j}$ can be written as in (4.108b). Therefore, the set of probability measures in Eq. (4.109) *does not* define a random relation. Compare Eq. (4.109) with the set of probability measures obtained using random set independence, Eq. (4.85). Since $P_1^i \cdot P_2^{j|s_1} \in P^{i,j}$, $\Psi_{\mathrm{R,E}} \subseteq \Psi_{\mathrm{R,R}}$ (Fetz and Oberguggenberger 2004).

As for numerical implementation, Eq. (4.109) does not suggest a more efficient algorithm than those introduced in Section 4.2.2. The reader is referred to Example 4.6, which already dealt with marginal random sets.

### *4.3.4 Strong Independence*

Let us construct the set $\Psi_{\mathrm{R,S}}$ of joint probability measures obtained with the choices given in Table 4.23 (fifth row), and the additional constraint that, for a fixed index *i*, the measure $P^{i,j}$ on the Cartesian product $A^{i,j} = A_1^i \times A_2^j$ has the same marginal, $P_1^i$, on focal set $A_1^i$ for all $j$ = 1,…,$n_2$, i.e.:

$$P_1^i = P_1^{i,1} =,\ldots,= P_1^{i,n_2} \tag{4.110}$$

Likewise, for a fixed index *j*, $P^{i,j}$ is forced to have the same marginal, $P_2^j$, on focal set $A_1^i$ for all $i$ =1,…,$n_1$, i.e.:

$$P_2^j = P_2^{1,j} =,\ldots,= P_2^{n_1,j} \tag{4.111}$$

One obtains:

$$\Psi_{\mathrm{R,S}} = \left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left( m_1\left(A_1^i\right) m_2\left(A_2^j\right) \cdot P_1^i \otimes P_2^j \right) \right\} \tag{4.112}$$

In Section 4.2.3, strong independence was characterized by the fact that the set of probability measures, $\Psi_S$, was composed of all product measures (Eq. (4.74)). Since, for the $j$-th marginal random set: $P_j = \sum_{i=1}^{n_j} m_j\left(A_j^i\right) P_j^i$ , the set $\Psi_S$ may be written as:

$$\begin{aligned}\Psi_S &= \left\{P_1 \otimes P_2\right\} = \left\{\left(\sum_{i=1}^{n_1} m_1\left(A_1^i\right) P_1^i\right) \otimes \left(\sum_{j=1}^{n_2} m_2\left(A_2^j\right) \cdot P_2^j\right)\right\} \\ &= \left\{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(m_1\left(A_1^i\right) m_2\left(A_2^j\right) \cdot P_1^i \otimes P_2^j\right)\right\} = \Psi_{R,S}\end{aligned} \tag{4.113}$$

Hence, $\Psi_S = \Psi_{R,S}$. Since $P^{i,j} = P_1^i \otimes P_2^j$ and since not all probability measures on $A^{i,j}$ may be written as such, $\Psi_{R,S}$ *does not* define a random relation with focal elements $A^{i,j} = A_1^i \times A_2^j$ .

Fetz and Oberguggenberger showed that the upper (lower, resp.) probability of an event, $T$, can be determined by solving the following optimization problem that is written in terms of basic probability assignments and focal elements of the marginals (Fetz and Oberguggenberger 2004):

$$\begin{aligned}&\text{maximize (minimize)} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} m_1^i \cdot m_2^j I_T\left(s_1^i, s_2^j\right) \\ &\text{subject to:} \\ &s_1^i \in A_1^i \qquad i = 1,\ldots,n_1 \\ &s_2^j \in A_2^j \qquad j = 1,\ldots,n_2\end{aligned} \tag{4.114}$$

where $I_T$ is the indicator (or characteristic function, see Section 2.2) of the set $T$. Typically, this is a hard problem to solve because there may be many local minima and because the objective function is not continuous. Such a difficulty in solving the problem in terms of marginal random sets is the counterpart to the non-convexity of $\Psi_S$ demonstrated in Section 4.2.3, where it was also shown that upper and lower previsions (and thus probabilities) are in any case achieved at the extreme points of $\Psi_S$. Alternatively to solving problem (4.114), it may thus be advantageous to first find the extreme points of the marginal distribution sets $\Psi_i$ and then proceed as described in Section 4.2.3.

Constraints $s_1^i \in A_1^i$, $i = 1,\ldots,n_1$ in Eq. (4.114) correspond to constraint (4.110) on the marginal probability measures. As shown in Figure 4.13a, they highlight that, given two Cartesian products $A^{i,j}$ and $A^{i,k}$ generated by the same marginal focal set $A_1^i$, $s_1$ must assume the same value on $A^{i,j}$ and $A^{i,k}$. In other words, the joint probabilities are particular selectors of a random relation with focal elements $A^{i,j} = A_1^i \times A_2^j$ and probabilistic assignement $m^{i,j} = m_1^i \cdot m_2^j$, i.e. of $\Psi_{R,R}$ (Eq. (4.85)). In this case, we speak of *consistent* selectors with *consistent* marginals. As a consequence, the set of consistent selectors is not convex because $\Psi_S$ is not convex.

Alternatively, for the selectors of a general random relation, when $s_1$ may assume different values on $A^{i,j}$ and $A^{i,k}$, as shown in Figure 4.13b, marginals will be called *non-consistent*. Likewise for $s_2$. Consistent and non-consistent marginals will be used extensively when dealing with correlation in Section 4.4 [N 4-2].

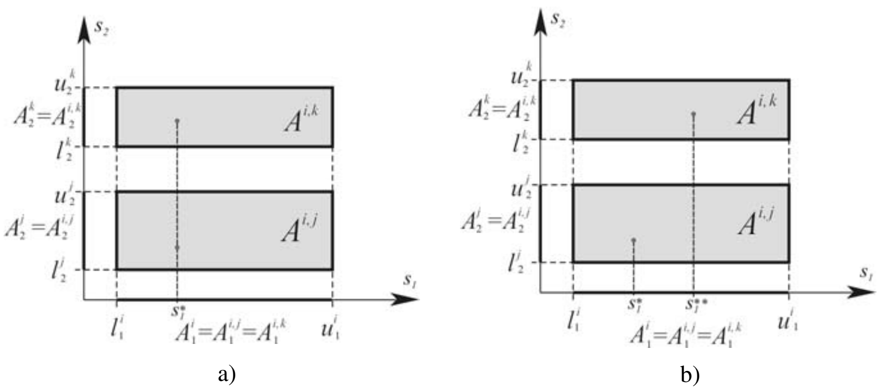



**Fig. 4.13** a) $s_1$ assumes the same value, $s_1$*, on $A^{i,j}$ and $A^{i,k}$ (consistent marginals); b) $s_1$ assumes two different values, $s_1$* and $s_1$**, on $A^{i,j}$ and $A^{i,k}$, respectively (non-consistent marginals)

**Example 4.15.** Let us consider the two boxes of resin as in Example 4.12 (page 160). In the current example, however, the relationship between picking a cartridge from the first box and picking one from the second box is still considered as stochastically independent. In addition, if two cartridges of unknown types are picked, then the types of the replacement cartridges are stochastically independent. Since the set of marginal probabilities generated by the marginal random sets are the same as those considered in Section 4.2.3, the results are exactly the same as in Section 4.2.3.

### 4.3.5 Fuzzy Cartesian Product or Consonant Random Cartesian Product

When marginal random sets $\mathcal{F}_1=\{(A_1, m_1)\}$ and $\mathcal{F}_2=\{(A_2, m_2)\}$ are consonant, they can be expressed as fuzzy sets $F_j$, with membership function $\mu_{F_j}\left(s_j\right)=\sum_{A_j^k : s_j \in A_j^k} m_j\left(A_j^k\right)$ (Eq. (3.24)). Focal element $A_1^i$ is then the α-cut for the α-level $\sum_{A_j^k : A_j^k \supseteq A_j^i} m_j\left(A_j^k\right)$. The fuzzy Cartesian product or consonant random Cartesian product of $F_1$ and $F_2$, $F := F_1 \times F_2$, is then defined as a fuzzy set on $S=S_1\times S_2$ with membership function

$$\mu_{\mathrm{F}}\left(s\right)=\mu_{\mathrm{F}}\left(s_1,s_2\right)=\min\left\{\mu_{F_1}\left(s_1\right),\mu_{F_2}\left(s_2\right)\right\} \tag{4.115}$$

It is easy to see that the α-cut, ${}^{\alpha}A$, of $F$ is the Cartesian product of the α-cuts ${}^{\alpha}A_1$ and ${}^{\alpha}A_2$: ${}^{\alpha}A={}^{\alpha}A_1 \times {}^{\alpha}A_2$. If $S_1$ and $S_2$ are sets of real numbers, than ${}^{\alpha}A$ is a two-dimensional box. Figure 4.14 presents an example of fuzzy Cartesian product: notice that the set, $M = \{\alpha^1,\ldots, \alpha^6\}$, of α-levels is equal to the union of the α-levels of $F_1$ , $M_{F_1}=\left\{\alpha^1,\alpha^3,\alpha^5,\alpha^6\right\}$, and $F_2$, $M_{F_2}=\left\{\alpha^1,\alpha^2,\alpha^4,\alpha^6\right\}$.

A Fuzzy Cartesian product, $F$, defines a consonant random Cartesian product $\left(\left\{A^i\right\},m\right)$ on $S$ by (Eq. (3.25b))

$$\begin{aligned} & m\left(F_1\left(\alpha^i\right)\times F_2\left(\alpha^i\right)\right)=\alpha^i-\alpha^{i+1} && \forall i=1,\ldots,\left|M\right| \\ & m\left(A\right)=0 && \text{otherwise} \end{aligned} \tag{4.116}$$

with $\alpha^{|M|+1} = 0$.

Eq. (4.116) points out that, in a consonant random Cartesian product, the focal sets are not *all* Cartesian products that can be formed with marginal focal elements. Indeed, the focal elements are only the Cartesian products of α-cuts having the same α-level, $\alpha^i$. For example, in Figure 4.14 the consonant random Cartesian product contains six focal elements, whereas there are 4·4=16 possible Cartesian products. Each of these six focal elements is given by a rectangle of uniform gray color and the darker gray rectangles included in it.

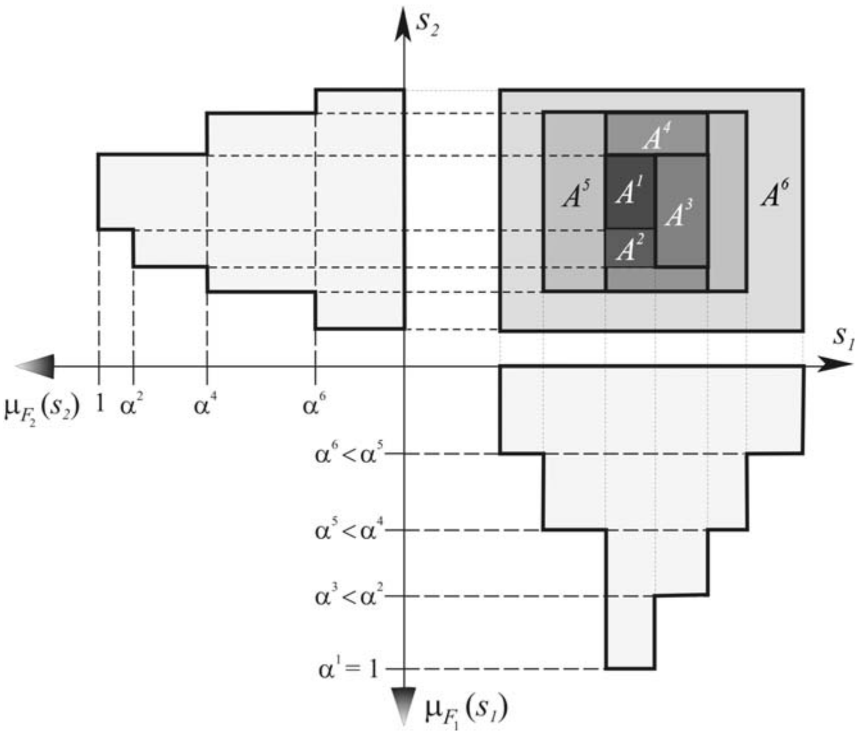



**Fig. 4.14** Fuzzy Cartesian product

This differs from all the other cases of independence introduced in Sections 4.3.1 through 4.3.4 above, where each possible Cartesian product of marginal focal sets was a focal element for the random relation. In this latter case, the obtained Cartesian products are not nested (in general); for example, ${}^{\alpha_3}A_1 \times {}^{\alpha_1}A_2$ is neither a subset nor a superset of ${}^{\alpha_2}A_1 \times {}^{\alpha_2}A_2$. Because of this special choice of the joint focal elements, $\Psi_{R,U} \supseteq \Psi_{R,F}$ ($\Psi_{R,F}$ is the set of probability measures/distributions compatible with $F$) but no general inclusion may be stated between $\Psi_{R,F}$ and $\Psi_{R,R}$, $\Psi_{R,U}$, $\Psi_{R,E}$, $\Psi_{R,S}$.

Since the focal elements for the random relation are only the Cartesian products of α-cuts having the same α-level, a fuzzy Cartesian product (or consonant random Cartesian product) is an appropriate model when $s_1 \in F_1(\alpha^i)$ is always observed at the same time as $s_2 \in F_2(\alpha^i)$, i.e. when the total observational dependence says that intervals of variables $s_1$ and $s_2$ are always jointly observed with the same degree of Possibility.

### 4.3.6 Relationships between the Five Types of Independence

Let us summarize the relationships between the different types of sets of probability measures/distributions introduced in this section for random sets:

$$\begin{aligned}
&\Psi_{\mathrm{R,U}} \supseteq \Psi_{\mathrm{R,R}} \supseteq^{*} \Psi_{\mathrm{R,E}} \supseteq \Psi_{\mathrm{R,S}} \\
&\Psi_{\mathrm{R,U}} \supseteq \Psi_{\mathrm{R,F}} \\
&{}^{*} = \text{proven only for } S \text{ of finite cardinality}
\end{aligned} \tag{4.117}$$

These relationships establish corresponding inclusions for upper and lower probability values calculated by using the different types of independence.

## 4.4 Correlation

In this section, subscripts 1 and 2 indicate two different marginal random sets or variables on two different real lines, $S_1$ and $S_2$. The concept of (linear) correlation plays an important role in probability, statistics, and their applications to systems. In reliability engineering, for example, correlation can have either a positive or a negative effect, leading to either an underestimation or an overestimation, respectively, of the reliability of a system.

To illustrate, consider the reliability of a simple series system composed of two elements, each having a probability of failure equal to 5% ((Ang and Tang 1975), page 48). If the failure events of the two elements are considered to be perfectly correlated (e.g., a chain whose two links are fabricated from the same steel bar by the same manufacturer), then the probability of failure of the chain is $p_{f1} = 0.05$. However, if the two events are independent (e.g., a chain whose two links are randomly selected from two suppliers), then the probability of failure of the chain is $p_{f2} = 0.05 + 0.05 - 0.05^2 = 0.0975 \approx 2\, p_{f1}$.

On the other hand, consider the load, L, on a ground-floor column of a multistory building ((Ang and Tang 1975), page 195). The load contribution from each floor to L is an increasing function of the correlation among floor loads; therefore, the influence of the load correlation on the failure probability is the opposite to that of the series elements.

From Section 2.4, recall that probability theory provides a clear definition for the correlation coefficient between two random variables $x_1$ and $x_2$

$$\rho_{x_1,x_2} = \frac{\sigma_{x_1,x_2}}{\sigma_{x_1}\sigma_{x_2}} \tag{4.118}$$

where $\sigma_{x_1,x_2}$ is the covariance of $x_1$ and $x_2$, and $\sigma_{x_1}$ (resp. $\sigma_{x_2}$) is the standard deviation of $x_1$ (resp. $x_2$) (Eq. (2.43)). Lack of correlation does not imply independence, whereas independence always implies lack of correlation.

However, when dealing with random relations, the concept of correlation is not well understood because focal sets are not singletons. In the following, we will try to shed some light by interpreting a random relation as multivalued mapping (Section 3.1.1) and as a convex set of probability measures (Section 3.2.3). In this section, we restrict ourselves to finite support random relations because topological difficulties would obscure the meaning of the concepts introduced below[N 4-3].

Let us start from two finite discrete random variables $x_1$ and $x_2$, with a joint probability distribution $p_{\mathbf{x}}(\mathbf{a})$ and with correlation coefficient

$\rho_{x_1,x_2}$ calculated with Eq. (4.118). In order to specify that $\rho_{x_1,x_2}$ was obtained with precise mass distribution $p_{\mathbf{x}}(\mathbf{a})$, the symbol $\rho_{x_1,x_2}{}^{(p)}$ will be used when needed. As shown in Figure 4.15, following (Dempster 1967) *imprecision* is introduced through a multi-valued mapping $G$: $(a_1, a_2) \mapsto A$ that maps a pair $\mathbf{a} \equiv (a_1, a_2) \in \mathbb{R}^2$ to a set $A \subseteq S_1 \times S_2 = \mathbb{R}^2$ ($A^{a_1,a_2} \neq \varnothing$ iff $p_{\mathbf{x}}(a_1, a_2) \neq 0$).

For simplicity, in this section $A$ is assumed to be compact and simply connected (i.e., any closed curve contained in $A$ can collapse to a point while remaining within $A$); these assumptions are similar to those of Kruse (Kruse 1987). Likewise, $G$ maps the joint probability mass function $p_{\mathbf{x}}(a_1, a_2)$ to the joint basic probability assignment $m(A)$ such that $\sum m(A) = 1$ and $m(\varnothing) = 0$.

Call $P(G)$ the set of probability measures induced by the selectors $V \in SCT$ (Section 3.2.3.2). Call $P(G)^{(p)}$ the class of measures included in $P(G)$, and whose correlation is $\rho_{x_1,x_2}{}^{(p)}$. In particular, given our assumptions above, the selectors collocate the entire probability measure on one point of each focal set, and thus they correspond to Dirac measures on $A^i$: $P^i = \delta^i$. As a consequence, $\Psi \supseteq P(G) \supseteq P(G)^{(p)}$.

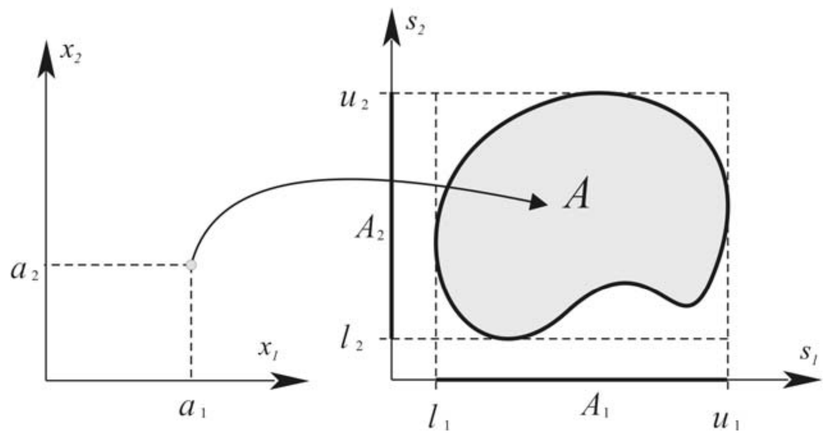



**Fig. 4.15** Multivalued mapping $G$ introduces imprecision

Correlation between two random variables, $x_1$ and $x_2$, is a simple measure of the global orientation of their joint density function. On the other hand, imprecision in $s_1$ and $s_2$ introduced by the multi-valued mapping reflects one's lack of knowledge about their exact values but does not change the underlying cause of this correlation. This suggests that correlation of set-valued random variables should be defined as a set-valued quantity that reflects the orientation of the joint basic probability assignment in $\{A\}$. As a natural extension to the definition of coefficient of correlation, $\rho_{x_1,x_2}$ in Eq. (4.118), it is thus proposed to calculate the coefficient of correlation for the random relation $\mathcal{F}$={(*A*, *m*)} as the set of values of $\rho_{s_1,s_2}$ obtained for all possible selections in *SCT* or Dirac measures in $\Psi^i$. This is equivalent to calculating the coefficient of correlation for all possible realizations of points ($a_1$, $a_2$), each constrained to lie in its focal set $A^i$ with joint mass distribution $p(a_1, a_2) = m(A^i)$

$$\boldsymbol{\rho}_{s_1,s_2} = \left\{ \frac{\sigma_{s_1,s_2}}{\sigma_{s_1}\sigma_{s_2}} : V \in SCT \right\} \tag{4.119}$$

A comparison of $\rho_{x_1,x_2}{}^{(p)}$ with $\boldsymbol{\rho}_{s_1,s_2}$ yields the importance of imprecision introduced by the multi-valued mapping. Since in this section the marginals are compact and simply connected intervals, the random set correlation given by Eq. (4.119) is the interval $\boldsymbol{\rho}_{s_1,s_2} = \left[ \left(\rho_{s_1,s_2}\right)^L, \left(\rho_{s_1,s_2}\right)^U \right]$. Since the random set correlation is an interval, the correlation matrix for $n$ variables is a symmetric interval $n\times n$ matrix with the additional constraint that its realizations must be positive definite.

Notice that one could define the coefficient of correlation for the random relation $\mathcal{F}$={(*A*, *m*)} as the set of values of $\rho_{s_1,s_2}$ obtained for all possible measures in $\Psi$ instead of restricting the search to the selectors. Since the selectors maximize the linear correlation (in the positive and negative directions), the proposed restriction to the selectors does not affect the interval bounds and makes calculations much simpler. A restriction to *EXT*, the set of extreme points of $\Psi$, may not yield the actual correlation bounds because correlation is a non-linear expression of the probability distributions. The random set correlation $\boldsymbol{\rho}_{s_1,s_2}$ can be calculated through the optimization problems described below.

### 4.4.1 The Entire Random Relation Is Given

Let $h_i(s_1, s_2) = 0$ be the implicit representation of the boundary to joint focal set $A^i$, and let $n_f$ be the number of joint focal sets. Recall that in a random relation the probability distributions (and hence Dirac measures, i.e.

selectors) over one joint focal element may be chosen independently from the distributions in another focal element. As a consequence, the optimization problem has $2n_f$ variables and reads:

$$\begin{aligned}\text{Find:}\quad & \left(\boldsymbol{\rho}_{s_1,s_2}\right)^L = \min g\left(a_1^1,\ldots,a_1^{n_f},a_2^1,\ldots,a_2^{n_f}\right)\\ & \left(\boldsymbol{\rho}_{s_1,s_2}\right)^U = \max g\left(a_1^1,\ldots,a_1^{n_f},a_2^1,\ldots,a_2^{n_f}\right)\\ \text{such that:}\quad & h_i\left(a_1^i,a_2^i\right)\le 0 \quad i=1,\ldots,n_f\end{aligned} \tag{4.120}$$

where

$$\begin{aligned}& g\left(a_1^1,\ldots,a_1^{n_f},a_2^1,\ldots,a_2^{n_f}\right)=\\ & =\frac{\sum_{i=1}^{n_f} m\left(A^i\right)\left(a_1^i-\sum_{i=1}^{n_f} m\left(A^i\right)a_1^i\right)\left(a_2^i-\sum_{i=1}^{n_f} m\left(A^i\right)a_2^i\right)}{\sqrt{\sum_{i=1}^{n_f} m\left(A^i\right)\left(a_1^i-\sum_{i=1}^{n_f} m\left(A^i\right)a_1^i\right)^2 \sum_{i=1}^{n_f} m\left(A^i\right)\left(a_2^i-\sum_{i=1}^{n_f} m\left(A^i\right)a_2^i\right)^2}}\end{aligned} \tag{4.121}$$

If some of the projections of sets $A^i$ coincide and available evidence suggests that marginals should be consistent, then the resulting information is no longer a random relation because selectors in different sets $A^i$ are now constrained to have the same projection. Consider the case in which projections $A_j^i$ coincide for $i$ in an index set $I_j$. The optimization problem is (there are $2n_f$ variables)

$$\begin{aligned}\text{Find:}\quad & \left(\boldsymbol{\rho}_{s_1,s_2}\right)^L = \min g\left(a_1^1,\ldots,a_1^{n_f},a_2^1,\ldots,a_2^{n_f}\right)\\ & \left(\boldsymbol{\rho}_{s_1,s_2}\right)^U = \max g\left(a_1^1,\ldots,a_1^{n_f},a_2^1,\ldots,a_2^{n_f}\right)\\ \text{such that:}\quad & \begin{cases} h_i\left(a_1^i,a_2^i\right)\le 0 & i=1,\ldots,n_f\\ a_j^i=a_j^k & i,k\in I_j\end{cases}\end{aligned} \tag{4.122}$$

where $g\left(a_1^1,\ldots,a_1^{n_f},a_2^1,\ldots,a_2^{n_f}\right)$ is given in Eq. (4.121). Likewise for other common projections.

### *4.4.2 Only the Marginals Are Given*

In all of the cases below, the joint probabilistic assignment is unknown.

#### 4.4.2.1 The Joint Mass Correlation, $\rho_{x_1,x_2}^{(p)}$, Is Known

If only the marginal focal sets $A_1$ and $A_2$ are known, then the actual shape of the joint focal sets $A$ is left unspecified, and further information must be introduced to completely define the random relation. For example, one could use the largest set consistent with the marginals: $A = A_1 \times A_2$ (Figure 4.16a).

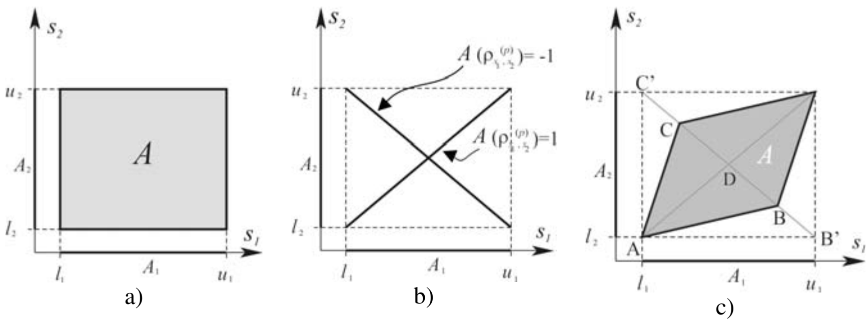



**Fig. 4.16** Focal set $A$ constructed from its marginals $A_1$ and $A_2$ for: a) $\rho_{x_1,x_2}^{(p)} = 0$; b) $\rho_{x_1,x_2}^{(p)} = 1$; and $\rho_{x_1,x_2}^{(p)} = -1$; **c)** $0 < \rho_{x_1,x_2}^{(p)} < 1$

Let us now consider the case of uncorrelated variables. Recall that the class of uncorrelated variables is a superset of the class of independent variables (e.g., Section 2.4 and (Ferson, Nelsen et al. 2004) Figure 10 on page 43) and that all definitions of independence for random sets (Section 4.2) use focal sets $A = A_1 \times A_2$ because joint focal sets that are subsets of $A_1 \times A_2$ "would describe specific types of dependence and thus will not enter [an] investigation of independence" ((Fetz 2005) page 85). Thus, if $A = A_1 \times A_2$ for independent variables, the case for uncorrelated variables should be even more so. For correlated variables, the fact that focal sets are not singletons opens the way for two possible interpretations

1) No focal set correlation
2) Focal set correlation

According to the first interpretation, each joint focal set, $A$, represents an imprecise observation of a point in $\mathbb{R}^2$, and thus correlation should be measured among different focal sets, and not within a particular joint focal set. Under the no focal set correlation interpretation, a joint focal set, $A_{nc}^i$ is just the Cartesian product of two marginal focal sets. Let $\Psi_{nc}^i$ indicate the set of probability measures, $P^{A_{nc}^i}$, which are zero outside the focal set $A_{nc}^i$, and let $\Psi_{nc}$ be the convex set

$$\Psi_{nc} = \left\{ P_{RS,nc} : P_{RS,nc} = \sum_{i=1}^{n} m\left(A_{nc}^i\right) P^{A_{nc}^i} \right\} \tag{4.123}$$

The second interpretation (focal set correlation)[N 4-5] hinges on the interpretation of a random set as a set of probability measures (Eq. (4.5)). This leads to a local interpretation of correlation, whereby each $P^{A^i} \in \Psi^i$ must also display correlation between $s_1$ and $s_2$ (focal set correlation). In the extreme case of perfect correlation, focal set $A$ should reduce to a segment. Since this segment must project as segments $A_1$ and $A_2$, the segment will join the following points (Figure 4.16b)

- $(l_1, l_2)$ and $(u_1, u_2)$ if variables are perfectly positively correlated; or
- $(l_1, u_2)$ and $(u_1, l_2)$ if variables are perfectly negatively correlated.

In the intermediate cases, it is proposed to linearly interpolate in a generalized manner between these extremes[N 4-4]. To illustrate, consider the case $0 < \rho_{x_1,x_2}{}^{(p)} < 1$ depicted in Figure 4.16c. Points $(l_1, l_2)$ and $(u_1, u_2)$ are in $A$ both in the case of uncorrelated variables and in the case of perfectly positively correlated variables; therefore, they should belong to $A$ in any intermediate case. In order to preserve symmetry with respect to these two points and in order to interpolate linearly, it is proposed to use the convex hull of these two points and the two points B and C on the diagonal $B'C'$ as illustrated in Figure 4.16c, in which $|BD| = \left(1 - \rho_{x_1,x_2}{}^{(p)}\right)|B'D|$ and $|CD| = \left(1 - \rho_{x_1,x_2}{}^{(p)}\right)|C'D|$[N 4-6]. Let $\Psi_c^i$ indicate the set of probability measures, which are zero outside the focal set $A_c^i$ constructed this way, and let $\Psi_c$ be the set

$$\Psi_c = \left\{ P_{RS,c} : P_{RS,c} = \sum_{i=1}^{n} m\left(A_c^i\right) P^{A_c^i} \right\} \tag{4.124}$$

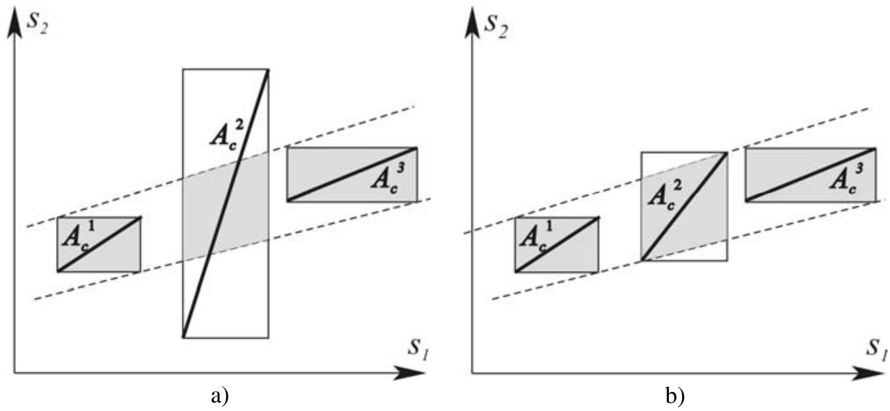



**Fig. 4.17** a) Random relation made up of three focal sets for $\rho_{x_1,x_2}{}^{(p)} = 1$; light dashed lines indicate the widths of the marginals; b) Coherent focal sets

For $\rho_{x_1,x_2}{}^{(p)} = 1$, Figure 4.17a illustrates the case of a random relation with three focal sets. The straight solid lines represent the focal sets, $A_c^i$, constructed according to focal set correlation. The hatched focal sets indicate all points that can be covered by selectors with $\rho_{x_1,x_2}{}^{(p)}=1$. Since the straight solid lines are not completely in the hatched areas, $\Psi_c \not\subset P(G)^{(p)}$, and since the solid lines do not include all hatched areas, $P(G)^{(p)} \not\subset \Psi_c$. This situation arises because the marginal focal sets for $A^2$ are not coherent with the hypothesis that $\rho_{x_1,x_2}{}^{(p)}=1$. Marginal focal sets for $A^2$ have been corrected in Figure 4.17b by reducing $A_2^2$, and thus $P(G)^{(p)} \supseteq \Psi_c$. The optimization problems that follow will enforce such reductions by considering only the distributions in *SCT* that satisfy the assigned correlation value $\rho_{x_1,x_2}{}^{(p)}$.

Let $\boldsymbol{\rho}_{x_1,x_2,nc}$ and $\boldsymbol{\rho}_{x_1,x_2,c}$ be the correlation intervals calculated using Eq. (4.119) under the hypotheses of no focal set correlation and focal set correlation, respectively. Since, for a common joint probabilistic assignment, $\Psi_{nc} \supseteq P(G) \supseteq P(G)^{(p)} \supseteq \Psi_c$, then $\boldsymbol{\rho}_{x_1,x_2,nc} \supseteq \boldsymbol{\rho}_{x_1,x_2,c}$.

From a computational viewpoint, firstly, consider the case in which $\rho_{x_1,x_2}{}^{(p)} = 0$, or the case of no focal set correlation. Let $\rho_{s_1,s_2}{}^{(m)}$ be the correlation coefficient calculated at the midpoints of the focal sets using Eq. (4.118). Let $n_1$ and $n_2$ be the number of marginal focal sets for $s_1$ and $s_2$, respectively. Then, the $n_f = n_1 \cdot n_2$ focal sets $A^{i,j}$, with unknown probabilistic assignements $m^{i,j} = m(A^{i,j})$, are the Cartesian products of marginals $A_1^i$ and $A_2^j$, and one can distinguish the two following cases

- Consistent marginals (the optimization problem has $n_1 + n_2 + n_1 \cdot n_2$ unknown variables)

$$
\begin{aligned}
&\text{Find:} \quad \left(\boldsymbol{\rho}_{s_1,s_2}\right)^L = \min g_c\left(a_1^1,\ldots,a_1^{n_1},a_2^1,\ldots,a_2^{n_2},m^{1,1},\ldots,m^{n_1,n_2}\right)\\
&\left(\boldsymbol{\rho}_{s_1,s_2}\right)^U = \max g_c\left(a_1^1,\ldots,a_1^{n_1},a_2^1,\ldots,a_2^{n_2},m^{1,1},\ldots,m^{n_1,n_2}\right)\\
&\text{such that:} \begin{cases}
l_1^i \le a_{1,i} \le u_1^i\ ,\ i=1,\ldots,n_1\\
l_2^j \le a_{2,j} \le u_j^j\ ,\ \ j=1,\ldots,n_2\\
m^{i,j} \ge 0\ ,\ i=1,\ldots,n_1,\ \ j=1,\ldots,n_2\\
\sum_{j=1}^{n_2} m^{i,j} = m_1\left(A_1^i\right),\ i=1,\ldots,n_1\\
\sum_{i=1}^{n_1} m^{i,j} = m_2\left(A_2^j\right),\ j=1,\ldots,n_2\\
\rho_{s_1,s_2}{}^{(m)}\left(m^{1,1},\ldots,m^{n_1,n_2}\right) = \rho_{x_1,x_2}{}^{(p)}
\end{cases}
\end{aligned}
\tag{4.125}
$$

- Non-consistent marginals (the optimization problem has $3(n_f = n_1 \cdot n_2)$ variables

Find:

$$
\begin{aligned}
&\left(\boldsymbol{\rho}_{s_1,s_2}\right)^L = \min g_{nc}\left(a_1^{1,1},\ldots,a_1^{n_1,n_2},a_2^{1,1},\ldots,a_2^{n_1,n_2},m^{1,1},\ldots,m^{n_1,n_2}\right)\\
&\left(\boldsymbol{\rho}_{s_1,s_2}\right)^U = \max g_{nc}\left(a_1^{1,1},\ldots,a_1^{n_1,n_2},a_2^{1,1},\ldots,a_2^{n_1,n_2},m^{1,1},\ldots,m^{n_1,n_2}\right)\\
&\text{such that:} \begin{cases}
l_1^i \le a_1^{i,j} \le u_1^i\ ,\ i=1,\ldots,n_1;\ j=1,\ldots,n_2\\
l_2^j \le a_2^{i,j} \le u_2^j,\ \ j=1,\ldots,n_2;\ i=1,\ldots,n_1\\
m^{i,j} \ge 0\ ,\ i=1,\ldots,n_1,\ \ j=1,\ldots,n_2\\
\sum_{j=1}^{n_2} m^{i,j} = m_1\left(A_1^i\right),\ i=1,\ldots,n_1\\
\sum_{i=1}^{n_1} m^{i,j} = m_2\left(A_2^j\right),\ j=1,\ldots,n_2\\
\rho_{s_1,s_2}{}^{(m)}\left(m^{1,1},\ldots,m^{n_1,n_2}\right) = \rho_{x_1,x_2}{}^{(p)}
\end{cases}
\end{aligned}
\tag{4.126}
$$

where

$$g_c\left(a_1^1,\ldots,a_1^{n_1},a_2^1,\ldots,a_2^{n_2},m^{1,1},\ldots,m^{n_1,n_2}\right)=$$
$$=\frac{\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}\left(a_1^i-\sum_{i=1}^{n_1}m_1\left(A_1^i\right)a_1^i\right)\left(a_2^j-\sum_{j=1}^{n_2}m_2\left(A_2^j\right)a_2^j\right)}{\sqrt{\sum_{i=1}^{n_1}m_1\left(A_1^i\right)\left(a_1^i-\sum_{i=1}^{n_1}m_1\left(A_1^i\right)a_1^i\right)^2\sum_{j=1}^{n_2}m_2\left(A_2^j\right)\left(a_2^j-\sum_{j=1}^{n_2}m_2\left(A_2^j\right)a_2^j\right)^2}} \tag{4.127}$$

$$g_{nc}\left(a_1^{1,1},\ldots,a_1^{n_1,n_2},a_2^{1,1},\ldots,a_2^{n_1,n_2},m^{1,1},\ldots,m^{n_1,n_2}\right)=$$
$$=\frac{\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}\left(a_1^{i,j}-\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}a_1^{i,j}\right)\left(a_2^{i,j}-\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}a_2^{i,j}\right)}{\sqrt{\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}\left(a_1^{i,j}-\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}a_1^{i,j}\right)^2\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}\left(a_2^{i,j}-\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}m^{i,j}a_2^{i,j}\right)^2}} \tag{4.128}$$

Secondly, consider the case of focal set correlation shown in Figure 4.18, in which $0 < \rho_{x_1,x_2}^{(p)} \le 1$. Then, it is possible to recover formulations similar to Eqs. (4.125) and (4.126) by using an affine transformation of planes that maps a distorted rectangle $A^{i,j}$ into a rectangle $\overline{A}^{i,j}$, and that preserves the lengths of the edges $\mathrm{O}^{i,j}\mathrm{B}^{i,j}$ and $\mathrm{O}^{i,j}\mathrm{C}^{i,j}$. Let us show this in the case of Eqs. (4.126). Let $\left(\overline{\mathrm{O}}^{i,j},\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right)$ be a reference system as in Figure 4.18 with unit vectors $\mathbf{v}_{\overline{a}_1^{i,j}}$ and $\mathbf{v}_{\overline{a}_2^{i,j}}$, and let $\mathbf{T}^{i,j}$ be the transformation matrix whose columns are the components of $\mathbf{v}_{a_1^{i,j}}$ and $\mathbf{v}_{a_2^{i,j}}$ written in the bases of the reference system (O, $s_1$, $s_2$). Then coordinates $\left(a_1^{i,j},a_2^{i,j}\right)$ of a point in $A^{i,j}$ can be obtained as functions $h_1^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right)$ and $h_2^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right)$, respectively, of coordinates $\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right)$

$$\begin{pmatrix} a_1^{i,j} \\ a_2^{i,j} \end{pmatrix} = \mathrm{OO}^{i,j} + \mathbf{T}^{i,j}\begin{pmatrix} \overline{a}_1^{i,j} \\ \overline{a}_2^{i,j} \end{pmatrix} := \begin{pmatrix} h_1^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right) \\ h_2^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right) \end{pmatrix} \tag{4.129}$$

Let *spr* $A_1^i = (u_1^i - l_1^i)/2$ be the spread of the projection and $A_1^i$ and let *mid* $A_1^i = (u_1^i + l_1^i)/2$ be its midpoint, with similar definitions valid for $A_2^j$. The components of the following vectors in the base of (O, $s_1$, $s_2$) can be easily determined from Figure 4.18

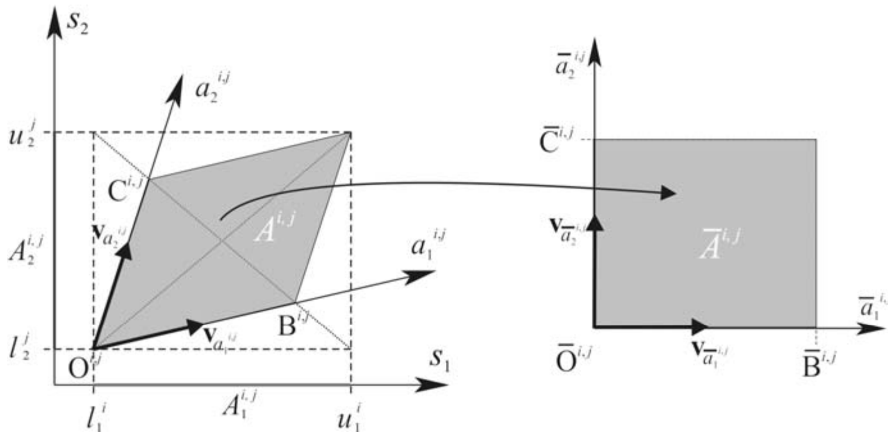



**Fig. 4.18** Affine transformation of focal set $A^{i,j}$ with $\left|\bar{O}^{i,j}\bar{B}^{i,j}\right| = \left|O^{i,j}B^{i,j}\right|$ and $\left|\bar{O}^{i,j}\bar{C}^{i,j}\right| = \left|O^{i,j}C^{i,j}\right|$

$$OO^{i,j} = \left(l_1^i, l_2^j\right)^{\mathrm{T}} = \left(mid\ A_1^i - spr\ A_1^i\ /\ 2,\ mid\ A_2^j - spr\ A_2^j\ /\ 2\right)^{\mathrm{T}} \tag{4.130}$$

$$OB^{i,j} = \left(mid\ A_1^i + \left(1-\rho_{x_1,x_2}{}^{(p)}\right) spr\ A_1^i\ /\ 2,\ mid\ A_2^j - \left(1-\rho_{x_1,x_2}{}^{(p)}\right) spr\ A_2^j\ /\ 2\right)^T \tag{4.131}$$

$$OC^{i,j} = \left(mid\ A_1^i - \left(1-\rho_{x_1,x_2}{}^{(p)}\right) spr\ A_1^i\ /\ 2,\ mid\ A_2^j + \left(1-\rho_{x_1,x_2}{}^{(p)}\right) spr\ A_2^j\ /\ 2\right)^T \tag{4.132}$$

from which one obtains

$$O^{i,j}B^{i,j} = OB^{i,j} - OO^{i,j} = \left(\left(2-\rho_{x_1,x_2}{}^{(p)}\right) spr\ A_1^i\ /\ 2,\ \rho_{x_1,x_2}{}^{(p)} \cdot spr\ A_2^j\ /\ 2\right)^{\mathrm{T}} \tag{4.133}$$

$$O^{i,j}C^{i,j} = OC^{i,j} - OO^{i,j} = \left(\rho_{x_1,x_2}{}^{(p)} \cdot spr\ A_1^i\ /\ 2,\ \left(2-\rho_{x_1,x_2}{}^{(p)}\right) \cdot spr\ A_2^j\ /\ 2\right)^{\mathrm{T}} \tag{4.134}$$

The unit vectors $\mathbf{v}_{a_1^{i,j}}$ and $\mathbf{v}_{a_2^{i,j}}$ (written in the basis of O, $s_1$, $s_2$) are simply

$$\mathbf{v}_{a_1^{i,j}} = \frac{O^{i,j}B^{i,j}}{\left|O^{i,j}B^{i,j}\right|}; \quad \mathbf{v}_{a_2^{i,j}} = \frac{O^{i,j}C^{i,j}}{\left|O^{i,j}C^{i,j}\right|} \tag{4.135}$$

As a result, Eqs. (4.126) become

$$
\begin{aligned}
&\text{Find:}\\
&\left(\boldsymbol{\rho}_{s_1,s_2}\right)^L = \min g_{nc}\left(h_1^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right),h_2^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right),m^{i,j}: i=1,\dots,n_1;\ j=1,\dots,n_2\right)\\
&\left(\boldsymbol{\rho}_{s_1,s_2}\right)^U = \max g_{nc}\left(h_1^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right),h_2^{i,j}\left(\overline{a}_1^{i,j},\overline{a}_2^{i,j}\right),m^{i,j}: i=1,\dots,n_1;\ j=1,\dots,n_2\right)\\
&\text{such that: } \begin{cases}
0 \le \overline{a}_1^{i,j} \le \left|\overline{\mathrm{O}}^{i,j}\mathrm{B}^{i,j}\right|,\ i=1,\dots,n_1;\ j=1,\dots,n_2\\
0 \le \overline{a}_2^{i,j} \le \left|\overline{\mathrm{O}}^{i,j}\mathrm{C}^{i,j}\right|,\ j=1,\dots,n_2;\ i=1,\dots,n_1\\
m^{i,j} \ge 0,\ i=1,\dots,n_1,\ \ j=1,\dots,n_2\\
\sum_{j=1}^{n_2} m^{i,j} = m_1\left(A_1^i\right),\ i=1,\dots,n_1\\
\sum_{i=1}^{n_1} m^{i,j} = m_2\left(A_2^j\right),\ j=1,\dots,n_2\\
\rho_{s_1,s_2}{}^{(m)}\left(m^{1,1},\dots,m^{n_1,n_2}\right) = \rho_{x_1,x_2}{}^{(p)}
\end{cases}
\end{aligned}
\tag{4.136}
$$

Problems (4.125), (4.126), and (4.136) are similar to the problem tackled by Ferson *et al*. (Ferson, Ginzburg et al. 2002; Ferson, Ginzburg et al. 2002; Ferson, Ginzburg et al. 2005), namely determining exact bounds for the correlation of interval data. Therefore, one can easily reproduce the steps of their proof to demonstrate that problems (4.125), (4.126), and (4.136) are NP-hard; moreover, currently there are no available algorithms that may work in many practical situations. Thus, the authors resorted to branch and bound techniques[N 4-7].

#### 4.4.2.2 The Joint Mass Correlation, $\rho_{x_1,x_2}{}^{(p)}$, Is Unknown

When starting from two marginal random sets $\mathcal{F}_1=\{(A_1, m_1)\}$ and $\mathcal{F}_2=\{(A_2, m_2)\}$ and when the precise joint mass correlation coefficient, $\rho_{x_1,x_2}{}^{(p)}$, is unknown, it is proposed to consider again two hypotheses:

1) No focal set correlation: $A = A_1 \times A_2$.

2) Focal set correlation: by symmetry, it is proposed to estimate $\rho_{x_1,x_2}{}^{(p)}$ based on the two sets of midpoints of the marginal intervals; $\rho_{x_1,x_2}{}^{(p)}$ is thus a measure of the correlation between the midpoints of the focal sets $A$, and of the correlation within each focal

set $A$. Focal sets $A$ are then constructed as specified above and illustrated in Figure 4.16. The proposed estimation of $\rho_{x_1,x_2}{}^{(p)}$ may be based on information-theoretic arguments as follows. Given the total ignorance on each marginal focal set, upper and lower probabilities are vacuous over each marginal focal set. According to Nguyen's interpretation of a membership function (Nguyen 1979), a unit rectangular membership function, say $\mu(s) = 1$, is superimposed on each focal set, say $A_1^i$. This leads to vacuous upper and lower probabilities on $A_1^i$. Choosing one element in $A_1^i$ is equivalent to the defuzzyfication process explained by Klir and Yuan (Klir and Yuan 1995). The following, most general, parametrized family of defuzzyfication methods is considered (Klir and Yuan 1995)

$$d_\alpha(\mu) = \frac{\int_{\left(A_1^i\right)^L}^{\left(A_1^i\right)^U} \left(\mu(s)\right)^\alpha s ds}{\int_{\left(A_1^i\right)^L}^{\left(A_1^i\right)^U} \left(\mu(s)\right)^\alpha ds}$$

The choice of parameter $\alpha$ should conform to general principles of uncertainty (Klir 2005); for example, the application of the invariance principle of uncertainty (Klir 2005) yields a unique value for $\alpha$. However, regardless of the value of $\alpha$, $d_\alpha(\mu)$ is always the interval midpoint when $\mu(x) = 1$.

When the joint correlation is unknown, the following modifications must be made to optimization problems in Section 4.4.2.1

- Under the no focal set correlation assumption, the problems are the same as in Eqs. (4.125) and (4.126), where the last constraint is removed.
- Under the focal set correlation assumption, the optimization problem can be divided into two uncoupled non-linear optimization problems. In the first problem, the range for $\rho_{s_1,s_2}{}^{(m)}$ is determined using the $n_1 \cdot n_2$ probability assignment values $m^{i,j}$ as the design variables:

$$\text{Find: } \min \rho_{s_1,s_2}^{(m)}, \max \rho_{s_1,s_2}^{(m)}$$

$$\text{such that: } \begin{cases} m^{i,j} \geq 0\,,\ i=1,\ldots,n_1,\ \ j=1,\ldots,n_2 \\ \sum_{j=1}^{n_2} m^{i,j} = m_1\left(A_1^i\right),\ i=1,\ldots,n_1 \\ \sum_{i=1}^{n_1} m^{i,j} = m_2\left(A_2^j\right),\ j=1,\ldots,n_2 \end{cases} \tag{4.137}$$

If $\min \rho_{s_1,s_2}^{(m)} \cdot \max \rho_{s_1,s_2}^{(m)} \leq 0$, $\rho_{s_1,s_2}$ can be equal to zero, and the second optimization problem is thus the same as for the no focal set correlation case above, i.e. Eqs. (4.125) and (4.126), where the last constraint is removed.

If $\min \rho_{s_1,s_2}^{(m)} \cdot \max \rho_{s_1,s_2}^{(m)} > 0$, then the interval $\boldsymbol{\rho}_{s_1,s_2}$ is calculated by solving Problem (4.136) for $\rho_{s_1,s_2}^{(p)}$ = SIGN($\min \rho_{s_1,s_2}^{(m)}$) min($|\min \rho_{s_1,s_2}^{(m)}|$, $|\max \rho_{s_1,s_2}^{(m)}|$).

**Example 4.16.** In order to gain insight into the correlation coefficient for random sets in the spirit of Figure 4.15, different multi-valued mappings were applied to a precise probability mass function whose correlation coefficient, $\rho_{x_1,x_2}^{(p)}$, was then changed to investigate its effect on the derived random relation.

A continuous bivariate normal distribution with means $\mu_1^{(c)} = 4$ and $\mu_2^{(c)} = 3$, standard deviations $\sigma_1^{(c)} = 0.8$ and $\sigma_2^{(c)} = 0.3$, and correlation coefficient $\rho_{x_1,x_2}^{(c)}$ was integrated in a 4 by 1.5 rectangle centered at $(\mu_1^{(c)}, \mu_2^{(c)})$. This rectangle was divided into $8 \cdot 10 = 80$ equal sub-rectangles with edges 0.5 by 0.15. The normalized integral of the bivariate normal distribution over a sub-rectangle centered at $(a_1, a_2)$ in the original space was used as discrete mass, $p_{\mathbf{x}}(a_1, a_2)$. Table 4.26 gives the statistics of the discrete distribution mass function $p_{\mathbf{x}}(a_1, a_2)$.

Two sets of examples were run: in Set 1, the random relation was assigned; in Set 2, only the marginal random sets were assigned, with no information on the correlation; marginal random sets were the same as the marginal random sets obtained from the random relation in Set 1. In order to exemplify the consequences of focal set correlation, in Set 1 the multivalued mapping was constructed using both focal set correlation and no focal set correlation.

**Table 4.26** Example 4.16: Parameters for the continuous bivariate normal distribution (superscript “c”) and for the discrete distribution obtained from it (superscript “p”)

| $\rho_{x_1,x_2}^{(c)}$ | $\mu_1^{(p)}$ | $\mu_2^{(p)}$ | $\sigma_1^{(p)}$ | $\sigma_2^{(p)}$ | $\rho_{x_1,x_2}^{(p)}$ |
|---|---|---|---|---|---|
| 0.8 | 4 | 3 | 0.775 | 0.289 | 0.723 |
| 0.0 | 4 | 3 | 0.786 | 0.293 | 0.000 |

*Example Set 1: assigned random relation*

Let *spr* $A_j^i$ be the half-length of interval $A_j^i$. As for the multivalued mapping $G$, three cases were considered with *spr* $A_1^i = 0.5\ \sigma_1^{(c)}$ (*spr* $A_2^j = 0.5\ \sigma_2^{(c)}$), *spr* $A_1^i = \sigma_1^{(c)}$ (*spr* $A_2^j = \sigma_2^{(c)}$), and *spr* $A_1^i = 2\ \sigma_1^{(c)}$ (*spr* $A_2^j = 2\ \sigma_2^{(c)}$), respectively, in such a way that point ($a_1$, $a_2$) in the original space was the midpoint of $A$=$G$($a_1$, $a_2$). As a consequence, $\rho_{s_1,s_2}{}^{(m)}$ calculated at the focal sets' mid points (focal set correlation) is the same as $\rho_{x_1,x_2}{}^{(p)}$. The marginal intervals so obtained indicate sufficiently accurate measurements (e.g., (Dantsin, Kreinovich et al. 2006)) because their "narrowed intervals" (*mid* $A_1^i$ - *spr* $A_1^i/8$, *mid* $A_1^i$ + *spr* $A_1^i/8$) do not overlap, likewise for the $s_2$-axis.

To illustrate, Figure 4.19 portrays points ($a_1$, $a_2$) in the original space together with focal sets $A$ derived from them through the multivalued mapping $G$ for focal set correlation.

In the calculation of $\boldsymbol{\rho}_{s_1,s_2}$, both no focal set correlation and focal set correlation assumptions were used, as well as consistent marginals and non-consistent marginals assumptions.

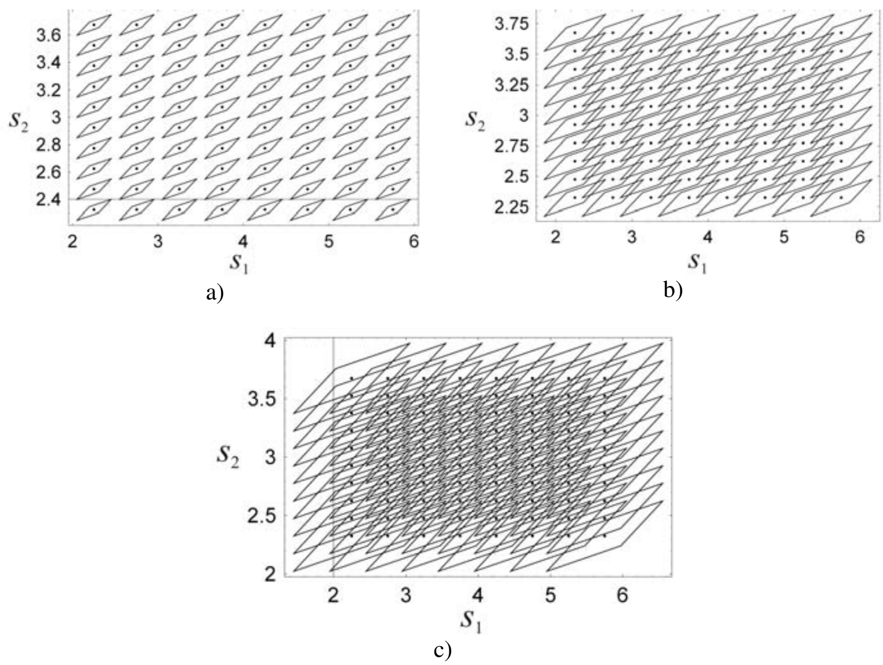



**Fig. 4.19** Points ($a_1$, $a_2$) together with focal sets $A$ derived from them through the multivalued mapping $G$ for $\rho_{x_1,x_2}{}^{(p)} = 0.723$ when: a) *spr* $A_1^i = 0.5\ \sigma_1^{(c)}$ (*spr* $A_2^j = 0.5\ \sigma_2^{(c)}$); b) *spr* $A_1^i = \sigma_1^{(c)}$ (*spr* $A_2^j = \sigma_2^{(c)}$); **c)** *spr* $A_1^i = 2\ \sigma_1^{(c)}$ (*spr* $A_2^j = 2\ \sigma_2^{(c)}$)

A branch-and-bound method (Hammer, Hocks et al. 1995; Kearfott 1996) was used to carry out the calculations, whose results are collected in Table 4.27 through Table 4.29. As expected, the interval $\boldsymbol{\rho}_{x_1,x_2}$ obtained with the no focal set correlation assumption always includes $\boldsymbol{\rho}_{s_1,s_2}$ obtained when focal set correlation is taken into account. For example, compare columns 2 and 3, and 4 and 5 in row 3 of Table 4.27 through Table 4.29. However, by comparing the columns obtained by using consistent marginals with the relevant columns obtained using non-consistent marginals, one notices that the marginal consistency hypothesis has a far greater effect on $\boldsymbol{\rho}_{s_1,s_2}$ than the effect exerted by focal set correlation.

This is especially evident when $\rho_{x_1,x_2}{}^{(p)} = 0$; for example, in Table 4.27 $\boldsymbol{\rho}_{s_1,s_2}$ bounds increased by four orders of magnitude when the hypothesis of consistent marginals was dropped.

Such a hierarchy between the correlation coefficient intervals obtained with non-consistent and consistent marginals parallels the hierarchy between probability bounds for unknown interaction and strong independence, respectively, which was proved by Fetz and Oberguggenberger (Fetz and Oberguggenberger 2004) and Fetz (Fetz 2005).

By analyzing the columns for $\rho_{x_1,x_2}{}^{(p)} = 0.723$ in Table 4.27 through Table 4.29, one notices that, even though $\rho_{x_1,x_2}{}^{(p)}$ was calculated at the midpoints of the focal sets, the interval $\boldsymbol{\rho}_{s_1,s_2}$ is not symmetric about $\rho_{x_1,x_2}{}^{(p)}$; $\rho_{x_1,x_2}{}^{(p)}$ is always much closer to $\left(\boldsymbol{\rho}_{s_1,s_2}\right)^U$ than to $\left(\boldsymbol{\rho}_{s_1,s_2}\right)^L$ regardless of the assumption on focal set correlation. This means that, in the cases examined, imprecision allows for a dispersed or non-linear configuration of the information much more than it allows for a linear interrelationship between the two variables. On the other hand, the intervals $\boldsymbol{\rho}_{s_1,s_2}$ for uncorrelated variables ($\rho_{x_1,x_2}{}^{(p)}=0$) are always symmetric about $\rho_{x_1,x_2}{}^{(p)} = 0$.

**Table 4.27** Example 4.16: Correlation intervals for focal sets with *spr* $A_1^i = 0.5\,\sigma_1^{(c)}$ and *spr* $A_2^j = 0.5\,\sigma_2^{(c)}$

| $\rho_{x_1,x_2}{}^{(p)}$ | Non-consistent marginals | | Consistent marginals | |
|---|---|---|---|---|
| | No focal set correlation | Focal set correlation | No focal set correlation | Focal set correlation |
| 0.723 | [0.3713368, 0.9163020] | [0.5563581, 0.8269218] | [0.6603782, 0.7273850] | [0.7155356, 0.7255974] |
| 0 | [-0.3994516, 0.3994516] | | [-0.0005418034, 0.0005418034] | |

**Table 4.28** Example 4.16: Correlation intervals for focal sets with $spr\ A_1^i = \sigma_1^{(c)}$ and $spr\ A_2^j = \sigma_2^{(c)}$

| $\rho_{x_1,x_2}^{(p)}$ | Non-consistent marginals | | Consistent marginals | |
|---|---|---|---|---|
| | No focal set correlation | Focal set correlation | No focal set correlation | Focal set correlation |
| 0.723 | [-0.074549, 0.979133] | [0.2965469, 0.8925333] | [0.4727766, 0.7273850] | [0.6970032, 0.7270029] |
| 0 | [-0.712030, 0.712030] | | [-0.002046394, 0.002046394] | |

**Table 4.29** Example 4.16: Correlation intervals for focal sets with $spr\ A_1^i = 2\ \sigma_1^{(c)}$ and $spr\ A_2^j = 2\ \sigma_2^{(c)}$

| $\rho_{x_1,x_2}^{(p)}$ | Non-consistent marginals | | Consistent marginals | |
|---|---|---|---|---|
| | No focal set correlation | Focal set correlation | No focal set correlation | Focal set correlation |
| 0.723 | [-0.7897353, 0.9977801] | [-0.4679445, 0.9604998] | [-0.02518543, 0.7273850] | [0.6316610, 0.7273850] |
| 0 | [-0.9681857, 0.9681857] | | [-0.007500380, 0.007500380] | |

*Example Set 2: Only the marginals are given, the joint mass correlation is unknown*

When only the marginals are given and the joint mass correlation is unknown, there is a loss of information with respect to the case in which the entire random relation is given. In order to appreciate this loss, the marginal random sets induced by the random relations of Example Set 1 were used. These random sets are detailed in Table 4.30 and Table 4.31.

As for focal set correlation, it is found that $\min \rho_{s_1,s_2}^{(m)} \cdot \max \rho_{s_1,s_2}^{(m)} \leq 0$, and therefore the Cartesian product of the marginal focal elements was used; as a consequence, there is no difference between focal set correlation and non-focal set correlation. The results given in Table 4.32 - Table 4.34 reveal that the correlation interval is nearly vacuous or vacuous (i.e. [-1, 1]) even for the smallest imprecision, when $spr\ A_1^i = 0.5\ \sigma_1^{(c)}$ ($spr\ A_2^j = 0.5\ \sigma_2^{(c)}$). When compared to the relevant results for Set 1 in Table 4.27 through Table 4.29, the results for Set 2 indicate that there is a complete loss of knowledge in the correlation when only marginals are available. It is notable that these results were obtained even if the marginal intervals were sufficiently accurate measurements because their "narrowed intervals" do not intersect.

We conclude that the large uncertainty in the correlation coefficient is not caused by the imprecision (focal sets are not singletons), but by the difficulty to reconstruct the joint probability assignment, which is a common problem to the theory of precise probability.

**Table 4.30** Example 4.16: Marginal random sets for $\rho_{x_1,x_2}^{(p)} = 0.723$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $mid\ A_1^i$ | 2.25 | 2.75 | 3.25 | 3.75 | 4.25 | 4.75 | 5.25 | 5.75 | | |
| $m_1(A_1^i)$ | 0.0244 | 0.0764 | 0.1626 | 0.2365 | 0.2365 | 0.1626 | 0.0764 | 0.0244 | | |
| $mid\ A_2^i$ | 2.325 | 2.475 | 2.625 | 2.775 | 2.925 | 3.075 | 3.225 | 3.375 | 3.525 | 3.675 |
| $m_2(A_2^i)$ | 0.0167 | 0.0447 | 0.0933 | 0.1518 | 0.1935 | 0.1935 | 0.1518 | 0.0933 | 0.0447 | 0.0167 |

**Table 4.31** Example 4.16: Marginal random sets for $\rho_{x_1,x_2}^{(p)} = 0$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $mid\ A_1^i$ | 2.25 | 2.75 | 3.25 | 3.75 | 4.25 | 4.75 | 5.25 | 5.75 | | |
| $m_1(A_1^i)$ | 0.0270 | 0.0774 | 0.1614 | 0.2342 | 0.2342 | 0.1614 | 0.0774 | 0.0270 | | |
| $mid\ A_2^i$ | 2.325 | 2.475 | 2.625 | 2.775 | 2.925 | 3.075 | 3.225 | 3.375 | 3.525 | 3.675 |
| $m_2(A_2^i)$ | 0.0188 | 0.0460 | 0.0932 | 0.1505 | 0.1915 | 0.1915 | 0.1505 | 0.0932 | 0.0460 | 0.0188 |

**Table 4.32** Example 4.16: Marginals only: correlation intervals for focal sets with $spr\ A_1^{\,i} = 0.5\ \sigma_1^{(c)}$ and $spr\ A_2^{\,j} = 0.5\ \sigma_2^{(c)}$

| $\rho_{x_1,x_2}^{(p)}$ | Non-consistent marginals | | Consistent marginals | |
|---|---|---|---|---|
| | No focal set correlation | Focal set correlation | No focal set correlation | Focal set correlation |
| 0.723 | [-1, 1] | [-1, 1] | [-0.987349, 0.987349] | [-0.987349, 0.987349] |
| 0 | [-1, 1] | | [-0.987542, 0.987542] | |

**Table 4.33** Example 4.16: Marginals only: correlation intervals for focal sets with $spr\ A_1^{\,i} = \sigma_1^{(c)}$ and $spr\ A_2^{\,j} = \sigma_2^{(c)}$

| $\rho_{x_1,x_2}^{(p)}$ | Non-consistent marginals | | Consistent marginals | |
|---|---|---|---|---|
| | No focal set correlation | Focal set correlation | No focal set correlation | Focal set correlation |
| 0.723 | [-1, 1] | [-1, 1] | [-0.995956, 0.995956] | [-0.995956, 0.995956] |
| 0 | [-1, 1] | | [-0.995906, 0.995906] | |

**Table 4.34** Example 4.16: Marginals only: correlation intervals for focal sets with $spr\, A_1^{\,i} = 2\,\sigma_1^{(c)}$ and $spr\, A_2^{\,j} = 2\,\sigma_2^{(c)}$

| $\rho_{x_1,x_2}^{(p)}$ | Non-consistent marginals | | Consistent marginals | |
|---|---|---|---|---|
| | No focal set correlation | Focal set correlation | No focal set correlation | Focal set correlation |
| 0.723 | [-1, 1] | [-1, 1] | [-1, 1] | [-1, 1] |
| 0 | [-1, 1] | | [-1, 1] | |

Figure 4.20a and Figure 4.20d show array plots of the basic probability assignment, $m$, for the complete random relations used in Example Set 1; Figure 4.20a refers to the case in which $\rho_{x_1,x_2}^{(p)} = 0.723$, and Figure 4.20b refers to the case in which $\rho_{x_1,x_2}^{(p)} = 0$. In an array plot, the values in the array $m$ are shown in a discrete array of squares generated in grayscale output, in which zero values are shown white, and the maximum value is shown black.

Figure 4.20b and Figure 4.20c show $m$ calculated by solving the optimization problems for Example Set 2 using the marginals for the random relation with $\rho_{x_1,x_2}^{(p)} = 0.723$ in Example Set 1. Likewise, Figure 4.20e and Figure 4.20f show $m$ calculated by solving the optimization problems for Example Set 2 using the marginals for the random relation with $\rho_{x_1,x_2}^{(p)} = 0$ in Example Set 1. It is notable that the obtained basic probability assignment was the same for both consistent and non-consistent marginals, and for all levels of imprecision considered.

By comparing Figure 4.20b with Figure 4.20e (and Figure 4.20c with Figure 4.20f) it appears that the distributions of the basic probability assignments are very similar even if the marginal random sets were different. Indeed, zero values of the basic probability assignment occur for the same focal elements. Regardless of the consistent or non-consistent marginals assumptions and regardless of the level of imprecision considered, the basic probability assignment that maximizes the correlation is denser around a straight line with positive slope; vice versa, the basic probability assignment that minimizes the correlation is denser around a straight line with negative slope.

Comparison of Figure 4.20a with Figure 4.20b and Figure 4.20c (Figure 4.20d with Figure 4.20e and Figure 4.20f) reveals that the mere knowledge of marginals leads to a wide range of possible distributions of the basic probability assignment, and that reconstructing the basic probability assignment of the original random relation is impossible.

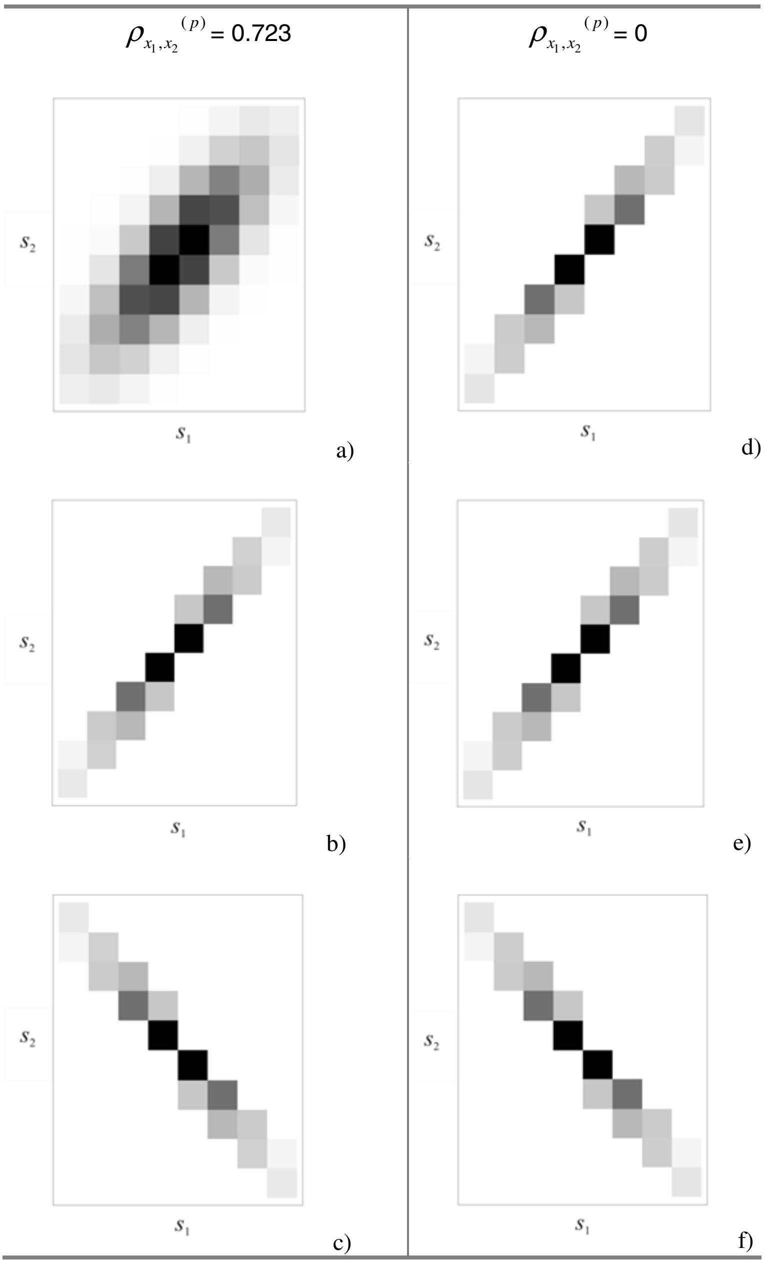



**Fig. 4.20** Array plots of the probability assignments, $m$. (a) (and (d)): $m$ for the complete random relation used in Example Set 1; (b) and (c) (and (e) and (f)): maximizing and minimizing $m$, respectively, for the optimization problems in Example Set 2

## 4.5 Conclusions

Within the theory of imprecise probability, the definitions for unknown interaction, epistemic irrelevance, epistemic independence, and strong independence have been investigated. Two approaches have been proposed to calculate upper and lower probabilities and expectations on the joint distribution: the choice between the two approaches mainly depends on the number of upper and lower probabilities and expectations to be calculated. All algorithms were designed to accommodate two types of constraints over marginal distributions: prevision bounds or convex hulls of extreme distributions. In all cases, linear optimization algorithms were derived: for epistemic irrelevance/independence this was achieved by rewriting the algorithm in terms of the joint distribution. Upper and lower conditional probabilities are always achieved at the extreme points of the set of joint distributions even though the objective function is not linear, when it is a conditional probability. Strong independence generates non-convex sets of joint distributions.

When marginals are random sets, it was explained how these definitions of independence may be recast to generate random relations as sets of convex combinations of probability measures centered at the focal sets and zero outside the focal sets. However, only random set independence yields a unique random relation on the joint space; unknown interaction generates infinite random relations, and epistemic independence and stochastic independence do not generate random relations.

The concept of correlation for two variables constrained by a random relation needs to take into account the imprecision conveyed by the focal sets and is therefore an interval. As a consequence, the correlation matrix for $n$ variables is a symmetric interval $n \times n$ matrix with the additional constraint that its realizations must be positive definite.

When a random relation on the joint space is available, information is affected by imprecision and marginals are intervals rather than singletons. As a consequence, the concepts of consistent and non-consistent marginals were introduced. In the first case, a variable assumes the same value when it belongs to a marginal generated by two or more focal sets; this parallels the situation encountered in strong independence for random sets. In the second case, a variable may assume different values when it belongs to a marginal generated by two or more focal sets; this parallels the situation encountered in unknown interaction and epistemic independence for random sets. Conversely, when only marginals are available as random sets, information on the marginals is affected by imprecision, and marginals do not uniquely define focal sets on the joint space because many focal sets may have the same projection on a given axis. The chapter presented a natural definition of focal sets starting from the marginals.

The calculation of random set correlation was reduced to solving two NP-hard optimization problems, which, however, proved very easy to solve with branch-and-bound algorithms in the numerical experiments carried out. When the entire random relation is available, the numerical examples presented showed that the hypothesis of non-consistent marginals leads to correlation bounds that are much larger (four orders of magnitude in some cases) than those obtained under the hypothesis of consistent marginals. This result parallels the hierarchy between strong independence and epistemic independence. When imprecision is generated through a multi-valued mapping with domain on joint random variables, the random set correlation bounds are not symmetric with respect to the correlation coefficient of the initial random variables. When only the marginals are available, there is a complete loss of correlation knowledge, and the correlation interval is nearly vacuous or vacuous (i.e. [-1, 1]) even if the measurements are sufficiently accurate in that their narrowed intervals do not overlap. Solutions to the optimization problems were found at the extremes of their feasible intervals 50% or less of the times.

## Notes

**N 4-1.** Within this context external to random relations, one may think of creating all possible combinations between definitions of independence for distribution sets $\Psi^{i,j}$ (unknown interaction, epistemic irrelevance, epistemic independence and strong independence) and all possible choices of the basic probability assignment *m*. Except for the combinations lited in Table 4.23, the resulting bewildering number of combinations has not received specific names in the literature, but inclusion relationships between distribution sets may be found in (Fetz and Oberguggenberger 2004).

**N 4-2.** Constraints (4.110) and (4.111) may be used with any of the combinations mentioned in Note N 4-1 to generate additional combinations (Fetz and Oberguggenberger 2004).

**N 4-3.** In the literature, Diamond (Diamond 1990) addressed the problem of fitting an affine function of the kind

$$\boldsymbol{Y} = a\boldsymbol{X}+\boldsymbol{B} \tag{4.138}$$

for interval-valued random data by extending the least-squares optimality criterion for two variables *X* and *Y*, and gave a sufficient condition for non-degenerate elements to admit a unique optimal solution. In Eq. (4.138), $a$ is a real number, $\boldsymbol{B}$ is an interval, $a\boldsymbol{X}$ is the product of the interval $\boldsymbol{X}$ by the scalar $a$, and + denotes the Minkowski addition on the set of real intervals, i.e. $a\boldsymbol{X}+\boldsymbol{B} = \{z + w, z \in a\boldsymbol{X}, w \in \boldsymbol{B}\}$. More recently, Gil *et al.* (Gil, Lubiano et al. 2002) generalized Diamond's study by extending the least-squares method to a generalized metric on the space of nonempty compact intervals, and by finding the optimal solutions $a^*$ and $\boldsymbol{B}^*$ in Eq. (4.138) for the general case of nondegenerate interval-valued random sets (i.e. random sets whose focal sets are all intervals) with necessary and sufficient conditions for the non-uniqueness of the solution. An extended determination coefficient was also defined by Gil *et al.* for the affine function in Eq. (4.138) as a real number between 0 and 1; it is equal to 0 if and only if (iff) $a^* = 0$ (affine independence) and is equal to 1 if the distance between $Y$ and $a^*\boldsymbol{X}+\boldsymbol{B}^*$ is zero for the chosen metric. Kruse (Kruse 1987) interpreted a random set as a measurable map, $G$, defined on a probability space $(\Omega, \mathscr{B}, P)$ and taking values in the set of non-empty, compact subsets of $\mathbb{R}$. A selection is a random variable, $V$, on the same probability space, and such that $V(\omega) \in G(\omega)$. He defined the variance of a random set as the set of variance values for all possible selectors, $V$.

In the fuzzy set literature, Chaudhuri and Bhattacharya (Chaudhuri and Bhattacharya 2001) proposed a definition for the correlation coefficient between two fuzzy sets defined on the same universal support. Chaudhuri and Bhattacharya compared this definition to an earlier different definition by Murthy *et al*. (Murthy, Pal et al. 1985); both definitions yield a real number. Feng *et al.* (Feng, Hu et al. 2001) defined the variance and covariance of fuzzy random variables (i.e. fuzzy numbers whose α-cut extremes are random variables) as two crisp numbers and then applied them to fuzzy stochastic processes. Finally, Liu and Kao (Liu and Kao 2002) used Zadeh's min-max extension principle to extend the definition in Eq. (4.118) to a sample of $n$ pairs of fuzzy numbers $\left(\tilde{X}_i, \tilde{Y}_i\right)$ as follows

$$\tilde{\rho}_{XY} = \frac{\sum_{i=1}^{n}\left(\tilde{X}_i - \sum_{i=1}^{n}\tilde{X}_i/n\right)\left(\tilde{Y}_i - \sum_{i=1}^{n}\tilde{Y}_i/n\right)}{\sqrt{\sum_{i=1}^{n}\left(\tilde{X}_i - \sum_{i=1}^{n}\tilde{X}_i/n\right)^2\left(\tilde{Y}_i - \sum_{i=1}^{n}\tilde{Y}_i/n\right)^2}} \tag{4.139}$$

Fuzzy set $\tilde{\rho}_{XY}$ is approximated by Liu and Kao by calculating a finite number of its α-cuts, $[(\tilde{\rho}_{XY})_\alpha^L, (\tilde{\rho}_{XY})_\alpha^U]$, whose extremes are obtained by solving the following two global optimization problems

$$\begin{aligned}
&\text{Find: } (\tilde{\rho}_{XY})_\alpha^L = \min f(x_1, \ldots, x_n, y_1, \ldots, y_n);\\
&\qquad\quad (\tilde{\rho}_{XY})_\alpha^U = \max f(x_1, \ldots, x_n, y_1, \ldots, y_n)\\
&\text{such that: } (X_i)_\alpha^L \le x_i \le (X_i)_\alpha^U \ \forall i \text{ AND } (Y_i)_\alpha^L \le y_i \le (Y_i)_\alpha^U \ \forall i
\end{aligned} \tag{4.140}$$

where

$$f(x_1, \ldots, x_n, y_1, \ldots, y_n) = \frac{\sum_{i=1}^{n}\left(x_i - \sum_{i=1}^{n}x_i/n\right)\left(y_i - \sum_{i=1}^{n}y_i/n\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \sum_{i=1}^{n}x_i/n\right)^2\left(y_i - \sum_{i=1}^{n}y_i/n\right)^2}} \tag{4.141}$$

The two global optimization problems (4.140) were solved using mathematical programming techniques and software Lingo (LINDO 1999).

In the fuzzy-random literature, Meyer and Kruse (Meyer and Kruse 1990) defined the covariance for two fuzzy-random variables in a similar way, i.e. by applying the min-max extension principle of Zadeh, and thus obtaining a fuzzy set. The authors gave a set representation of the covariance, but found it extremely difficult to calculate. Finally, Meyer and Kruse proved a limit theorem for the estimation of the covariance.

In the interval analysis literature (see the website "http://cs.utep.edu/interval-comp/main.html" for an up-to-date bibliography), Ferson *et al.* (Ferson, Ginzburg et al. 2002; Ferson, Ginzburg et al. 2002; Ferson, Ginzburg et al. 2005) showed that calculating exact bounds for the correlation of two sets of interval data is NP-hard, i.e. there is no feasible algorithm that would always compute the desired bounds for the correlation in polynomial time. Although Ferson *et al.* (Ferson, Ginzburg et al. 2002; Ferson, Ginzburg et al. 2002; Ferson, Ginzburg et al. 2005), Xiang (Xiang 2006), and Dantsin *et al.* (Dantsin, Kreinovich et al. 2006) were able to present quadratic-time or $O(n \cdot \log(n))$ algorithms for calculating the upper bound of the variance of interval data (also an NP-hard problem) that works in many practical cases, at present there are no similar algorithms for calculating correlation bounds.

**N 4-4.** This proposal was put forward by the authors in (Tonon and Pettit 2004; Tonon and Pettit 2005; Tonon and Pettit 2005); as this book was being prepared, Ferson and Kreinovich (Ferson and Kreinovich 2006) gave an extensive taxonomy of one-parameter models of correlation within intervals.

**N 4-5.** Focal set correlation is conceptually coherent with the methods proposed by Chatillon (Chatillon 1984) and Schilling (Schilling 1984) for calculating the coefficient of correlation for singletons. In these methods, the cluster of data points is enclosed by an ellipse, rather than a quadrilateral as proposed here. When the coefficient of correlation tends to zero, Chatillon's and Schilling's methods retrieve a circle, whereas the method proposed here retrieves a box with edges parallel to the axes (see justification above for uncorrelated variables). When the coefficient of correlation is equal to one, both Chatillon's and Schilling's methods and the method proposed here retrieve a straight line.

**N 4-6.** There is another interesting analogy with the coefficient of correlation between two (linearized) failure modes in the normalized space introduced by Hasofer and Lind (Hasofer and Lind 1974) for reliability analysis. This coefficient of correlation is in fact equal to the scalar product of the two unit normals to the hyperplanes defining the two limit states (Ditlevsen 1979). Therefore, the correlation coefficient is equal to zero when the two normals are orthogonal, and equal to one when they are parallel to one another. Likewise, in the proposed method, unit vectors

$\mathrm{AC}/|\mathrm{AC}|$ and $\mathrm{AB}/|\mathrm{AB}|$ (Figure 4.16c) are orthogonal if $\rho_{x_1,x_2}{}^{(p)} = 0$ and they are parallel if $\rho_{x_1,x_2}{}^{(p)}=1$. It is easy to show that the scalar product between $\mathrm{AC}/|\mathrm{AC}|$ and $\mathrm{AB}/|\mathrm{AB}|$ is equal to

$$\mathrm{AC}/|\mathrm{AC}|\cdot\mathrm{AB}/|\mathrm{AB}| = \frac{\rho_{x_1,x_2}{}^{(p)}\left(2-\rho_{x_1,x_2}{}^{(p)}\right)\left(|A_1|^2+|A_2|^2\right)}{\sqrt{\rho_{x_1,x_2}{}^{(p)}|A_1|^2+\left(2-\rho_{x_1,x_2}{}^{(p)}\right)|A_2|^2}\sqrt{\rho_{x_1,x_2}{}^{(p)}|A_2|^2+\left(2-\rho_{x_1,x_2}{}^{(p)}\right)|A_1|^2}}$$

It can be shown that the scalar product $\mathrm{AC}/|\mathrm{AC}|\cdot\mathrm{AB}/|\mathrm{AB}|$ is always greater than $\rho_{x_1,x_2}{}^{(p)}$ and that this is even more so when the ratio $|A_1|/|A_2|$ increases.

**N 4-7.** It is interesting to compare the formulations developed in this section with Liu and Kao's (2002) formulation for the correlation coefficient of a sample of *n* pairs of fuzzy numbers $\left(\tilde{X}_i,\tilde{Y}_i\right)$ (Eqs. (4.140) and (4.141)). If all fuzzy sets have a rectangular membership function, so that they collapse to intervals (i.e., $\left(\tilde{X}_i\right)_1=\left(\tilde{X}_i\right)_0=A_{X,i}$), and one assumes $m_{\mathrm{XY}}(A_{ij}) = 1/n$, then Liu and Kao's formulation yields the same result as the non-consistent marginal formulation (Eqs. (4.126)) if one additionally assumes that the variables are non-interactive. This is because Liu and Kao implicitly use a cylindric extension (e.g., Klir and Yuan, 1995, page 123) to determine a fuzzy relation on $\mathbb{R}^2$ starting from a pair of fuzzy numbers $\left(\tilde{X}_i,\tilde{Y}_i\right)$ defined on the *X* and *Y* axes, respectively. Cylindric extension maximizes the nonspecificity (e.g., Klir and Yuan, 1995, page 123 and Chapter 9) in deriving a fuzzy relation starting from one of its projections, which is consistent with the derivations and discussion presented above in Section 4.4; that is, maximizing the nonspecificity is consistent with the goal of not imposing unjustified assumptions on the available information. In a generalized sense, this is analogous with the philosophy of maximum entropy in probability.

# Chapter 5
# Inclusion and Mapping of Random Sets/Relations

This chapter investigates the concepts of weak and strong inclusion between random sets or relations. Approximations to random sets and relations are constructed by including given random sets or relations into random sets and relations that are easier to deal with from a computational viewpoint: these approximations yield validated outer bounds on the probability of events. Finally, mappings of random sets are investigated along with the monotonicity of inclusions.

## 5.1 Inclusion of Random Sets

### *5.1.1 Weak Inclusion*

The idea of including one random set in another plays a significant role in the ensuing theory and applications of random sets and relations. Let $\mathcal{F}$ and $\overline{\mathcal{F}}$ be two random sets or relations. Since a set of probability distributions and measures, $\Psi$, can be associated to a random set (Section 3.2.3), it is natural to define $\mathcal{F} \subseteq_{weak} \overline{\mathcal{F}}$ if and only if $\Psi \subseteq \overline{\Psi}$. From the basic definitions (Eqs. (3.3), (3.10) and (3.11)), it is clear that this condition is equivalent to the inclusion of Belief-Plausibility bounds, i.e.:

$$\Psi \subseteq \overline{\Psi} \Leftrightarrow \left[ Bel(T), Pl(T) \right] \subseteq \left[ \overline{Bel}(T), \overline{Pl}(T) \right] \forall T \subseteq S \tag{5.1}$$

Using the reservoir-bathtub analogy introduced in Example 3-4, if $\Psi \subseteq \overline{\Psi}$, there are more pipe arrangements in $\overline{\mathcal{F}}$ than in $\mathcal{F}$, and thus the bounds on the possible flow rates into the bathtub are larger, regardless of the size and position of the bathtub. On the other hand, if, for all possible sizes and positions of the bathtub, the bounds on the flow rates are larger, then there are more pipe arrangements in $\overline{\mathcal{F}}$ than in $\mathcal{F}$.

This definition of inclusion is called *weak* because it does not keep track of which focal sets $\overline{A}^j$ include focal sets $A^i$, and of how their basic probability assignments are related. From an operative viewpoint, checking that two random sets are weakly included entails checking either the right-hand-side of Eq. (5.1) for all possible subsets $T \subseteq S$ or the left-hand-side of Eq. (5.1). While the first check is straightforward (one only needs to calculate the Belief, and can then use Eq. (3.7) to calculate the Plausibility), let us expand on the second check.

Since $\Psi$ and $\overline{\Psi}$ are convex, the left-hand-side of Eq. (5.1) is equivalent to checking that all extreme points of $\Psi$ are in $\overline{\Psi}$. Let $v_i$ $i$=1,.., $q$ ($\overline{v}_i$ $i$=1,.., $\overline{q}$, resp.) be the vertices of $\Psi$ ($\overline{\Psi}$, resp.) found by using the algorithm in Section 3.2.3.5. The reader is invited to read again the initial discussion in Section 3.2.3.5 on the need for a reduced probability space of dimension $s_{imp}$ and for projecting vertices onto a $p$-dimensional projected imprecise probability space, where $p = s_{imp} - 1$.

Two different algorithms are now introduced: the first algorithm may be adopted when only the vertices of $\Psi$ and $\overline{\Psi}$ are known; the second algorithm may be used when the entire convex hulls $\Psi$ and $\overline{\Psi}$ (including facets) are known.

#### 5.1.1.1 First Weak Inclusion Algorithm

If only the vertices of $\Psi$ and $\overline{\Psi}$ are known, in order to decide whether $v_i \in \overline{\Psi}$, one has to check whether $v_i$ is a convex combination of the $\overline{v}_i$ or not ((Grötschel, Lovász et al. 1988), page 49), i.e.

$$v_i = \sum_{j=1}^{\overline{q}} \lambda_j \overline{v}_j \text{ with } \sum_{j=1}^{\overline{q}} \lambda_j = 1 \text{ and } \lambda_j \geq 0 \tag{5.2}$$

Figure 5.1a illustrates a 2-D example ($p$ = 2), where $\overline{\Psi}$ is a pentagon ($\overline{q}$ = 5). Notice that $\overline{\Psi}$ is the union of all triangles whose vertices are in $\overline{EXT} = \{\overline{v}_1, \ldots, \overline{v}_5\}$; there are $\begin{pmatrix} \overline{q} \\ p+1 \end{pmatrix}$=10 such triangles, each of which is a convex combination of its vertices. Thus, $v_i \in \overline{\Psi}$ if and only if it is in at least one such triangle. In Figure 5.1a, $v_i$ is in triangles $\overline{v}_2\overline{v}_3\overline{v}_5$ and

$\overline{v}_2\overline{v}_3\overline{v}_4$, and thus there are non-negative triples $\{\lambda_2, \lambda_3, \lambda_5\}^*$ and $\{\lambda_2, \lambda_3, \lambda_4\}^{**}$, which are also the solutions to the following systems of equations (5.3) when $I_k = \{2, 3, 5\}$ and $I_k = \{2, 3, 4\}$, respectively:

$$\begin{cases} v_i = \sum_{j \in I_k} \lambda_j \overline{v}_j \\ \sum_{j \in I_k} \lambda_j = 1 \end{cases} \tag{5.3}$$

In general, $v_i \in \overline{\Psi}$ if and only if a non-negative solution exists to a system (5.3) of $p$+1 maximal equations in $p$+1 coefficients $\lambda_i$ for at least one of the $\binom{\overline{q}}{p+1}$ sets $I_k$ of $p$+1 indexes that can be formed from elements of the set $\{1,..,\overline{q}\}$.

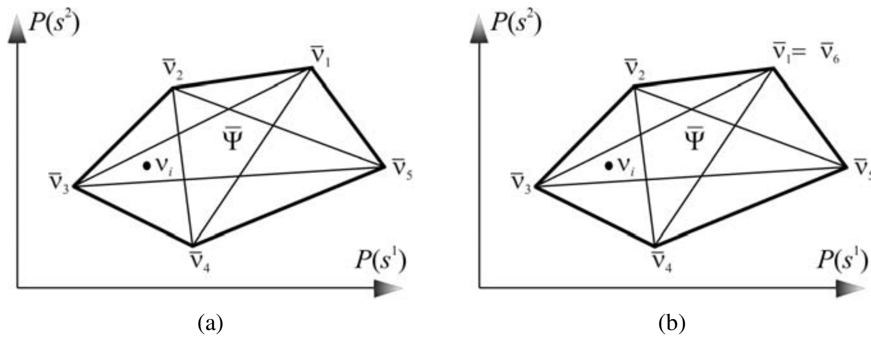



**Fig. 5.1** Point in polygon check: (a) vertex numbering for Algorithm 1; (b) vertex numbering for Algorithm 2

The system (5.3) is singular if and only if vectors in $\{\overline{v}_j : j \in I_k\}$ are linearly dependent, i.e. they all lie on the same facet, $f$, of $\overline{\Psi}$. Two cases may arise:

1) When $v_i \in f$, there are infinite solutions. Figure 5.2a shows a 3-D example in which $I_k$ is any 4-combination of {1, 2, 3, 4, 5}, i.e. any degenerate tetrahedron with vertices in the polygon $\overline{v}_1, \ldots, \overline{v}_5$.

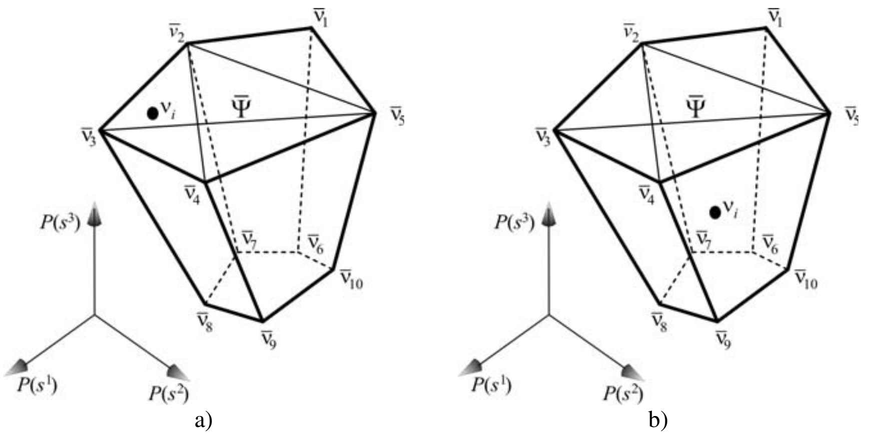



**Fig. 5.2** a) Vertex $v_i$ lies on facet $\overline{v}_1 ... \overline{v}_5$ ; b) Vertex $v_i$ does not lie on facet $\overline{v}_1 ... \overline{v}_5$

2) When $v_i \notin f$, there are no solutions. Figure 5.2b shows a 3-D example in which, when $I_k$ is any 4-combination of {1, 2, 3, 4, 5}, the corresponding degenerate tetrahedron with vertices in the polygon $\overline{v}_1, \dots, \overline{v}_5$ cannot include $v_i$.

Singularity may be detected with no overhead during the LU factorization of non-symmetric system (5.3). Once detected, combination $I_k$ is discarded.

Since the LU factorization of (5.3) accounts for most of the computational cost, it is advisable to first cycle on the combinations $I_k$, and have an inner loop on the vertices of $\Psi$. Let $p = s_{imp}$-1. The general algorithm is thus:

Calculate vertices of $\Psi$ and $\overline{\Psi}$ using the algorithm in Section 3.2.3.5.
Project vertices onto the $p$-dimensional projected imprecise probability space

$$n_k \leftarrow \binom{\overline{q}}{p+1}$$

Calculate combinations $I_k$
presence(1:$q$)=0
outer: DO $k$ = 1 to $n_k$;
    LU-factor (5.3) and store L and U
    IF (5.3) is singular, CYCLE outer
        inner: DO $i$ = 1 to $q$;
            IF presence($i$)==0 THEN
                Solve (5.3) by LU backsubstitution

```
                    IF λ_j ≥ 0,   j ∈ I_k
                    THEN
                          v_i ∈ Ψ̄
                          presence(i) ←1
                          CYCLE inner
                    END IF
                END IF
            END DO inner
END DO outer
IF ∏_{i=1}^{q} presence(i)==1 THEN
      Ψ ⊆ Ψ̄
ELSE
      Ψ ⊄ Ψ̄
END IF
```

In case 1) above, $v_i$ is detected as belonging to the first 4-combination $I_{\mathrm{k}}$ corresponding to a non-degenerate tetrahedron with base $\overline{v}_2\overline{v}_3\overline{v}_5$ or $\overline{v}_2\overline{v}_3\overline{v}_4$ and vertex in $\{\overline{v}_6,\ldots,\overline{v}_{10}\}$. In case 2), $v_i$ is detected as belonging to the first 4-combination $I_{\mathrm{k}}$ corresponding to a non-degenerate tetrahedron that contains $v_i$.

Let $\mathbf{c}$ be a non-zero $\overline{q}$-dimensional column vector, and let $\boldsymbol{\lambda} = \left(\lambda_1,\ldots,\lambda_{\overline{q}}\right)^{\mathrm{T}}$. The problem of minimizing $\mathbf{c}^{\mathrm{T}}\boldsymbol{\lambda}$ subject to constraints (5.2) is a standard linear programming problem. The simplex method searches for a basic solution by solving a problem such as (5.3), and then attempts to improve on it (if it is not an optimal solution and the solution is not unbounded) or it solves a new system of equations (5.3) if the previous basic solution contains some negative components. Thus, the computational complexity of the algorithm above is the same as the simplex algorithm, i.e. exponential in the size of the problem: for $\overline{q}=300$ and $p = 100$, the number of sets $I_k$ is in the order of $10^{81}$. However, the actual cost of the simplex method is much less, and interior point methods (Nemirovsky and Yudin 1994) such as Karmarkar's (Karmarkar 1984) or Mehotra's (Mehrotra 1992) that run in polynomial time become superior only when the number of vertices reaches 15,000 or more ((Kinkaid and Cheney 2002), page 709). The following example shows the effectiveness of the above algorithm.

**Example 5.1.** Let $S = \{a, b, c, d, e\}$, and consider the two random sets: $\mathcal{F}$ = { ({a, b}, 0.3), ({a, c}, 0.3), ({c, d}, 0.3), ({e}, 0.1) }, $\overline{\mathcal{F}}$ = { ({a, b, c}, 0.4), ({a, b, d}, 0.3), ({a, c, d}, 0.2), ({c, d, e}, 0.1) }; these two random sets were also considered in (Dubois and Prade 1986). Table 5.1 gives the calculated Belief values for all possible subsets of $S$; by using Eq. (3.7) to calculate the relevant Plausibility values, they show that the right-hand-side of Eq. (5.1) is satisfied. Let us now check the left-hand-side of Eq. (5.1).

In this example, $|S|=5$. Since $A^4$ = {e} and {e} is not in any other focal set, the projection of $\Psi$ onto the 5-th dimension degenerates to a point (the upper and lower probabilities of singleton "e" are equal to 0.1). Thus, the imprecise possibility space $S_{imp}$ is given by the first 4 elements of $S$, and the projected imprecise probability space is given by the projection onto 3 of the first 4 dimensions. By using Algorithm 3.2.3.5, one finds 8 distinct points in $\Psi$, which are the extreme distributions, i.e. the vertices of the convex hull of $\Psi$ (listed in Table 5.2).

As for $\overline{\Psi}$, $S = S_{imp}$ because all focal sets contain more than one singleton, and thus $p$= 4. Algorithm 3.2.3.5 finds 22 distinct points in $\overline{\Psi}$, which are vertices of its convex hull, i.e. extreme distributions (listed on the left-hand-side of Table 5.3). The right-hand-side of Table 5.3 gives the vertices of the convex hull obtained by projecting onto the first 3 dimensions: notice that only 12 vertices were obtained because several vertices overlapped on the 3-D projection.

Vertices in the left-hand-side of Table 5.3 were projected onto the first 4 dimensions, and the first inclusion algorithm above yielded the results in Table 5.4. As expected, $\Psi \subset \overline{\Psi}$ because all vertices of $\Psi$ are a linear combination of vertices in $\overline{\Psi}$ (given in the third column of Table 5.4) with coefficients given in the right-most column of Table 5.4. Notice that:

- Of the 26,334 possible combinations $I_k$, very few (from 13 to 691) were used before finding a non-negative solution to Eq. (5.3). This confirms the comment that the practical cost of the algorithm is much less than the worst case scenario.
- Some of these used combinations yielded a singular system (5.3). Indeed, about one third of the first 300 combinations were singular. Once detected as singular, these systems are not solved, and are not considered in the next cycle through $v_i$. This further reduces the computational time.
- The algorithm correctly finds that $v_2 \equiv \overline{v}_8$ and $v_4 \equiv \overline{v}_{17}$ (rows 3 and 5 in Table 5.4).

Figure 5.3a and b show the projection of $\Psi$ and $\overline{\Psi}$, respectively, onto the first three dimensions: $P_i = P(\{i\})$, $i = a, b, c$. Notice that $\overline{\Psi}$ has many more faces than $\Psi$ because $\overline{\Psi}$ has 22 vertices, whereas $\Psi$ is defined by 8 vertices. This is due to the fact that $\overline{\Psi}$ is not degenerate and that focal sets are composed of 3 elements instead of 2, which makes matrix **B**, and thus vector $\mathbf{p}_j$, less sparse in Algorithm 3.2.3.5.

**Table 5.1** Example 5.1: Belief values for $\mathcal{F}$ and $\overline{\mathcal{F}}$

| Events | $Bel(T)$ | $Bel(T)$ |
|---|---|---|
| $\{a\}$ | 0 | 0 |
| $\{b\}$ | 0 | 0 |
| $\{c\}$ | 0 | 0 |
| $\{d\}$ | 0 | 0 |
| $\{e\}$ | 0.1 | 0. |
| $\{a, b\}$ | 0.3 | 0 |
| $\{a, c\}$ | 0.3 | 0 |
| $\{a, d\}$ | 0 | 0 |
| $\{a, e\}$ | 0.1 | 0 |
| $\{b, c\}$ | 0 | 0 |
| $\{b, d\}$ | 0. | 0 |
| $\{b, e\}$ | 0.1 | 0 |
| $\{c, d\}$ | 0.3 | 0 |
| $\{c, e\}$ | 0.1 | 0 |
| $\{d, e\}$ | 0.1 | 0 |
| $\{a, b, c\}$ | 0.6 | 0.4 |
| $\{a, b, d\}$ | 0.3 | 0.3 |
| $\{a, b, e\}$ | 0.4 | 0 |
| $\{a, c, d\}$ | 0.6 | 0.2 |
| $\{a, c, e\}$ | 0.4 | 0 |
| $\{a, d, e\}$ | 0.1 | 0 |
| $\{b, c, d\}$ | 0.3 | 0 |
| $\{b, c, e\}$ | 0.1 | 0 |
| $\{b, d, e\}$ | 0.1 | 0. |
| $\{c, d, e\}$ | 0.4 | 0.1 |
| $\{a, b, c, d\}$ | 0.9 | 0.9 |
| $\{a, b, c, e\}$ | 0.7 | 0.4 |
| $\{a, b, d, e\}$ | 0.4 | 0.3 |
| $\{a, c, d, e\}$ | 0.7 | 0.3 |
| $\{b, c, d, e\}$ | 0.4 | 0.1 |

**Table 5.2** Example 5.1: Convex hull vertices for polytope $\Psi$

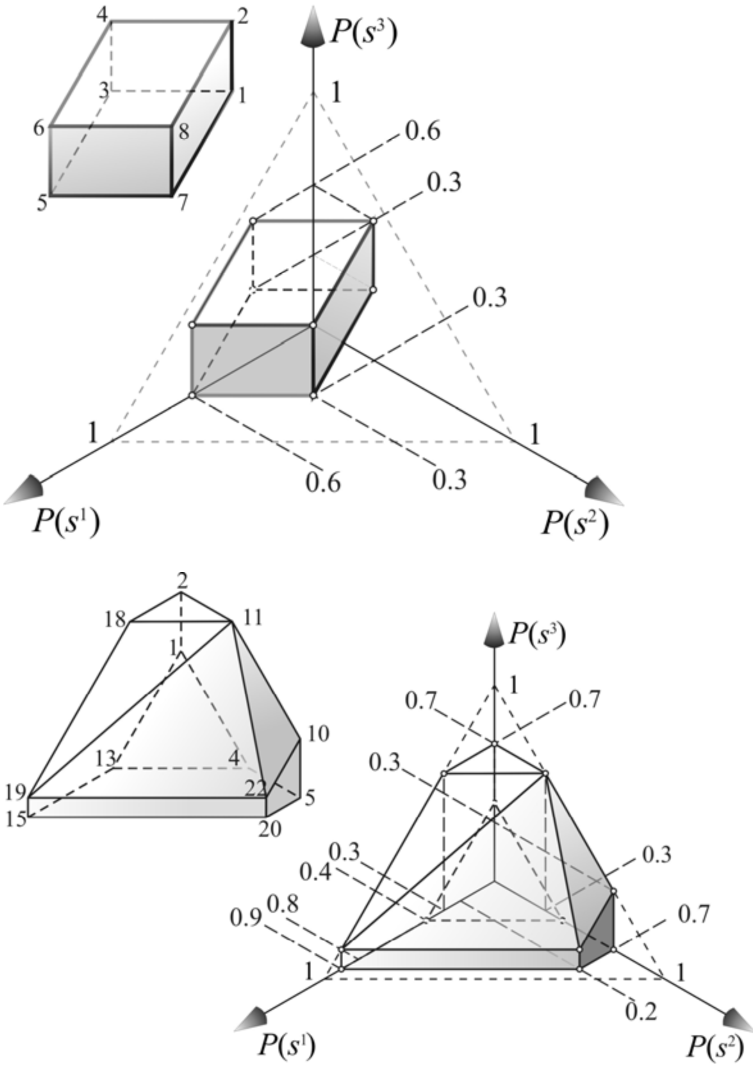
| No. | Perm. | Distribution | No. | Perm. | Distribution |
|---|---|---|---|---|---|
| 1 | 1 | {0., 0.3, 0.3, 0.3, 0.1} | 5 | 9 | {0.6, 0., 0., 0.3, 0.1} |
| 2 | 2 | {0., 0.3, 0.6, 0., 0.1} | 6 | 12 | {0.6, 0., 0.3, 0., 0.1} |
| 3 | 7 | {0.3, 0., 0.3, 0.3, 0.1} | 7 | 13 | {0.3, 0.3, 0., 0.3, 0.1} |
| 4 | 8 | {0.3, 0., 0.6, 0., 0.1} | 8 | 23 | {0.3, 0.3, 0.3, 0., 0.1} |

**Table 5.3** Example 5.1: Convex hull vertices for polytope $\overline{\Psi}$

| Vertices found by projecting on 4 dimensions | | | Vertices found by projecting on 3 dimensions | | |
|---|---|---|---|---|---|
| No. | Perm. | Distribution | No. | Perm. | Distribution |
| 1 | 1 | {0., 0., 0.4, 0.5, 0.1} | 1 | 1 | {0., 0., 0.4, 0.5, 0.1} |
| 2 | 4 | {0., 0., 0.7, 0.3, 0.} | 2 | 4 | {0., 0., 0.7, 0.3, 0.} |
| 3 | 5 | {0., 0., 0.4, 0.6, 0.} | 3 | 8 | {0., 0.4, 0., 0.6, 0.} |
| 4 | 7 | {0., 0.4, 0., 0.5, 0.1} | 4 | 9 | {0., 0.7, 0., 0.2, 0.1} |
| 5 | 9 | {0., 0.7, 0., 0.2, 0.1} | 5 | 14 | {0., 0.3, 0.7, 0., 0.} |
| 6 | 11 | {0., 0.4, 0., 0.6, 0.} | 6 | 18 | {0., 0.7, 0.3, 0., 0.} |
| 7 | 12 | {0., 0.7, 0., 0.3, 0.} | 7 | 31 | {0.4, 0., 0., 0.5, 0.1} |
| 8 | 13 | {0., 0.3, 0.6, 0., 0.1} | 8 | 33 | {0.9, 0., 0., 0., 0.1} |
| 9 | 16 | {0., 0.7, 0.2, 0., 0.1} | 9 | 38 | {0.3, 0., 0.7, 0., 0.} |
| 10 | 18 | {0., 0.7, 0.3, 0., 0.} | 10 | 48 | {0.9, 0., 0.1, 0., 0.} |
| 11 | 23 | {0., 0.3, 0.7, 0., 0.} | 11 | 61 | {0.2, 0.7, 0., 0., 0.1} |
| 12 | 27 | {0., 0., 0.6, 0.3, 0.1} | 12 | 95 | {0.2, 0.7, 0.1, 0., 0.} |
| 13 | 31 | {0.4, 0., 0., 0.5, 0.1} | | | |
| 14 | 32 | {0.4, 0., 0., 0.6, 0.} | | | |
| 15 | 33 | {0.9, 0., 0., 0., 0.1} | | | |
| 16 | 36 | {0.9, 0., 0., 0.1, 0.} | | | |
| 17 | 37 | {0.3, 0., 0.6, 0., 0.1} | | | |
| 18 | 47 | {0.3, 0., 0.7, 0., 0.} | | | |
| 19 | 48 | {0.9, 0., 0.1, 0., 0.} | | | |
| 20 | 61 | {0.2, 0.7, 0., 0., 0.1} | | | |
| 21 | 71 | {0.2, 0.7, 0., 0.1, 0.} | | | |
| 22 | 95 | {0.2, 0.7, 0.1, 0., 0.} | | | |

**Table 5.4** Example 5.1: Results of first inclusion algorithm for $\Psi$ and $\overline{\Psi}$

| $v_i$ | No. of singular systems (5.3) | Combination number, $k$ | Combination of vertices $\overline{v}_i$, $I_k$ | Convex Combination Coefficients $\{\lambda_{I_{k,1}},\ldots,\lambda_{I_{k,p+1}}\}$ |
|---|---|---|---|---|
| 1 | 102 | 224 | {1, 2, 4, 8, 13} | {0.15, 0, 0.45, 0.4, 0} |
| 2 | 30 | 71 | {1, 2, 3, 8, 13} | {0, 0, 0, 1., 0} |
| 3 | 108 | 272 | {1, 2, 4, 12, 15} | {0.5, 0, 0, 0.166667, 0.333333} |
| 4 | 8 | 13 | {1, 2, 3, 4, 17} | {0, 0, 0, 0, 1.} |
| 5 | 108 | 281 | {1, 2, 4, 13, 15} | {0, 0, 0, 0.6, 0.4} |
| 6 | 108 | 298 | {1, 2, 4, 15, 17} | {0, 0, 0, 0.5, 0.5} |
| 7 | 77 | 181 | {1, 2, 4, 5, 15} | {0, 0, 0.555556, 0.111111, 0.333333} |
| 8 | 171 | 691 | {1, 2, 8, 9, 15} | {0, 0, 0.416667, 0.25, 0.333333} |




a)

b)

**Figure 5.3** Example 5.1: Convex hulls projected onto the first 3 coordinate space: (a) $\Psi$ ; (b) $\overline{\Psi}$ . $s_1$ = a, $s_2$ = b, $s_3$ = c. Vertex numbering in (b) refers to the numbering in the left-hand side of Table 5.3

#### 5.1.1.2 Second Weak Inclusion Algorithm

If the full convex hull (including facets and counterclockwise, from outside, *triangulation* of *non-simplicial* facets (see (O'Rourke 1998)) is known for polytope $\overline{\Psi}$, more efficient algorithms exist in computational geometry.

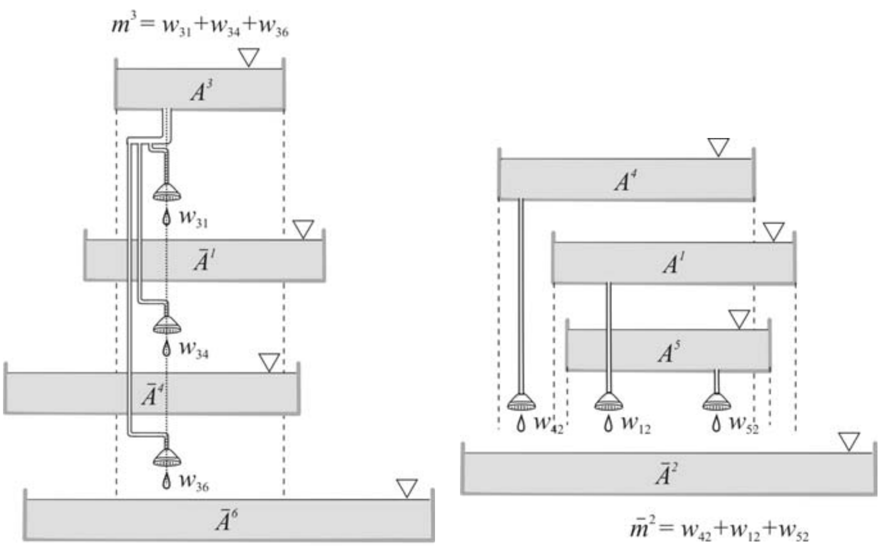
These algorithms are typically sub-algorithms to convex hull algorithms (e.g., incremental algorithm (O'Rourke 1998)). In the example of Figure 5.1b, $v_i \in \overline{\Psi}$ if and only if $v_i$ is to the left of each simplex making up the boundary of $\overline{\Psi}$ (in this case directed edges $\left(\overline{v}_k, \overline{v}_l\right)$, $k< l = k+1$). This condition is equivalent to requiring that the signs of the determinants $\left|\overline{\xi}_k\ \overline{\xi}_l\ \xi_i\right|$ must all be positive, where $\overline{\xi}$ ($\xi$, resp.) is the column vector of $\overline{v}$ ($v$, resp.) in the projected imprecise probability space $\tilde{S}$ with 1 appended as last coordinate. Likewise, in the $p$-dimensional case, the signs of the determinants $\left|\overline{\xi}_k \ldots \overline{\xi}_l\ \xi_i\right|$ must all be positive, where $\left(\overline{v}_k,\ldots,\overline{v}_l\right)$ is the counterclockwise list of vertices for a simplicial facet, and where 1 has been appended to the column vectors of the vertices' coordinates. Very efficient algorithms have been devised for calculating the sign of a determinant e.g., (Clarkson 1992).

### *5.1.2 Strong Inclusion*

The weak-inclusion definition is awkward to use because it is not defined in terms of the probability assignment. Additionally, finding an including random set is not trivial, especially if additional constraints, e.g., consonance, must be satisfied because the definition of weak inclusion is not constructive. A stronger, more versatile notion of inclusion may easily be introduced by using the reservoir-bathtub analogy (Example 3-4).

Recall that, in this analogy, water can just flow downwards through vertical pipes. Think of the focal sets in $\overline{\mathcal{F}}$ as a set of reservoirs underlying the focal sets in $\mathcal{F}$. The basic idea is that if $\overline{\mathcal{F}}$ contains $\mathcal{F}$, then each focal set $A^i$ should discharge into at least one focal set $\overline{A}^j$ regardless of where the pipe is located in $A^i$, i.e. there must be at least one $\overline{A}^j$ such that $A^i \subseteq \overline{A}^j$. Call $w_{ij}$ the flow rate from $A^i$ to $\overline{A}^j$. Since the flow can only be downward, $w_{ij} \geq 0$, and since $A^i$ discharges only into reservoirs that contain $A^i$, if $A^i$ is not contained in $\overline{A}^j$, then $w_{ij} = 0$. As exemplified in Figure 5.4a, by conservation of mass, the outgoing flow rate from $A^i$, $m\left(A^i\right)$, must be equal to the sum of the flow rates from $A^i$ into reservoirs $\overline{A}^j$ that contain $A^i$, i.e., $m\left(A^i\right) = \sum_{j:A^i \subseteq \overline{A}^j} w_{ij}$.

On the other hand, if we focus our attention on reservoir $\bar{A}^j$ as in Figure 5.4b, $\bar{A}^j$ must receive water from all reservoirs $A^i$ contained in it. If no reservoir $A^i$ is contained in it, then $\bar{A}^j$ is dry, and it cannot be considered as a reservoir (focal set). Therefore, there must be at least one $A^i$ such that $A^i \subseteq \bar{A}^j$. Again, by mass conservation, the outgoing flow rate from $\bar{A}^j$, $\bar{m}\left(\bar{A}^j\right)$, must be equal to the flow rates from all reservoirs $A^i$ contained in it, i.e., $\bar{m}\left(\bar{A}^j\right) = \sum_{i:A^i \subseteq \bar{A}^j} w_{ij}$ .




(a) (b)

**Fig. 5.4** Example of strong random set inclusion: (a) water from $A^3$ flows only into reservoirs $\bar{A}^j$ that contain $A^3$; (b) water flowing into $\bar{A}^2$ only comes from reservoirs $A^i$ that are contained in $\bar{A}^2$

We can summarize the above discussion in the following definition of inclusion (in order not to burden the notation, from now on we will drop the adjective "strong") (Dubois and Prade 1986; Yager 1986; Delgado and

Moral 1987; Dubois and Prade 1991): $\mathcal{F} \subseteq \overline{\mathcal{F}}$ if and only if the three following conditions hold:

$$\begin{aligned}
&(i)\ \forall A^i \quad \exists\, \overline{A}^j : A^i \subseteq \overline{A}^j \\
&(ii)\ \forall \overline{A}^j \quad \exists\, A^i : A^i \subseteq \overline{A}^j \\
&(iii)\ \exists\ n \times \overline{n}\ matrix\ \mathbf{w} : w_{ij} \geq 0 \text{ and} \\
&\qquad \forall A^i,\ m(A^i) = \sum_{j:A^i \subseteq \overline{A}^j} w_{ij} \\
&\qquad \forall \overline{A}^j,\ \overline{m}(\overline{A}^j) = \sum_{i:A^i \subseteq \overline{A}^j} w_{ij} \\
&\qquad A^i \not\subset \overline{A}^j \Rightarrow w_{ij} = 0
\end{aligned} \tag{5.4}$$

In words, the weights $m(A^i)$ can only be shared among the supersets of $A^i$ that are focal sets in $\overline{\mathcal{F}}$, and the weight $\overline{m}(\overline{A}^j)$ is the sum of the shares allocated to the focal sets in $\mathcal{F}$ that are subsets of $\overline{A}^j$. Zadeh's definition of inclusion for fuzzy sets is a special case of Eq. (5.4) and (5.1) because if $F_1$ and $F_2$ are fuzzy sets equivalent to consonant random set $\mathcal{F}$ and $\overline{\mathcal{F}}$, respectively, then

$$\mathcal{F} \subseteq \overline{\mathcal{F}} \text{ and } \Psi \subseteq \overline{\Psi} \text{ if and only if } \mu_{F_1} \leq \mu_{F_2} \tag{5.5}$$

**Example 5.2.** Let $S = \mathbb{R}$. Consider random sets $\mathcal{F}$ : (([0.5, 0.6], 0.5), ([0.3, 0.9], 0.5)); and $\overline{\mathcal{F}}$ : (([0.3, 0.7], 0.1), ([0.4, 0.9], 0.1), ([0.3, 1], 0.8)). Focal set inclusions and a matrix **w** are given in Table 5.5; notice that $w_{21} = w_{22} = 0$ because $A^2$ is only a subset of $\overline{A}^3$. Figure 5.5 shows the relevant reservoir-bathtub analogy: notice that no pipes connect $A^2$ to either $\overline{A}^1$ or $\overline{A}^2$.

**Table 5.5** Example 5.2: Matrix **w** (entries are not italicized) and probability assignments (italicized)

| | $\overline{A}^1$ | $\overline{A}^2$ | $\overline{A}^3$ | Total $m(A^i)$ |
|---|---|---|---|---|
| $A^1 \subseteq (\overline{A}^1, \overline{A}^2, \overline{A}^3)$ | 0.1 | 0.1 | 0.3 | *0.5* |
| $A^2 \subseteq \overline{A}^3$ | 0.0 | 0.0 | 0.5 | *0.5* |
| Total $\overline{m}\left(\overline{A}^i\right)$ | *0.1* | *0.1* | *0.8* | ***1*** |

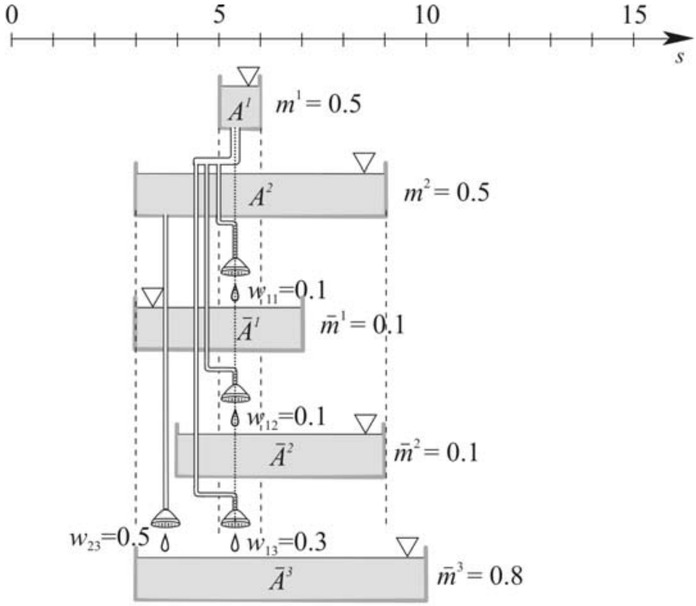



**Fig. 5.5** Example 5.2: Reservoir-bathtub analogy

One may now wonder what the relationship is between weak inclusion and inclusion. Consider again the example in Figure 5.4a, which shows the reservoirs $\bar{A}^j$ that contain reservoir $A^3$, and the relevant flow rates $w_{3j}>0$. With reference to Figure 5.6a, the maximum flow rate provided by $A^3$ into bathtub $T_1$ is equal to $m(A^3)$. The reservoirs $\bar{A}^j$ that contain reservoir $A^3$ also provide a flow rate at least equal to $m(A^3)$ because they may also receive flow rate from other reservoirs $A^i$, $i\neq3$, and all water flow rates $w_{3j}>0$ can be diverted into $T_1$.

On the other hand, since bathtub $T_2$ does not intersect the footprint of $A^3$, it receives no water flow from $A^3$. However, $T_2$ intersects the footprint of $\bar{A}^6$, which contains $A^3$. Thus, the maximum flow rate from the reservoirs $\bar{A}^j$ that contain $A^3$ into $T_2$ is at least $m(A^3)$ because all water flow rates $w_{3j}>0$ can be diverted into $T_2$ through $\bar{A}^j$. By repeating this reasoning for all reservoirs $A^i$, one obtains that the maximum flow rate into any bathtub in $\mathcal{F}$ is always lower than or equal to the maximum flow rate provided by $\bar{\mathcal{F}}$, i.e. $Pla(T) \leq \overline{Pla}(T)\ \forall T$.

As for the minimum flow rate into a bathtub, consider again the example in Figure 5.4b, which shows the reservoirs $A^i$ contained in reservoir $\overline{A}^2$, and the relevant flow rates $w_{i2}>0$. Figure 5.6b shows that any bathtub $T$ (whose footprint is in $\overline{A}^2$) may receive no water from $\overline{A}^2$, and thus flow rates $w_{i2}>0$ may not go into bathtub $T$. Additionally, reservoirs $A^i$ contained in reservoir $\overline{A}^2$ may give flow rate to other reservoirs $\overline{A}^j$, $j\neq 2$. However, $T$ has to receive water from all reservoirs $A^i$ contained in reservoir $\overline{A}^2$ because they are contained in $T$. By repeating this reasoning for all reservoirs $\overline{A}^j$, one obtains that the minimum flow rate into any bathtub $T$ in $\overline{\mathcal{F}}$ is always lower than or equal to the minimum flow rate provided by $\mathcal{F}$, i.e. $\overline{Bel}(T) \leq Bel(T)\ \forall T$. As a consequence, (strong) inclusion always implies weak inclusion; however, the reverse is not true.

In order to understand why the reverse is not true, consider a set, $\mathcal{U}$, of reservoirs $A^i$. Let $\overline{\mathcal{U}}$ be the set of reservoirs $\overline{A}^j$ onto which all $A^i \in \mathcal{U}$ discharge, i.e. $\overline{\mathcal{U}} = \left\{\overline{A}^j : w_{ij} > 0, A^i \in \mathcal{U}\right\}$. In the example of Figure 5.6a, let us assume that $A^3$ is in $\mathcal{U}$, then $\overline{A}^1$, $\overline{A}^4$, and $\overline{A}^6$ are in $\overline{\mathcal{U}}$. Recall that the reservoirs $\overline{A}^j$ that contain reservoir $A^3$ may also receive flow rate from other reservoirs $A^i$, $i\neq 3$. As a consequence:

$$\sum_{i:A^i\in\mathcal{U}} m\left(A^i\right) \leq \sum_{j:\overline{A}^j\in\overline{\mathcal{U}}} m\left(\overline{A}^j\right)\quad \forall\mathcal{U} \tag{5.6}$$

On the other hand, consider a set, $\overline{\mathcal{Z}}$, of reservoirs $\overline{A}^j$. Let $\mathcal{Z}$ be the set of reservoirs $A^i$ that discharge into $\overline{A}^j \in \overline{\mathcal{Z}}$, i.e. $\mathcal{Z} = \left\{A^i : w_{ij} > 0, \overline{A}^j \in \overline{\mathcal{Z}}\right\}$. In the example of Figure 5.6b, let us assume that $\overline{A}^2$ is in $\overline{\mathcal{Z}}$, then $A^1$, $A^4$, and $A^5$ are in $\mathcal{Z}$. Recall that the reservoirs $A^i$ contained in reservoir $\overline{A}^2$ may discharge into other reservoirs $\overline{A}^j \notin \overline{\mathcal{Z}}$. As a consequence:

$$\sum_{j:\overline{A}^j\in\overline{\mathcal{Z}}} m\left(\overline{A}^j\right) \leq \sum_{i:A^i\in\mathcal{Z}} m\left(A^i\right)\quad \forall\overline{\mathcal{Z}} \tag{5.7}$$

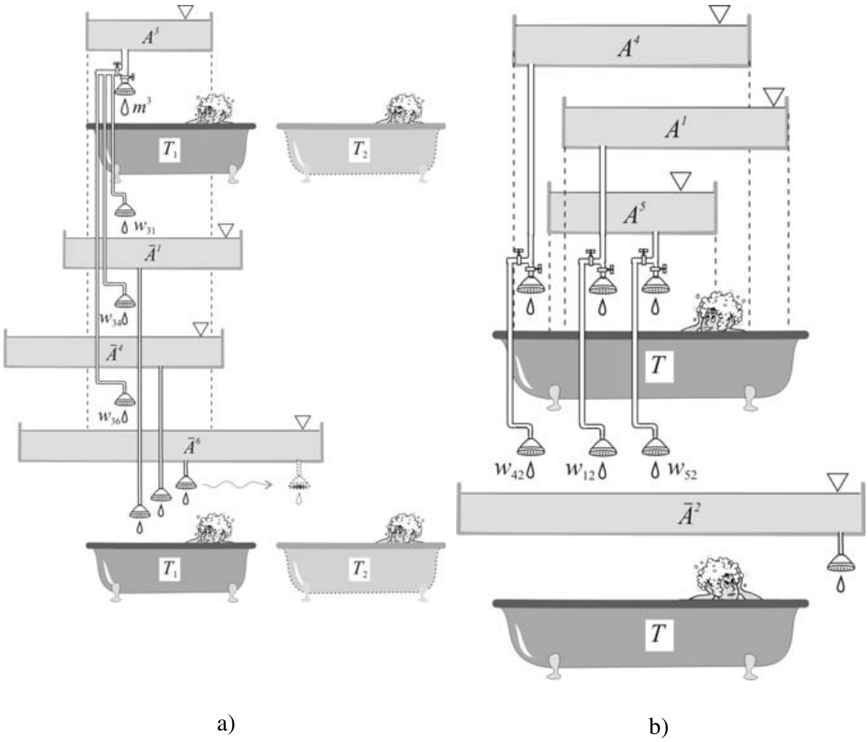



**Fig. 5.6** Relationship between weak inclusion and (strong) inclusion: a) plausible flow rate for bathtubs $T_1$ and $T_2$; b) believed flow rate for bathtub $T$

Eqs. (5.6) and (5.7) are necessary and sufficient for mass conservation. Notice that Eq. (5.7) for $\bar{\mathcal{Z}} = \left\{\bar{A}^j\right\}$ (a single reservoir) gives $\overline{Bel}\left(\bar{A}^j\right) \le Bel\left(\bar{A}^j\right)$ and that, for every bathtub $T$, there is a $\bar{\mathcal{Z}}$ such that

$$\overline{Bel}(T) = \sum_{j:\bar{A}^j \in \bar{\mathcal{Z}}} m\left(\bar{A}^j\right) \tag{5.8}$$

This entails that, for every bathtub $T$, the minimum and maximum flows are constrained as in Eq. (5.1). However, mass conservation is much more stringent, and requires Eqs. (5.6) and (5.7) to be simultaneously satisfied for all combinations of reservoirs. Therefore, even though Eq. (5.1) is satisfied, in general, one may find a combination of reservoirs $\bar{\mathcal{Z}}$ such that there is no bathtub $T$ whose minimum inward flow rate is given by Eq. (5.8). This is why weak inclusion does not imply (strong) inclusion.

Let us now formally summarize the results obtained so far in a theorem.

**Theorem 5.1.** Let $\mathcal{F}$ and $\overline{\mathcal{F}}$ be two random sets or relations defined on $S$ with compatible sets of probability measures $\Psi$ and $\overline{\Psi}$, respectively. Then:

$$\begin{aligned}
&(i)\ \Psi \subseteq \overline{\Psi} \Leftrightarrow \left[Bel(T), Pl(T)\right] \subseteq \left[\overline{Bel}(T), \overline{Pl}(T)\right]\ \forall T \subseteq S \\
&(ii)\ \mathcal{F} \subseteq \overline{\mathcal{F}} \Rightarrow \left[Bel(T), Pl(T)\right] \subseteq \left[\overline{Bel}(T), \overline{Pl}(T)\right]\ \forall T \subseteq S \\
&(iii)\ \mathcal{F} \subseteq \overline{\mathcal{F}} \Rightarrow \Psi \subseteq \overline{\Psi} \\
&(iv)\ \left[Bel(T), Pl(T)\right] \not\subset \left[\overline{Bel}(T), \overline{Pl}(T)\right] \Rightarrow \mathcal{F} \not\subset \overline{\mathcal{F}}
\end{aligned} \tag{5.9}$$

but the reverse of (*ii*) and (*iii*) is not necessarily true, i.e.:

$$(v)\left[Bel(T), Pla(T)\right] \subseteq \left[\overline{Bel}(T), \overline{Pla}(T)\right] \forall T \subseteq S \not\Rightarrow \mathcal{F} \subseteq \overline{\mathcal{F}} \tag{5.10}$$

*Proof.* (*i*) Immediate from the definitions (3.10) and (3.11) on page 34.

(*ii*) (Dubois and Prade 1986) Let us use Eq. ((5.4) (*iii*)) within the definition (3.3) of Plausibility for $T \subseteq S$:

$$\overline{Pla}(T) = \sum_{T \cap \overline{A}^j \neq \varnothing} \overline{m}\left(\overline{A}^j\right) = \sum_{T \cap \overline{A}^j \neq \varnothing} \sum_{i: A^i \subseteq \overline{A}^j} w_{ij} = \sum_{i,j} \left\{ w_{ij} : A^i \subseteq \overline{A}^j, T \cap \overline{A}^j \neq \varnothing \right\} \tag{5.11}$$

By Eq. (5.4) (*ii*)):

$$\left\{ (i,j) : A^i \subseteq \overline{A}^j, T \cap \overline{A}^j \neq \varnothing \right\} \supseteq \left\{ (i,j) : A^i \subseteq \overline{A}^j, T \cap A^i \neq \varnothing \right\} \tag{5.12}$$

Hence:

$$\overline{Pla}(T) \geq \sum_{i,j} \left\{ w_{ij} : A^i \subseteq \overline{A}^j, T \cap A^i \neq \varnothing \right\} = \sum_{T \cap A^i \neq \varnothing} m\left(A^i\right) = Pla(T) \tag{5.13}$$

Likewise, one can show that $\overline{Bel}(T) \leq Bel(T)$ by using Eq. (3.7).

(*iii*) Immediate from (*ii*) using (*i*). However, it is a good exercise to prove it independently. Let $P^{A_i} \in \Psi^i$. From ((5.4) (*ii*) and (*iii*)), for any $\overline{A}^j$, one can form a probability measure on $\overline{A}^j$ by defining:

$$P*^{\overline{A}_j} := \left( \sum_{i: A^i \subseteq \overline{A}_j} w_{ij} P^{A_i} \right) / \overline{m}\left(\overline{A}_j\right) \in \overline{\Psi}^j \tag{5.14}$$

Let $\overline{\Psi}*^j := \left\{ P*^{\overline{A}_j} : P^{A_i} \in \Psi^i \right\}$. Since $\bigcup_{i:A^i \subseteq \overline{A}_j} A^i \subseteq \overline{A}_j$, $\overline{\Psi}*^j \subseteq \overline{\Psi}^j$, and hence $\Psi \subseteq \overline{\Psi}$.

(*iv*) Immediate from (*ii*)

(*v*) (Dubois and Prade 1986). Consider the following counterexample. As in Example 5.1, let $S = \{a, b, c, d, e\}$, and consider the two random sets: $\left\{\left(A^i, m^i\right)\right\}$ = { ({a, b}, 0.3), ({a, c}, 0.3), ({c, d}, 0.3), ({e}, 0.1) }, $\left\{\left(\overline{A}^j, \overline{m}^j\right)\right\}$ = { ({a, b, c}, 0.4), ({a, b, d}, 0.3), ({a, c, d}, 0.2), ({c, d, e}, 0.1) }. Table 5.1 gives the calculated Belief values for all possible subsets of $S$; by using Eq. (3.7) to calculate the relevant Plausibility values, they show that the first inclusion in Eq. (5.10) is satisfied. Eq. ((5.4) (*iii*)) yields the following system of 8 linear equations in 7 of the entries of matrix **w**:

$$\begin{cases} w_{11} + w_{12} = 0.3 \\ w_{21} + w_{23} = 0.3 \\ w_{33} + w_{34} = 0.3 \\ w_{44} = 0.1 \\ w_{11} + w_{21} = 0.4 \\ w_{12} = 0.3 \\ w_{23} + w_{33} = 0.2 \\ w_{34} + w_{44} = 0.1 \end{cases} \tag{5.15}$$

to be solved in [0, 1] with the additional equation $\sum_{i,j} w_{ij} = 1$. Eqs. (5.15) are inconsistent because they yield the following solutions: (8)→$w_{34}$ = 0; (3)→$w_{33}$ = 0.3; (1)→$w_{11}$ = 0.0; (5)→$w_{21}$ = 0.4; (2)→$w_{23}$ = -0.1; (7)→$w_{23}$ = 0.0. Figure 5.7 shows the reservoir-bathtub analogy; notice that conservation of mass (5.7) is violated when taking, for example, $\overline{\mathcal{Z}} = \left\{\overline{A}^1, \overline{A}^2\right\}$ because $\overline{m}\left(\overline{A}^1\right) + \overline{m}\left(\overline{A}^2\right) = 0.7 > m\left(A^1\right) + m\left(A^2\right) = 0.6$. As a consequence, the mass balance equation for reservoir $A^2$ (second equation in (5.15)) would give a negative flow rate of -0.1 from $A^2$ to $\overline{A}^3$ (which violates our definition of flow direction). This result is incompatible with mass equation for $\overline{A}^3$ (seventh equation in (5.15)), which requires a zero flow rate from $A^2$ to $\overline{A}^3$. ■

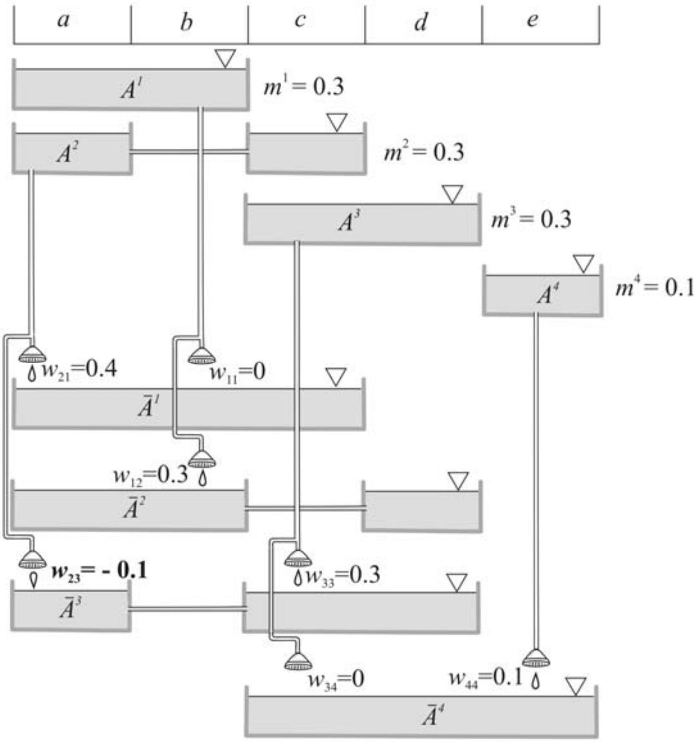



**Fig. 5.7** Reservoir representation for the counterexample in Theorem 5.1. (*v*)

Let $\mathcal{F}$ be the random set corresponding to the actual data. If random set $\overline{\mathcal{F}}$, such that $\overline{\mathcal{F}} \supseteq \mathcal{F}$, is easier to elicit or to compute with, then it is possible to calculate (on the safe side) upper and lower bounds on the probability of every subset (event) in the same space of interest. In particular, if $\overline{\mathcal{F}}$ is consonant (i.e. a fuzzy set or relation with membership function $\mu(x)$), then the probability interval for set $T$ is very easy to compute by making use of Eq. (3.25) (for infinite sets, the "sup" operator should be used)

$$\left[\overline{Bel}(T), \overline{Pl}(T)\right] = \left[\overline{Nec}(T), \overline{Pos}(T)\right] = \left[1 - \max_{x \in T^c} \mu(x), \max_{x \in T} \mu(x)\right] \; \forall T \qquad (5.16)$$

The remainder of this section will introduce algorithms to include a random set in a consonant random set, and investigate inclusion and monotonicity relationships under the hypothesis of random set independence and non-interactivity. Section 5.2.2 will explore how monotonicity properties are preserved through mappings.

### 5.1.3 Including a Random Set in a Consonant Random Set

Consider a variable, *s*, constrained by a non-consonant random set defined on a completely ordered set $S$ (e.g., $S = \mathbb{R}$, so that focal sets are real intervals). At present, there seems to be no preferred algorithm capable of performing this inclusion without any loss of information (Joslyn and Klir 1992; Klir 1995). We may distinguish four cases:

a) The minimum values of *s* yield worst-case scenarios in design situations: for example this is often the case for resistances, rock mass classifications (RMR, Q, or GSI), cohesion, and the friction angle of a fracture or a soil. In this case, $F_{UPP}$ should be preserved.
b) The maximum values of *s* are of interest: for example, the water pressure in the ground, and load effects. In this case, $F_{LOW}$ should be preserved.
c) No indication is available on which values of *s* are crucial, and equal weight should be given to preserving $F_{UPP}$ and $F_{LOW}$.
d) No indication is available on which values of *s* are crucial, and one wants to minimize the cardinality of the consonant inclusion (applicable if $S$ is finite).

In order not to be too conservative, in the first (second) case we will try to minimize the difference between the information relative to the left (right) extremes of the focal sets.

#### 5.1.3.1 Case (a)

Let $A^i = [l^i, u^i]$. The following is the procedure proposed in (Tonon, Bernardini et al. 2000):

1) Intervals $A^i$ $i$=1,…, $n$, are ordered starting from the interval whose left extreme is maximum:

$$A^1 : l^1 = \max_i \{l^i\} \tag{5.17}$$

$$A^k : l^k = \max_{i \geq k} \{l^i\} \quad k = 2,\ldots,n \tag{5.18}$$

   If two or more intervals are encountered that have the same value of the left extreme, then they are ordered starting from the interval having the minimum value of the right extreme; this allows the loss of information regarding the right extremes to be minimized as well, whenever possible.
2) Set:

$$\bar{A}^1 = \left[\bar{l}^1, \bar{u}^1\right] := A^1 \tag{5.19}$$

3) For each interval $A^i$ with $i \geq 2$, check if:

$$u^i \geq \overline{u}^{i-1} \quad i \geq 2 \tag{5.20}$$

- IF Eq. (5.20) is true, THEN set $\overline{u}^i := u^i \quad i \geq 2$, case (1) in Figure 5.8.
- ELSE, set $\overline{u}^i := u^{i-1} \quad i \geq 2$, case (2) in Figure 5.8.

In this way, the focal elements of $\overline{\mathcal{F}}$ are all nested. Now it is straightforward to construct the matrix **w** introduced in Eq. (5.4)).

4) For each interval $A^i$, $i$=1,...,$n$ let $k_i$ be the number of intervals $\overline{A}^j$ that include $A^i$, i.e. $\overline{A}^j : \overline{A}^j \supseteq A^i$; it turns out that $k_i = n\text{-}i$.
5) Set :

$$w_{ii} = m\left(A^i\right) - k_i \cdot \beta = m\left(A^i\right) - \left(n-i\right) \cdot \beta \tag{5.21}$$

$$w_{ij} = \begin{cases} \beta & if \quad \overline{A}^j \supset A^i \quad i.e., if \ i < j \\ 0 & if \quad \overline{A}^j \subseteq A^i \ i.e., if \ i > j \end{cases} \tag{5.22}$$

where β is a non-negative real number. The smaller β, the better the approximation of $\overline{\mathcal{F}}$ to $\mathcal{F}$, with best approximation achieved when β = 0.

Notice that **w** so defined is upper tridiagonal when β > 0, and diagonal when β = 0.

6) The basic probability assignment $\overline{m}$ is calculated as:

$$\overline{m}\left(\overline{A}^j\right) = \sum_{i=1}^{n} w_{ij} \tag{5.23}$$

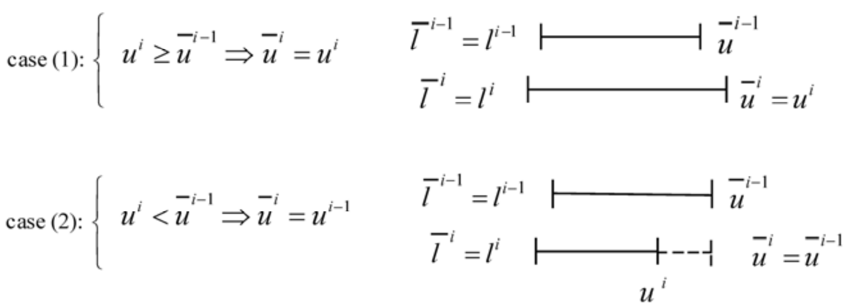



**Fig. 5.8** Two different cases in the inclusion of a non-consonant random set in a consonant random set

**Example 5.3.** Let $S = \mathbb{R}$. Consider random set $\mathcal{F}$ : {([3, 7], 0.5), ([5, 8], 0.2), ([2, 4], 0.3)} depicted in Figure 5.9a together with its upper/lower CDFs. Let us re-order the focal elements per Eqs. (5.17) and (5.18) so that the random set is {([5, 8], 0.2), ([3, 7], 0.5), ([2, 4], 0.3)}. After applying Step 3, the focal sets are {[5, 8], [3, 8], [2, 8]}. Table 5.6 gives matrix **w**. Figure 5.9b shows the upper/lower CDFs of the including random set $\overline{\mathcal{F}}$ ; when β = 0, the upper CDF coincides with the upper CDF of $\mathcal{F}$ .

**Table 5.6** Example 5.3: Matrix **w** (entries are not italicized) and probability assignments (italicized)

| | $\overline{A}^1$ | $\overline{A}^2$ | $\overline{A}^3$ | Total $m(A^i)$ |
|---|---|---|---|---|
| $A^1 \subseteq (\overline{A}^1, \overline{A}^2, \overline{A}^3)$ | *0.2-2β* | *β* | *β* | 0.2 |
| $A^2 \subseteq (\overline{A}^2, \overline{A}^3)$ | *0* | *0.5- β* | *β* | 0.5 |
| $A^3 \subseteq \overline{A}^3$ | *0* | *0* | *0.3* | 0.3 |
| Total $\overline{m}\left(\overline{A}^i\right)$ | 0.2-2 β | 0.5 | 0.3+2 β | 1 |

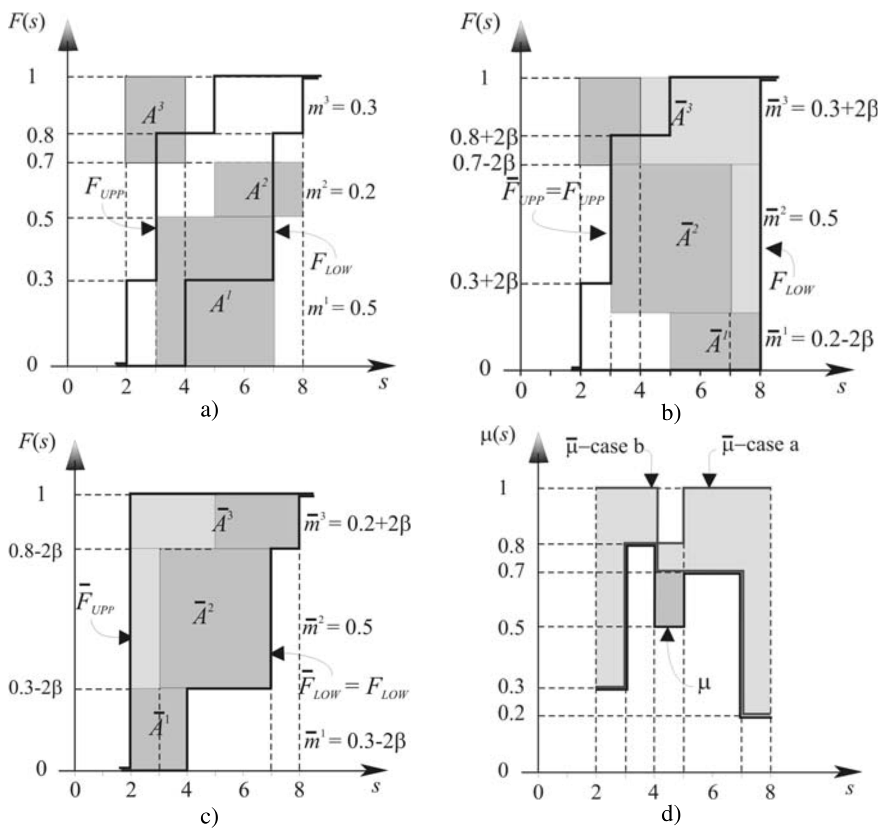



**Fig. 5.9** Example 5.3: (a) Original focal elements and upper/lower CDFs; (b) consonant approximation, case (a); (c) consonant approximation, case (b); (d) contour functions of the original included random set (μ) and of the inclusive consonant random sets for cases a) and b)

**Table 5.7** Example 5.4: Matrix **w**

| | $\bar{A}^1$ | $\bar{A}^2$ | $\bar{A}^3$ | Total $m(A^i)$ |
|---|---|---|---|---|
| $A^1 \subseteq (\bar{A}^1, \bar{A}^2, \bar{A}^3)$ | *0.3-2β* | *β* | *β* | 0.3 |
| $A^2 \subseteq (\bar{A}^2, \bar{A}^3)$ | *0* | *0.5- β* | *β* | 0.5 |
| $A^3 \subseteq \bar{A}^3$ | *0* | *0* | *0.2* | 0.2 |
| Total $\bar{m}(\bar{A}^i)$ | 0.3-2 β | 0.5 | 0.2+2 β | 1 |

#### 5.1.3.2 Case (b)

The same procedure as in case (a) can be used, provided upper and lower extremes ("*u*" and "*l*") are interchanged and the words "right" and "left" are also interchanged.

**Example 5.4.** Consider again the same random sets as in Example 5.3, $\mathcal{F}$ : {([3, 7], 0.5), ([5, 8], 0.2), ([2, 4], 0.3)}. Re-order the focal elements $\mathcal{F}$ : {([2, 4], 0.3), ([3, 7], 0.5), ([5, 8], 0.2)}. By applying Step 3 with apexes "*u*" and "*l*" interchanged, the focal sets are {[2, 4], [2, 7], [2, 8]}. Table 5.7 gives matrix **w.** Figure 5.9c shows upper and lower CDFs of $\bar{\mathcal{F}}$ ; when $\beta$=0, the lower CDF coincides with the lower CDF of $\mathcal{F}$ . Figure 5.22d displays the contour function μ of $\mathcal{F}$ and the contour functions of the inclusive consonant random sets for the cases a) and b), which closely match the contour function of $\mathcal{F}$ to the left and right, respectively, of its maximum value, i.e. 0.8.

#### 5.1.3.3 Case (c)

To preserve symmetry, consider the consonant projection produced by the contour function of $\mathcal{F}$ . With slight modifications, the algorithm below was suggested by Dubois and Prade (1986):

1. Let $M$ be the set containing the values of the contour function (Plausibility of the singletons, Eqs. (3.24)) of $\mathcal{F}=(\mathcal{A}, m)$: $M = \{\alpha = Pla(\{s\}) : s \in S\}$. Order $M$ in decreasing order: $\alpha^1 > \ldots > \alpha^p$.
2. Define a nested set $\mathcal{A}_*$ of focal elements $A_*^i$ as the family of $\alpha$–cuts (Eq.(3.27)) induced by $M$ so that $A_*^i \subset A_*^{i+1}$:

$$\mathcal{A}_* = \{ A_*^i = \{s : Pl(\{s\}) \geq \alpha^i\} : \alpha^i \in M \} \tag{5.24}$$

3. Define the candidate focal sets by using a mapping $f: \mathcal{A} \rightarrow \mathcal{A}_*$

$$f(A) = A_*^i : \quad A \subseteq A_*^i, A \not\subset A_*^{i-1} \tag{5.25}$$

4. Let the including random set $\overline{\mathcal{F}} = \left(\overline{\mathcal{A}}, \overline{m}\right)$ be defined as:

$$\overline{\mathcal{A}} = f\left(\mathcal{A}\right); \qquad \overline{m}\left(\overline{A}^i\right) = \sum_{A:\overline{A}^i = f(A)} m\left(A\right) \tag{5.26}$$

The inclusion matrix has entries $w_{ij}=m(A^i)$ if $\overline{A}^j = f(A^i)$, $w_{ij}=0$ otherwise.

**Table 5.8** Example 5.5: Matrix **w**

| | $\overline{A}^1$ | $\overline{A}^2$ | $\overline{A}^3$ | Total $m(A^i)$ |
|---|---|---|---|---|
| $A^1 \subseteq (\overline{A}^1, \overline{A}^2, \overline{A}^3)$ | *0.5* | *0* | *0* | 0.5 |
| $A^2 \subseteq \overline{A}^3$ | *0* | *0* | *0.2* | 0.2 |
| $A^3 \subseteq (\overline{A}^2, \overline{A}^3)$ | *0* | *0.3* | *0* | 0.3 |
| Total $\overline{m}\left(\overline{A}^i\right)$ | 0.5 | 0.3 | 0.2 | 1 |

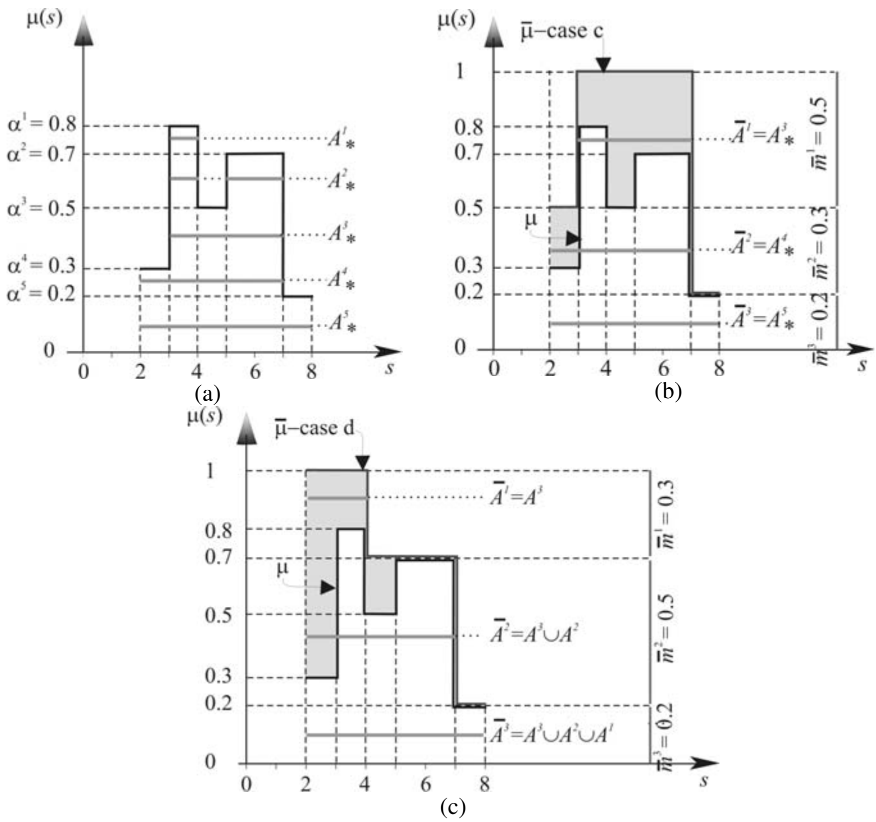



**Fig. 5.10** Example 5.5: (a) Included random set $\mathcal{F}$ and family of sets $\mathcal{A}_*$; (b) $\overline{\mathcal{F}}$ and contour function for the consonant approximation for case c); (c) $\overline{\mathcal{F}}$ and contour function for the consonant approximation in case d)

**Example 5.5.** Consider again the random set in Example 5.3: $\mathcal{F}$: {([3, 7], 0.5), ([5, 8], 0.2), ([2, 4], 0.3)}. Figure 5.10a shows the random set $\mathcal{F}$, its contour function and the sets in family $\mathcal{A}*$ generated by the α-cuts. The mapping in Eq. (5.25) is as follows: $A^1 \mapsto A_*^3$; $A^2 \mapsto A_*^5$; $A^3 \mapsto A_*^4$. Consequently, $\bar{\mathcal{A}} = \left\{ \bar{A}^1 = A_*^3, \bar{A}^2 = A_*^4, \bar{A}^3 = A_*^5 \right\}$; $m$={0.5, 0.3, 0.2} as displayed in Figure 5.10b. Table 5.8 gives matrix **w**. Figure 5.10b demonstrates that Case (c) treats the entire support of the included random set equally.

### 5.1.3.4 Case (d)

When $S$ is finite, a different optimality criterion proposed by Dubois and Prade (1990) consists of minimizing the cardinality of the including consonant random set $\bar{\mathcal{F}}$, which is defined as:

$$\left|\bar{\mathcal{F}}\right| := \sum_{\bar{A}^i \in \bar{\mathcal{F}}} \bar{m}\left(\bar{A}^i\right)\left|\bar{A}^i\right| = \sum_{s \in S} \bar{\mu}_{\bar{F}}\left(s\right) := \left|\bar{F}\right| \tag{5.27}$$

where $\left|\bar{F}\right|$ is the cardinality of the fuzzy set $\bar{F}$ (De Luca and Termini 1972) equivalent to $\bar{\mathcal{F}}$ through Eq. (3.24). The equality can easily be verified using Eqs. (3.26) and (3.27).

The candidate focal sets are now made up of unions of focal sets in $\mathcal{F}$. In order to calculate the cardinality generated by all such possible unions, let π be a permutation of the indexes {1,…, $n$ = |$\mathcal{A}$|}. For each permutation π, the following including consonant random set is generated:

$$\bar{\mathcal{A}} = \left\{ \bar{A}^i : \bar{A}^i = \bigcup_{j=1}^{i} A^{\pi(j)} \right\}; \qquad \bar{m}\left(\bar{A}^i\right) = m\left(A^{\pi(i)}\right) \qquad i=1,\dots,n \tag{5.28}$$

Strong inclusion is ensured because $A^i \subseteq \bar{A}^{\pi^{-1}(i)}$, and thus $w_{i\pi^{-1}(i)} = m\left(A^i\right)$; 0 otherwise. The optimum random set is the one that minimizes $\left|\bar{\mathcal{F}}\right|$. Dubois and Prade (1990, page 436) give a heuristic algorithm to carry out this minimization, which, however, is not guaranteed to converge to the optimal solution.

A possible extension to $S = \mathbb{R}$ when the number of focal elements is finite and no focal set degenerates to a point is to consider the (Lebesgue) measure $\left|\bar{A}^i\right| := \int_{\bar{A}^i} ds$.

**Example 5.6.** Consider again the random set in Example 5.3 and Example 5.5: $\mathcal{F}$: {([5, 8], 0.2), ([3, 7], 0.5), ([2, 4], 0.3)} with $|A^1|=4$; $|A^2|=5$; $|A^3|=3$. Table 5.9 gives all possible including random sets, from which the last permutation yields the minimum cardinality. This is the random set $\overline{\mathcal{F}}$ : {([2, 4], 0.3), ([2, 7], 0.5), ([2, 8], 0.2)}, which coincides with the including random set found in Case (b). Figure 5.10c shows its contour function (Case d).

**Table 5.9** Example 5.6: Enumeration of all possible including random sets

| π(1) | π(2) | π(3) | $\overline{A}^1$ | $\overline{A}^2$ | $\overline{A}^3$ | $\lvert\overline{A}^1\rvert$ | $\lvert\overline{A}^2\rvert$ | $\lvert\overline{A}^3\rvert$ | $\overline{m}^1$ | $\overline{m}^2$ | $\overline{m}^3$ | $\lvert\overline{\mathcal{F}}\rvert$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 1 | 1∪2 | 1∪2∪3 | 4 | 6 | 7 | 0.2 | 0.5 | 0.3 | 5.9 |
| 1 | 3 | 2 | 1 | 1∪3 | 1∪3∪2 | 4 | 7 | 7 | 0.2 | 0.3 | 0.5 | 6.4 |
| 2 | 1 | 3 | 2 | 2∪1 | 2∪1∪3 | 5 | 6 | 7 | 0.5 | 0.2 | 0.3 | 5.8 |
| 2 | 3 | 1 | 2 | 2∪3 | 2∪3∪1 | 5 | 6 | 7 | 0.5 | 0.3 | 0.2 | 5.7 |
| 3 | 1 | 2 | 3 | 3∪1 | 3∪1∪2 | 3 | 7 | 7 | 0.3 | 0.2 | 0.5 | 5.8 |
| 3 | 2 | 1 | 3 | 3∪2 | 3∪2∪1 | 3 | 6 | 7 | 0.3 | 0.5 | 0.2 | 5.3 |

**Table 5.10** Example 5.6: Matrix **w**

| | $\overline{A}^1 = A^3$ | $\overline{A}^2 = A^3 \cup A^2$ | $\overline{A}^3 = A^3 \cup A^2 \cup A^1$ | Total $m(A^i)$ |
|---|---|---|---|---|
| $A^1 \subseteq \overline{A}^3$ | *0* | *0* | *0.2* | 0.2 |
| $A^2 \subseteq (\overline{A}^2, \overline{A}^3)$ | *0* | *0.5* | *0* | 0.5 |
| $A^3 \subseteq (\overline{A}^1, \overline{A}^2, \overline{A}^3)$ | *0.3* | *0* | *0* | 0.3 |
| Total $\overline{m}\left(\overline{A}^i\right)$ | 0.3 | 0.5 | 0.2 | 1 |

### *5.1.4 Inclusion Properties for Random Relations under the Hypotheses of Random Set Independence and Non-interactivity*

In Section 4.3 (page 158), several notions of independence were introduced for random sets; in particular, Section 4.3.1 (page 159) introduced the definition of random set independence, whereas Section 4.3.5 (page 174) introduced the definition of fuzzy Cartesian product. In this section, we bring together these two notions of independence and the notion of inclusion to study the following problem.

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two random sets on $S_1$ and $S_2$, respectively, and let $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$ be two consonant inclusions of theirs, i.e. $\mathcal{F}_i \subseteq \overline{\mathcal{F}}_i$ ($i$ = 1, 2). As illustrated in Figure 5.11, let $(\mathcal{Z}_1, z_1)$ be the random relation on $S_1 \times S_2$ obtained from $\mathcal{F}_1$ and $\mathcal{F}_2$ under the hypothesis of random set independence, let $(\mathcal{Z}_2, z_2)$ be the fuzzy Cartesian product on $S_1 \times S_2$ obtained from $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$, and let $(\mathcal{Z}_3, z_3)$ be the random relation on $S_1 \times S_2$ obtained from $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$ under the hypothesis of random set independence. The purpose of this Section is to investigate whether:

1) $(\mathcal{Z}_1, z_1) \subseteq (\mathcal{Z}_2, z_2)$. In this case, we will say that fuzzy Cartesian product preserves inclusion.
2) $(\mathcal{Z}_1, z_1) \subseteq (\mathcal{Z}_3, z_3)$. In this case, we will say that random set independence preserves inclusion.

The practical consequence of this study (Tonon and Chen 2005) is as follows: if the inclusions are true, then the Belief-Plausibility intervals calculated with $(\mathcal{Z}_2, z_2)$ and $(\mathcal{Z}_3, z_3)$ include the Belief-Plausibility intervals calculated with $(\mathcal{Z}_1, z_1)$ (Theorem 5.1 (*ii*), page 218). If these bounds are interpreted as upper and lower probabilities, then probability bounds calculated with $(\mathcal{Z}_2, z_2)$ and $(\mathcal{Z}_3, z_3)$ include the probability bounds calculated with $(\mathcal{Z}_1, z_1)$. For example, the authors used the hypothesis of fuzzy Cartesian product to constrain parameters in the formulation of single and multi-objective optimizations of engineering systems (Tonon and Bernardini 1998; Tonon and Bernardini 1999).

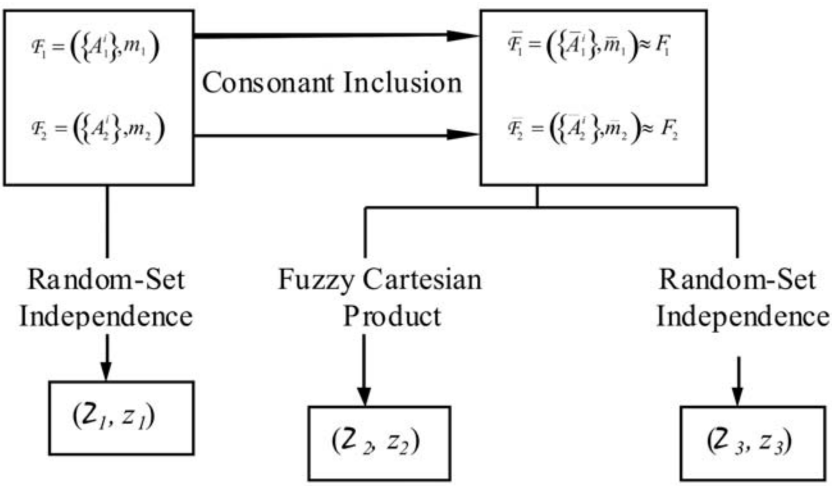



**Fig. 5.11** Schematic of the random sets used in the section: $\approx F_i$ indicates equivalent fuzzy set $F_i$

### 5.1.4.1 Fuzzy Cartesian Product

We will use a counterexample to show that the hypotheses above do not necessarily imply that $(\mathcal{Z}_1, z_1) \subseteq (\mathcal{Z}_2, z_2)$. In this example, $S = S_1 \times S_2$ where $S_1 = S_2 = \mathbb{R}$ as illustrated in Table 5.11, which gives the marginal random sets $\mathcal{F}_1$ and $\mathcal{F}_2$ defined on $S_1$ and $S_2$, respectively. Following the procedure for inclusion of 1-D random sets described in Section 5.1.3.1 (page 221), one obtains consonant random sets $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$, which include $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. This is detailed in Table 5.12, which gives matrices **w** for $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and in Table 5.13. Notice that $\mathcal{F}_1$ and $\overline{\mathcal{F}}_1$ are the random set and consonant random set, respectively, in Example 5.3. In order to understand the effect of the parameter β introduced in Section 5.1.3 (page 221), let $\beta = 10^{-6}$. As explained in Sections 3.2.4, 4.1, and 4.3.5, consonant random set $\mathcal{F}_i$ is equivalent to a fuzzy set, $F_i$. Hence, in Table 5.13 focal elements $\overline{A}$ are also considered as $\alpha$-cuts with $\alpha$-levels $\mu_{F_k}\left(\overline{A}_k^i\right)$ for $F_k$. For example: $\mu_{F_1}\left(\overline{A}_1^1\right) = \sum_{i=1}^{3}\overline{m}_1\left(\overline{A}_1^i\right) = 1$, $\mu_{F_1}\left(\overline{A}_1^2\right) = \sum_{i=2}^{3}\overline{m}_1\left(\overline{A}_1^i\right) =$ 0.800002, $\mu_{F_1}\left(\overline{A}_1^3\right) = \overline{m}_1\left(\overline{A}_1^3\right) =$ 0.300002.

**Table 5.11** Marginal random sets

| $\mathcal{F}_1 = \left(\left\{A_1^i\right\}, m_1\right)$ | | $\mathcal{F}_2 = \left(\left\{A_2^i\right\}, m_2\right)$ | |
|---|---|---|---|
| $A_1^i$ | $m_1(A_1^i)$ | $A_2^i$ | $m_2(A_2^i)$ |
| $A_1^1$ =[5, 8] | 0.2 | $A_2^1$ =[3, 7] | 0.7 |
| $A_1^2$ =[3, 7] | 0.5 | $A_2^2$ =[2, 5] | 0.1 |
| $A_1^3$ =[2, 4] | 0.3 | $A_2^3$ =[1, 8] | 0.2 |

**Table 5.12** Matrices **w** and consonant marginal random sets

| | $\overline{A}_1^1$ | $\overline{A}_1^2$ | $\overline{A}_1^3$ | Total $m(A_1^i)$ |
|---|---|---|---|---|
| $A_1^1 \subseteq \overline{A}_1^1, \overline{A}_1^2, \overline{A}_1^3$ | 0.2-2β | β | β | **0.2** |
| $A_1^2 \subseteq \overline{A}_1^2, \overline{A}_1^3$ | 0 | 0.5- β | β | **0.5** |
| $A_1^3 \subseteq \overline{A}_1^3$ | 0 | 0 | 0.3 | **0.3** |
| Total $\overline{m}\left(\overline{A}_1^i\right)$ | **0.2-2 β** | **0.5** | **0.3+2 β** | **1** |

| | $\overline{A}_2^1$ | $\overline{A}_2^2$ | $\overline{A}_2^3$ | Total $m(A_2^i)$ |
|---|---|---|---|---|
| $A_2^1 \subseteq \overline{A}_2^1, \overline{A}_2^2, \overline{A}_2^3$ | 0.7-2β | β | β | **0.7** |
| $A_2^2 \subseteq \overline{A}_2^2, \overline{A}_2^3$ | 0 | 0.1- β | β | **0.1** |
| $A_2^3 \subseteq \overline{A}_2^3$ | 0 | 0 | 0.2 | **0.2** |
| Total $\overline{m}\left(\overline{A}_2^i\right)$ | **0.7-2 β** | **0.1** | **0.2+2 β** | **1** |

**Table 5.13** Consonant marginal random sets and relevant α-levels

| $\overline{A}_1^i$ | $\overline{m}\left(\overline{A}_1^i\right)$ | $\mu_{F_1}\left(\overline{A}_1^i\right)$ | $\overline{A}_2^i$ | $\overline{m}\left(\overline{A}_2^i\right)$ | $\mu_{F_2}\left(\overline{A}_2^i\right)$ |
|---|---|---|---|---|---|
| $\overline{A}_1^1$ =[5, 8] | 0.199998 | 1 | $\overline{A}_2^1$ =[3, 7] | 0.699998 | 1 |
| $\overline{A}_1^2$ =[3, 8] | 0.5 | 0.800002 | $\overline{A}_2^2$ =[2, 7] | 0.1 | 0.300002 |
| $\overline{A}_1^3$ =[2, 8] | 0.300002 | 0.300002 | $\overline{A}_2^3$ =[1, 8] | 0.200002 | 0.200002 |

Now, as depicted in Figure 5.11, let us calculate random relation ($\mathcal{Z}_1$, $z_1$) from the marginals using the hypothesis of random-set independence (Table 5.14). Consonant random relation ($\mathcal{Z}_2$, $z_2$) was calculated in two steps:

1) The $\alpha$-levels for the decomposable fuzzy relationship $F$ equivalent to ($\mathcal{Z}_2$, $z_2$) were calculated using Eq. (4.115) and were reported in Table 5.15.
2) Notice that some $\alpha$-cuts in Table 5.15 have the same $\alpha$-level; for example $\alpha$-cuts $\left(\bar{A}_1^1 \times \bar{A}_2^3\right)$, $\left(\bar{A}_1^2 \times \bar{A}_2^3\right)$ and $\left(\bar{A}_1^3 \times \bar{A}_2^3\right)$ all have $\alpha$-level 0.200002. They can be grouped into one $\alpha$-cut as shown in Table 5.16, and the non-interactive random Cartesian product ($\mathcal{Z}_2$, $z_2$) equivalent to the decomposable fuzzy relationship $F$ can eventually be calculated using Eq. (4.116) (fourth column in Table 5.16).

**Table 5.14** Random-set independent random relation ($Z_1$, $z_1$) with focal elements $C_i$ obtained from marginals $\mathbf{F}_1$ and $\mathbf{F}_2$ given in Table 5.11

| $C_i$ | $A_1^i \times A_2^j$ | $z_1(C_i) = m_1(A_1^i)\ m_2(A_2^j)$ |
|---|---|---|
| $C_1$ | 1×1; [5,8]×[3,7] | 0.14 |
| $C_2$ | 1×2; [5,8]×[2,5] | 0.02 |
| $C_3$ | 1×3; [5,8]×[1,8] | 0.04 |
| $C_4$ | 2×1; [3,7]×[3,7] | 0.35 |
| $C_5$ | 2×2; [3,7]×[2,5] | 0.05 |
| $C_6$ | 2×3; [3,7]×[1,8] | 0.10 |
| $C_7$ | 3×1; [2,4]×[3,7] | 0.21 |
| $C_8$ | 3×2; [2,4]×[2,5] | 0.03 |
| $C_9$ | 3×3; [2,4]×[1,8] | 0.06 |

**Table 5.15** Fuzzy Cartesian product $F$ obtained from fuzzy sets $F_1$ and $F_2$ given in Table 5.13

| $D_i$ | $\bar{A}_1^i \times \bar{A}_2^j$ | $\mu_F(D_i) = \min\{(\mu(\bar{A}_1^i), \mu(\bar{A}_2^j)\}$ |
|---|---|---|
| $D_1$ | 1×1; [5,8]×[3,7] | 1 |
| $D_2$ | 1×2; [5,8]×[2,7] | 0.300002 |
| $D_3$ | 1×3; [5,8]×[1,8] | 0.200002 |
| $D_4$ | 2×1; [3,8]×[3,7] | 0.800002 |
| $D_5$ | 2×2; [3,8]×[2,7] | 0.300002 |
| $D_6$ | 2×3; [3,8]×[1,8] | 0.200002 |
| $D_7$ | 3×1; [2,8]×[3,7] | 0.300002 |
| $D_8$ | 3×2; [2,8]×[2,7] | 0.300002 |
| $D_9$ | 3×3; [2,8]×[1,8] | 0.200002 |

**Table 5.16** Fuzzy Cartesian product $F$ obtained from fuzzy sets $F_1$ and $F_2$ given in Table 5.13 and its equivalent consonant random Cartesian product ($Z_2$, $z_2$) with focal elements $D_i$

| $D_i$ | $\overline{A}_1^i \times \overline{A}_2^j$ | $\mu_F(D_i)$ | $z_2(D_i)$ |
|---|---|---|---|
| $D_1$ | $\overline{A}_1^1 \times \overline{A}_2^1 = [5,8]\times[3,7]$ | 1 | 0.199998 |
| $D_2$ | $\overline{A}_1^2 \times \overline{A}_2^1 = [3,8]\times[3,7]$ | 0.800002 | 0.5 |
| $D_3$ | $\left(\overline{A}_1^1 \times \overline{A}_2^2\right) \cup \left(\overline{A}_1^2 \times \overline{A}_2^2\right) \cup \left(\overline{A}_1^3 \times \overline{A}_2^1\right) \cup \left(\overline{A}_1^3 \times \overline{A}_2^2\right) = [2,8]\times[2,7]$ | 0.300002 | 0.1 |
| $D_4$ | $\left(\overline{A}_1^1 \times \overline{A}_2^3\right) \cup \left(\overline{A}_1^2 \times \overline{A}_2^3\right) \cup \left(\overline{A}_1^3 \times \overline{A}_2^3\right) = [2,8]\times[1,8]$ | 0.200002 | 0.200002 |

**Table 5.17** Focal elements for the calculation of Belief and Plausibility of $C_1 = D_1$

| Random relation ($\mathcal{Z}_1$, $z_1$), Table 5.14 | | Random relation ($\mathcal{Z}_2$, $z_2$), Table 5.16 | |
|---|---|---|---|
| $C_i : C_i \subseteq C_1$ | $C_i : C_i \cap C_1 \neq \varnothing$ | $D_i : D_i \subseteq C_1 = D_1$ | $D_i : D_i \cap C_1 \neq \varnothing$ |
| $C_1$ | $C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9$ | $D_1$ | $D_1, D_2, D_3, D_4$ |

For $C_1 = D_1$, Table 5.17 gives the focal elements needed to calculate Belief and Plausibility using ($\mathcal{Z}_1$, $z_1$) and ($\mathcal{Z}_2$, $z_2$), and one obtains $[Bel_{Z_1}(C_1), Pl_{Z_1}(C_1)] = [0.14, 1]$ and $[Bel_{Z_2}(C_1), Pl_{Z_2}(C_1)] = [0.19998, 1]$. Since $Bel_{Z_2}(C_1) > Bel_{Z_1}(C_1)$, then $[Bel_{Z_1}(C_1), Pl_{Z_1}(C_1)] \not\subset [Bel_{Z_2}(C_1), Pl_{Z_2}(C_1)]$, and from Theorem 5.1(*iv*), one concludes that ($\mathcal{Z}_1$, $z_1$) $\not\subset$ ($\mathcal{Z}_2$, $z_2$). Thus, this counterexample shows that non-interactivity does not necessarily preserve inclusion.

### 5.1.4.2 Random Set Independence

When random set independence is used on $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$ (Figure 5.11), inclusion is preserved, i.e. ($\mathcal{Z}_1$, $z_1$) $\subseteq$ ($\mathcal{Z}_3$, $z_3$). In order to show this property, let us consider each of the three conditions in Eq. (5.4)):

(i) $\forall\left(A_1^i \times A_2^j\right) \in \mathcal{Z}_1$, $\exists\ \overline{A}_1^k, \overline{A}_2^l : A_1^i \subseteq \overline{A}_1^k; A_2^j \subseteq \overline{A}_2^l$ because $\mathcal{F}_1 \subseteq \overline{\mathcal{F}}_1$ and $\mathcal{F}_2 \subseteq \overline{\mathcal{F}}_2$, hence $\left(A_1^i \times A_2^j\right) \subseteq \left(\overline{A}_1^k \times \overline{A}_2^l\right)$;

(ii) Similar to (i)

(iii) Let $\mathbf{w}_1$, $\mathbf{w}_2$ be the inclusion matrixes for the marginals. Consider matrix $\mathbf{w}$: $w(A_1^i \times A_2^j, \bar{A}_1^k \times \bar{A}_2^l) = w_{1,ik} \cdot w_{2,jl}$. Now, $\forall \left(A_1^i \times A_2^j\right)$:

$$\begin{aligned} \sum_{\bar{A}_1^k \times \bar{A}_2^l : A_1^i \times A_2^j \subseteq \bar{A}_1^k \times \bar{A}_2^l} w\left(A_1^i \times A_2^j, \bar{A}_1^k \times \bar{A}_2^l\right) &= \sum_{\bar{A}_1^k \times \bar{A}_2^l : A_1^i \times A_2^j \subseteq \bar{A}_1^k \times \bar{A}_2^l} w_{1,ik} \cdot w_{2,jl} \\ = \sum_{\bar{A}_1^k : A_1^i \subseteq \bar{A}_1^k} w_{1,ik} \cdot \sum_{\bar{A}_2^l : A_2^j \subseteq \bar{A}_2^l} w_{2,jl} &= m_1\left(A_1^i\right) \cdot m_2\left(A_2^j\right) \end{aligned} \tag{5.29}$$

Likewise for $\bar{A}_1^k \times \bar{A}_2^l$.

Although $(\mathcal{Z}_1, z_1) \subseteq (\mathcal{Z}_3, z_3)$, $(\mathcal{Z}_3, z_3)$ is not necessarily consonant. As a counterexample, consider the consonant marginal inclusions in Table 5.13: $\mathcal{Z}_3$ includes the focal elements $A^1 = [5, 8] \times [3, 7]$; $A^2 = [5, 8] \times [2, 7]$; $A^3 = [3, 8] \times [3, 7]$. Although $A^1 \subseteq A^2$ and, $A^1 \subseteq A^3$, $A^2$ and $A^3$ satisfy no inclusion relationship. An optimal consonant inclusion (i.e a consonant *outer approximation*; see Dubois and Prade 1990 p. 425 for the definitions of *optimal inner and outer approximations* of a random set) to $(\mathcal{Z}_3, z_3)$ can nevertheless be uniquely determined as specified in the following theorem.

**Theorem 5.2** (Dubois and Prade 1990)**.** *Let $F_1$ and $F_2$ be the fuzzy sets equivalent to consonant random sets $\bar{\mathcal{F}}_1 \supseteq \mathcal{F}_1$ and $\bar{\mathcal{F}}_2 \supseteq \mathcal{F}_2$, respectively. Let $\mathcal{L}(F_1 \times F_2)$ be the set of $\alpha$-cuts of the fuzzy Cartesian product $F_1 \times F_2$. The minimal outer approximation, H, to $(\mathcal{Z}_3, z_3)$ whose focal sets are in $\mathcal{L}(F_1 \times F_2)$ has membership function*:

$$\mu_H(s_1, s_2) = \min\left\{\mu_{F_1}(s_1) \cdot \left(2 - \mu_{F_1}(s_1)\right), \mu_{F_2}(s_2) \cdot \left(2 - \mu_{F_2}(s_2)\right)\right\} \tag{5.30}$$

*Proof* : Let $M_H = M_{F_1} \cup M_{F_2} = \{\alpha^1, \ldots, \alpha^{n+1}\}$ with $\alpha^1 = 1 > \alpha^2 > \ldots > \alpha^{n+1} = 0$ be the union of the $\alpha$-values attained by $F_1$, $M_{F_1}$, and $F_2$, $M_{F_2}$. Let $A_j^i = {}^{\alpha_i}A_j$, $j = 1, 2$, so that $\mathcal{L}(F_1 \times F_2) = \{\bar{B}^i = A_1^i \times A_2^i, i = 1, \ldots, n\}$. Let $m^i = \alpha^i - \alpha^{i+1}$, $i = 1, \ldots, n$. $\mathcal{Z}_3$ is composed of $n^2$ focal sets of the kind $B^{i+(j-1)\cdot n} = A_1^i \times A_2^j$, $i,j = 1, \ldots, n$, with assignment $m^i \cdot m^j$. Therefore, some focal sets are not in $\mathcal{L}(F_1 \times F_2)$ (Section 4.3.5); in particular, the focal sets included in $A_1^i \times A_2^i$ but not in $A_1^{i-1} \times A_2^{i-1}$ are $\{A_1^i \times A_2^j : j < i\} \cup \{A_1^j \times A_2^i : j < i\}$. Let $k \pmod n$ be the remainder of $k/n$. Since $B^k$ is generated by $A_1^i \times A_2^j$ where

$$i = \begin{cases} k \pmod n & \text{if } k \pmod n > 0 \\ n & \text{otherwise} \end{cases}; \quad j = \begin{cases} INT(k/n) + 1 & \text{if } k \pmod n > 0 \\ k/n & \text{otherwise} \end{cases}, \tag{5.31}$$

the inclusion is ensured by matrix **w** with entries $w_{ki} = m^i \cdot m^j$ if $i,j \leq i$; 0 otherwise. This entails that the probability assignment of the including random set is:

$$\begin{aligned} &\overline{m}\left(\overline{B}^1 = A_1^1 \times A_2^1\right) = \left(m^1\right)^2 \\ &\overline{m}\left(\overline{B}^i = A_1^i \times A_2^i\right) = \left(m^i\right)^2 + 2m^i \sum_{j<i} m^j \quad , i=2, ..n \end{aligned} \tag{5.32}$$

Let $(s_1, s_2) \in \left(A_1^i \times A_2^i\right) - \left(A_1^{i-1} \times A_2^{i-1}\right)$, that is, $\min\left\{\mu_{F_1}(s_1), \mu_{F_2}(s_2)\right\} = \alpha^i$. Then,

$$\begin{aligned} \mu_H(s_1, s_2) &= \sum_{j \geq i} \overline{m}\left(A_1^j \times A_2^j\right) && \text{Eqs. (3.23) and (3.20)} \\ &= \sum_{j \geq i}\left[\left(m^j\right)^2 + 2m^j \sum_{k<j} m^k\right] && \text{Eq. (5.32)} \\ &= \sum_{j \geq i}\left[\left(m^j\right)^2 + 2m^j\left(1 - \sum_{k \geq j} m^k\right)\right] && \text{Eq. (3.3)} \\ &= 2\sum_{j \geq i} m^j + \sum_{j \geq i}\left(m^j\right)^2 - 2\sum_{k \geq j \geq i} m^j m^k \\ &= 2\alpha^i - \sum_{j \geq i}\left(m^j\right)^2 - 2\sum_{k > j \geq i} m^j m^k \\ &= 2\alpha^i - \left[\sum_{j \geq i}\left(m^j\right)\right]^2 = 2\alpha^i - \left(\alpha^i\right)^2 \\ &= h\left(\alpha^i\right) = h\left(\min\left\{\mu_{F_1}(s_1), \mu_{F_2}(s_2)\right\}\right) \text{ where } h(x) := 2x - (x)^2 \end{aligned} \tag{5.33}$$

Eq. (5.30) follows because $h(x)$ is strictly increasing in [0, 1], i.e. $H = F_1{*} \times F_2{*}$ where

$$\mu_{F_i*}(s_i) = h\left(\mu_{F_i*}(s_i)\right) \tag{5.34}$$

■

A weak inner consonant inclusion, $L \subseteq_{\text{weak}} (\mathcal{Z}_3, z_3)$ can also be easily calculated as detailed in the following Theorem 5.4. Before proving Theorem 5.4, we need the result in Theorem 5.3.

**Theorem 5.3** ((Dubois and Prade 1986), page 214; (Dubois and Prade 1990), page 427)

(i) *Call a consistent random set (relation)* $\mathcal{F}=(\mathcal{A}, m)$ *a random set such that* $\bigcap_{A\in\mathcal{A}} A \neq \varnothing$. *The weak optimal inner consonant approximation of a consistent random set (relation)* $\mathcal{F}$ *is the fuzzy set (relation)* $F_*$ *whose membership function is the contour function of* $\mathcal{F}$.

(ii) *If* $\mathcal{F}$ *is not consistent, then* $F_*$ *yields an inner approximation to the Plausibility of* $\mathcal{F}$.

*Proof*: (i) Notice that Plausibility can also be written as:

$$\forall T,\ Pla_{\mathcal{F}}\left(T\right)=\sum_{A\subseteq S} m\left(A\right)\cdot\sup_{s\in T}\ I_A\left(s\right) \tag{5.35}$$

where $I_A$ is the characteristic function of $A$. Let $\mathcal{F}_*$ be the random set equivalent to $F_*$.The weak inclusion is ensured by the inequality:

$$\begin{aligned} &Pla_{\mathcal{F}}\left(T\right)\geq\sup_{s\in T}\sum_{A\subseteq S} m\left(A\right)\cdot I_A\left(s\right)=\sup_{s\in T} Pla_{\mathcal{F}}\left(\left\{s\right\}\right)\\ &\overset{(Eq.(3.24))}{=}\ \sup_{s\in T}\mu_{F_*}\left(s\right)\overset{(Eq.(3.25))}{=}\ Pla_{\mathcal{F}_*}\left(T\right)\end{aligned} \tag{5.36}$$

As for Belief, since the random set is consistent, $\sup_{s\in S} Pla_{\mathcal{F}_*}\left(\left\{s\right\}\right)=1$, Eqs. (3.4) and (3.7) hold, and thus

$$Bel_{\mathcal{F}_*}\left(T\right)\overset{Eq.(3.7)}{=}1-Pla_{\mathcal{F}_*}\left(T^C\right)\overset{Eq.(5.36)}{\geq}1-Pla_{\mathcal{F}}\left(T^C\right)\overset{Eq.(3.7)}{=}Bel_{\mathcal{F}}\left(T\right) \tag{5.37}$$

Let *Pla* be the Plausibility (or Possibility) measure of any consonant random set such that $Pla \leq Pla_{\mathcal{F}}$, and let $\mu$ be the membership function of its equivalent fuzzy set. Then $\forall s\in S$, $\mu(s) \leq Pla_{\mathcal{F}}(s)=\mu_{F_*}\left(s\right)$.

(ii) If $\mathcal{F}$ is not consistent, Eqs. (3.4) and (3.7), and thus (5.37) are no longer valid because $m(\varnothing)>0$. As a result, an inner bound on the Belief cannot be ensured. ■

**Theorem 5.4** (Dubois and Prade 1990). *Let* $F_1$ *and* $F_2$ *be the fuzzy sets equivalent to consonant random sets (relations)* $\overline{\mathcal{F}}_1\supseteq\mathcal{F}_1$ *and* $\overline{\mathcal{F}}_2\supseteq\mathcal{F}_2$, *respectively. The best weak inner consonant approximation,* $L\subseteq_{\text{weak}}(\mathcal{Z}_3, z_3)$ *has membership function*:

$$\mu_L\left(s_1,s_2\right)=\mu_{F_1}\left(s_1\right)\cdot\mu_{F_2}\left(s_2\right) \tag{5.38}$$

*Proof* : Notice that $(\mathcal{Z}_3, z_3)$ is consistent, i.e. $\bigcap_{B^k \in \mathcal{Z}} B^k \neq \varnothing$, because $A_1^1 \times A_2^1 = B^1 \subseteq B^k \; \forall k$. In order to show that the contour function of $(\mathcal{Z}_3, z_3)$ is given by Eq. (5.38), let $M_H = M_{F_1} \cup M_{F_2} = \{\alpha^1, \ldots, \alpha^{n+1}\}$ with $\alpha^1=1>\alpha^2> \ldots >\alpha^{n+1}=0$ being the union of the α-values attained by $F_1$, $M_{F_1}$, and $F_2$, $M_{F_2}$. Let $m^i=\alpha^i-\alpha^{i+1}$, $i=1,\ldots,n$. $\mathcal{Z}_3$ is composed of focal sets of the kind $A_1^i \times A_2^j$, $i,j=1,\ldots,n$, with assignment

$$\begin{aligned} m^i m^j &= \left(\alpha^i - \alpha^{i+1}\right)\left(\alpha^j - \alpha^{j+1}\right) = \\ &\alpha^i \cdot \alpha^j - \alpha^i \cdot \alpha^{j+1} - \alpha^{i+1} \cdot \alpha^j + \alpha^{i+1} \cdot \alpha^{j+1} \end{aligned} \tag{5.39}$$

Let $p\geq 1$, $q\geq 1$ be the smallest indexes such that $(s_1, s_2) \in A_1^p \times A_2^q$ (i.e. $(s_1, s_2) \notin A_1^{p+1} \times A_2^q$ and $(s_1, s_2) \notin A_1^p \times A_2^{q+1}$), then

$$\begin{aligned} &\mu_L(s_1, s_2) = Pla\left(\{s_1, s_2\}\right) = \sum_{i\geq p;\, j\geq q} m^i m^j = \\ &= \sum_{i\geq p;\, j\geq q} \alpha^i \cdot \alpha^j - \alpha^i \cdot \alpha^{j+1} - \alpha^{i+1} \cdot \alpha^j + \alpha^{i+1} \cdot \alpha^{j+1} = \alpha^p \cdot \alpha^q \end{aligned} \tag{5.40}$$

The theorem follows by using the contour function given by Eq. (5.38) within Theorem 5.3(i). ■

It is important to notice that the α-cuts of $L$ are not necessarily Cartesian products. Also, from Theorem 5.3(ii), one obtains that, if $(\mathcal{Z}_1, z_1)$ is not consistent, then the random relation defined by its contour function only defines an inner bound on the Plausibility of $(\mathcal{Z}_1, z_1)$.

In summary, under the hypothesis of random set independence, the following inclusions hold (refer also to Figure 5.11):

$$\begin{aligned} &(\mathcal{Z}_1, z_1) \subseteq (\mathcal{Z}_3, z_3) \subseteq H \\ &L \subseteq_{\text{weak}} (\mathcal{Z}_3, z_3) \subseteq H \end{aligned} \tag{5.41}$$

However, in general, $L \not\subset_{\text{weak}} (\mathcal{Z}_1, z_1)$ because $L$ is constructed by using $F_1$ and $F_2$, which include the original marginals $\mathcal{F}_1$ and $\mathcal{F}_2$. The Plausibility of $(\mathcal{Z}_1, z_1)$ is bounded from below by its contour function.

**Example 5.7.** Consider the random sets in the counterexample of Section 5.1.4.1 (Table 5.11), whose consonant inclusions are given in Table 5.13, and repeated here in Table 5.18 together with the membership functions of the transformed fuzzy sets $F_i$* calculated using Eq. (5.34) (columns 1 and 4, and

5 and 8). The set of α-values is $M_H = M_{F_1} \cup M_{F_2}$ = {1, 0.800002, 0.300002, 0.200002}, and $\{m^1, m^2, m^3, m^4\}$={0.199998, 0.5, 0.1, 0.200002} The set of focal elements for the including consonant random set is $\mathcal{L}(F_1 \times F_2)$ = $\{ \bar{B}^1$ = [5, 8]×[3, 7], $\bar{B}^2$= [3, 8]×[3, 7], $\bar{B}^3$ =[2, 8]×[2, 7], $\bar{B}^4$= [2, 8]× [1, 8]}. Table 5.19 gives the marginal focal elements together with the combined focal sets $B^k$: notice that several $B^k$s are not in $\mathcal{L}(F_1 \times F_2)$ (e.g., $B^3$), and several $B^k$s coincide because $M_{F_1} \subset M_H$ and $M_{F_2} \subset M_H$ . Table 5.20 gives the coefficients $i$, $j$ calculated with Eq. (5.31) for non-zero entries $w_{ki} = m^i \cdot m^j$, from which Eq. (5.32) can be checked. Also given in Table 5.20 (rightmost column) are the probability assignment of the included random relation ($\mathcal{Z}_3$, $z_3$), and the probability assignment for the including consonant random relation (bottom row); both assignments have been calculated using Eq. (5.4)).

Table 5.21 reports the including fuzzy relation $H$ calculated using Eq. (5.30), or, equivalently, as $H = F_1{*} \times F_2{*}$, where fuzzy sets $F_i$* are given in Table 5.18 (columns 1 and 4, and 5 and 8). The rightmost column in Table 5.21 reports the probability assignment for the including consonant random set equivalent to $H = F_1{*} \times F_2{*}$, which is the same as the bottom row in Table 5.20, as is to be expected by the derivation in Eq. (5.33).

As for inner approximation, Theorem 5.4 applied to point ($s_1$, $s_2$)=(3, 2) yields $\mu_L(3,2) = \mu_{F_1}(3) \cdot \mu_{F_2}(2) = 0.800002 \cdot 0.300002 = 0.2400022$. On the other hand, the smallest indexes $p$, $q$ such that $(2,3) \in A_1^p \times A_2^q$ are $p = 3$ and $q = 2$, and thus:

$$\begin{aligned} \mu_L(2,3) &= Pla(\{s_1, s_2\}) = \sum_{i \geq 2; j \geq 3} \bar{m}^i \bar{m}^j = \\ &= \sum \{ z_3^k : k = 7,8,11,12,15,16 \} = 0.2400022 \end{aligned} \tag{5.42}$$

In order to appreciate outer and inner approximations, let us calculate the bounds on Plausibility and Belief for set $T$=[1, 4]×[2, 6] (Table 5.20). Plausibility and Belief for the original random relation ($\mathcal{Z}_1$, $z_1$) (Table 5.14) are equal to:

$$\begin{aligned} &Bel(T) = z_1(C_8) = 0.03 \\ &Pla(T) = \sum_{i=4}^{9} z_1(C_i) = 0.35+0.05+0.1+0.21+0.03+0.06=0.8, \end{aligned} \tag{5.43}$$

whereas ($\mathcal{Z}_3$, $z_3$) yields (Table 5.19 and Table 5.20):

$$\begin{aligned} &Bel(T) = 0.00 \\ &Pla(T) = \sum_{i=5}^{16} z_3(B^k) = 0.800002 \end{aligned} \tag{5.44}$$

The outer approximation $H$ (Theorem 5.2) gives:

$$\begin{aligned} Bel_H(T) &\overset{Eq.(3.25)}{=} 1-\max_{s\in T^c}\mu_H(s) = 1\text{-}1\text{=}0.00 \\ Pla_H(T) &\overset{Eq.(3.25)}{=} \max_{s\in T}\mu_H(s) = 0.9600008 \end{aligned} \tag{5.45}$$

The inner approximation $L$ (Theorem 5.4) gives:

$$\begin{aligned} Bel_L(T) &\overset{Eq.(3.25)}{=} 1-\max_{s\in T^c}\mu_L(s) = 1-1 = 0.00 \\ Pla_L(T) &\overset{Eq.(3.25)}{=} \max_{s\in T}\mu_L(s) = \max_{s_1\in[1,4]}\mu_{F_1}(s_1)\cdot\max_{s_2\in[2,6]}\mu_{F_2}(s_2) \\ &= 0.800002\cdot 1\text{=}0.800002 \end{aligned} \tag{5.46}$$

These results exemplify the inclusions in Eq. (5.41); inner approximation $L$ turned out to yield the same values for Belief and Probability as $(\mathcal{Z}_3, z_3)$, whereas $H$ yielded a higher value for the Plausibility. $L$ is not an inner inclusion for $(\mathcal{Z}_1, z_1)$ because the Belief value calculated by using $L$ (Eq. (5.46)) is smaller than the relevant value calculated by using $(\mathcal{Z}_1, z_1)$ (Eq. (5.43)).

**Table 5.18** Example 5.7: Marginal consonant random sets, equivalent fuzzy sets, and transformed fuzzy sets

| $\overline{A}_1^i$ | $\overline{m}_1\left(\overline{A}_1^i\right)$ | $\mu_{F_1}\left(\overline{A}_1^i\right)$ | $\mu_{F_1*}\left(\overline{A}_1^i\right)$ | $\overline{A}_2^i$ | $\overline{m}_2\left(\overline{A}_2^i\right)$ | $\mu_{F_2}\left(\overline{A}_2^i\right)$ | $\mu_{F_2*}\left(\overline{A}_2^i\right)$ |
|---|---|---|---|---|---|---|---|
| $\overline{A}_1^1$ = [5, 8] | 0.199998 | 1 | 1 | $\overline{A}_2^1$ = [3, 7] | 0.699998 | 1 | 1 |
| $\overline{A}_1^2$ = [3, 8] | 0.5 | 0.800002 | 0.9600008 | $\overline{A}_2^2$ = [2, 7] | 0.1 | 0.300002 | 0.5100028 |
| $\overline{A}_1^3$ = [2, 8] | 0.300002 | 0.300002 | 0.5100028 | $\overline{A}_2^3$ = [1, 8] | 0.200002 | 0.200002 | 0.3600032 |

**Table 5.19** Example 5.7: Marginal focal elements and combined focal sets $B^k$

| | $A_2^1$ = [3, 7] | $A_1^2$ = [3, 7] | $A_1^3$ = [2, 7] | $A_1^4$ = [1, 8] |
|---|---|---|---|---|
| $A_1^1$ = [5, 8] | $B^1$=[5, 8] ×[3, 7] | $B^2$=[5, 8] ×[3, 7] | $B^3$=[5, 8] ×[2, 7] | $B^4$=[5, 8] ×[1, 8] |
| $A_1^2$ = [3, 8] | $B^5$=[3, 8] ×[3, 7] | $B^6$=[3, 8] ×[3, 7] | $B^7$=[3, 8] ×[2, 7] | $B^8$=[3, 8] ×[1, 8] |
| $A_1^3$ = [2, 8] | $B^9$=[2, 8] ×[3, 7] | $B^{10}$=[2, 8] ×[3, 7] | $B^{11}$=[2, 8] ×[2, 7] | $B^{12}$=[2, 8] ×[1, 8] |
| $A_1^4$ = [2, 8] | $B^{13}$=[2, 8] ×[3, 7] | $B^{14}$=[2, 8] ×[3, 7] | $B^{15}$=[2, 8] ×[2, 7] | $B^{16}$=[2, 8] ×[1, 8] |

**Table 5.20** Example 5.7: matrix **w**. Each non-zero entry gives the coefficients $i$, $j$ of product $m^i \cdot m^j$

| $B^k$ | 1 | 2 | 3 | 4 | $z_3^k$ |
|---|---|---|---|---|---|
| 1 | 1, 1 | | | | 0.0399992 |
| 2 | | 1, 2 | | | 0.099999 |
| 3 | | | 1, 3 | | 0.0199998 |
| 4 | | | | 1, 4 | 0.04 |
| 5 | | 2, 1 | | | 0.099999 |
| 6 | | 2, 2 | | | 0.25 |
| 7 | | | 2, 3 | | 0.05 |
| 8 | | | | 2, 4 | 0.100001 |
| 9 | | | 3, 1 | | 0.0199998 |
| 10 | | | 3, 2 | | 0.05 |
| 11 | | | 3, 3 | | 0.01 |
| 12 | | | | 3, 4 | 0.0200002 |
| 13 | | | | 4, 1 | 0.04 |
| 14 | | | | 4, 2 | 0.100001 |
| 15 | | | | 4, 3 | 0.0200002 |
| 16 | | | | 4, 4 | 0.0400008 |
| $\overline{m}\left(\overline{B}^i\right)$ | **0.0399992** | **0.449998** | **0.1499996** | **0.3600032** | **1** |

**Table 5.21** Example 5.7: Including consonant random relation $H$ (Eq. (5.30))

| $i$ | α-level | $\overline{B}^i = A_1^i \times A_2^i$ | $\overline{m}\left(\overline{B}^i\right) = \alpha^i\text{-}\alpha^{i+1}$ |
|---|---|---|---|
| 1 | 1.0000000 | [5, 8]×[3, 7] | 0.0399992 |
| 2 | 0.9600008 | [3, 8]×[3, 7] | 0.4499980 |
| 3 | 0.5100028 | [2, 8]×[2, 7] | 0.1499996 |
| 4 | 0.3600032 | [2, 8]×[1, 8] | 0.3600032 |

## 5.2 Mappings of Sets/Relations

### *5.2.1 Extension Principle*

Let $\mathcal{F} = \{(A^i, m)$ be a random set on $S$, and let $G$: $S \to Z$ be a single- or multi-valued mapping (e.g., a mathematical model of an engineering system). The information available on $S$ is *extended* to $Z$ by mapping each focal set (together with its probability assignment) to $Z$, so that a new random set, $\mathcal{R} = \left\{\left(R^i, \rho\right)\right\}$, is defined on $Z$. In formulas, the extension principle is:

$$\begin{aligned} R^i &:= G\left(A^i\right) \\ \rho\left(R^i\right) &:= \sum_{j:R^i = G\left(A^j\right)} m\left(A^j\right), \end{aligned} \tag{5.47}$$

which takes into account the possibility that more than one focal set $A^j$ be mapped on the same set $R^i$.

**Example 5.8.** Consider the random set {([1, 3], 0.1), ( [3, 6], 0.3), ([-3, -1], 0.6)}, on $S = \mathbb{R}$, and the function $G$: $\mathbb{R} \to \mathbb{R}$, $a \mapsto a^2$. The new random set $\left\{\left(R^i, \rho\right)\right\}$ is {([1, 9], 0.1), ([9, 36], 0.3), ([1, 9], 0.6)}. Since $A^1$ and $A^3$ both map into [1, 9], the final range is {{[1, 9], 0.7), ([9, 36], 0.3).

Since the extension principle generates a random set on $Z$, one of the strengths of random set theory lies in the fact that the extension principle (5.47) holds for both single-valued mappings and multi-valued mappings. The reason being that a focal set in the range $Z$ can be generated by:

- A focal set mapped by a single-valued mapping.
- A singleton mapped by a multi-valued mapping.
- A focal set mapped by a multi-valued mapping.

Depending on the structure of the focal sets, the extension principle in Eq. (5.47) has several important specializations, which will be explored in the next sections. The following theorems, although originally proven for single-valued mappings ($f$ in the following) are valid for multi-valued mappings ($G$).

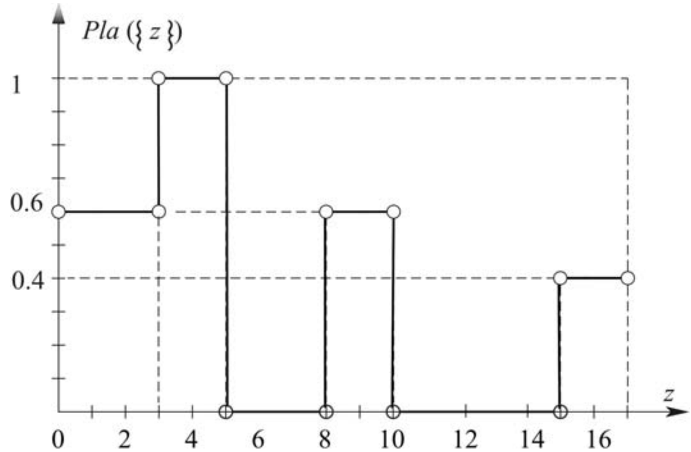



**Fig. 5.12** Example 5.9: Contour function of the random set induced on $Z$

**Example 5.9.** Let us assume: $S_1 = S_2 = \{1, 2, 3\}$; $Z = \mathbb{N}$; $G: S_1 \times S_2 \to \mathbb{N}$, $(a_1, a_2) \mapsto \left[(a_1 \cdot a_2)^2 - 1, (a_1 \cdot a_2)^2 + 1\right]$.

The random set on $S_1 \times S_2$ is assigned as:

$\{(A^1 = \{1, 2, 3\} \times \{1\}, m(A^1) = 0.6);\quad (A^2 = \{2\}\times\{1, 2\},\ m(A^2) = 0.4)\}$.

The random set $\mathcal{R}$ on $Z$ is obtained as:

$\{(R^1 = [0, 2]\cup[3, 5]\cup[8, 10], \rho(R^1) = 0.6);\ (R^2 = [3, 5]\cup[15, 17],\ \rho(R^2) = 0.4)\}$.

Figure 5.12 shows the contour function of $\mathcal{R}$.

#### 5.2.1.1 Consonant Random Relation

Consonant random relations were introduced in Section 4.1, Figure 4.2, page 106. The following theorem gives a quick way to calculate the image of consonant random relations, and clarifies that only one version of fuzzy extension principles is compatible with random set theory.

**Theorem 5.5.** *Given a consonant random relation $\mathcal{F}$ equivalent to a fuzzy relation F, its extension $\mathcal{R}$ obtained by using Eq.* (5.47) *is consonant and is equivalent to the fuzzy set R defined by:*

$$\forall z, \mu_R(z) = \begin{cases} \sup\left\{\mu_F(s_1, s_2) : (s_1, s_2) \in f^{-1}(z)\right\} \\ 0 \text{ if } f^{-1}(z) = \varnothing \end{cases} \tag{5.48}$$

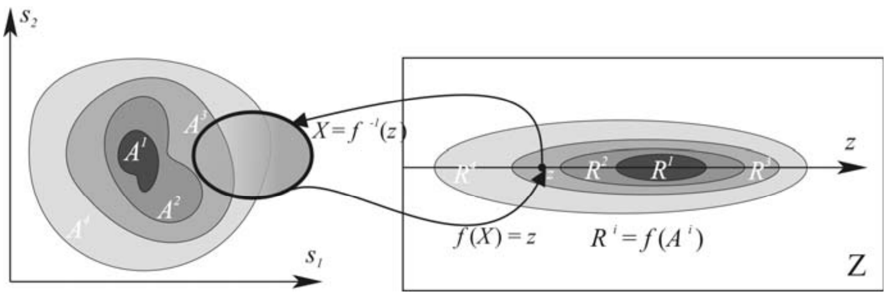



**Fig. 5.13** Mapping of a consonant random relation

*Proof.* To prove Eq. (5.48) in a general setting (Dubois and Prade 1991), let $X = f^{-1}(z) \neq \varnothing$ (Figure 5.13). Let us consider consonance first. Since $A^1 \subset A^2 \subset \ldots \subset A^n$ and $A^i \subset A^j \Rightarrow f(A^i) \subseteq f(A^j)$, $f(A^1) \subseteq f(A^2) \subseteq \ldots \subseteq f(A^n)$, hence the range of focal elements is nested. Now, let $R$ be the fuzzy relation equivalent to $\mathcal{R}$ per Eq. (4.10). The membership function of any point $z$ is:

$$\mu_R(z) \overset{Eq.(4.10)}{=} \sum_{R^i : z \in R^i} \rho(R^i) \overset{Eq.(5.47)}{=} \sum_{A^i : z \in f(A^i)} m(A^i) = \sum\left\{m(A^i) : \exists (s_1, s_2) \in A^i \cap X\right\} \tag{5.49}$$

In the last passage, the membership calculation has been "pulled back" onto the initial $S_1 \times S_2$ space. Because $\{A^i\}$ is consonant, let $j = \min\left\{i : \exists (s_1^*, s_2^*) \in A^i \cap X\right\}$, then $(s_1^*, s_2^*) \in A^i \ \forall i \geq j$, and

$$\mu_R(z) = \sum_{A^i : (s_1^*, s_2^*) \in A^i} m(A^i) = \mu_F(s_1^*, s_2^*) \tag{5.50}$$

with the convention $\mu_R(z) = 0$ if $\left\{i : \exists (s_1^*, s_2^*) \in A^i \cap X\right\} = \varnothing$. Clearly $(s_1^*, s_2^*)$ maximizes $\mu_F$ over $X$. ■

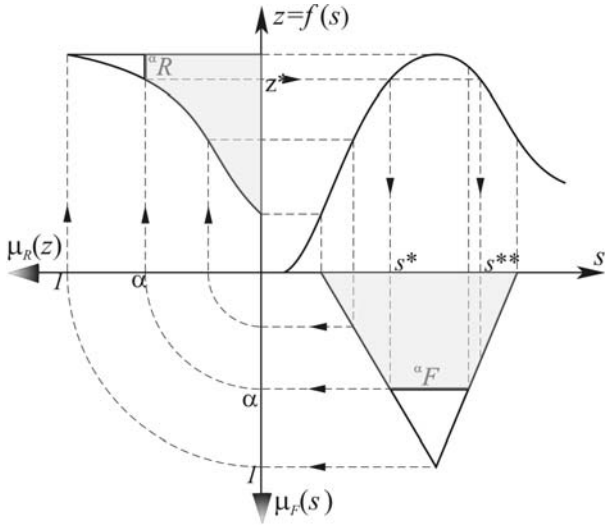



**Fig. 5.14** 1-D example of extension principle for consonant random relation defined on the real line

Figure 5.14 shows an example where $S = \mathbb{R}$: the starting point is a value $z^* \in Z = \mathbb{R}$, whose inverse through $f$ is $\{s^*, s^{**}\}$. Finally, $\mu_R(z^*) = \max\{\mu_F(s^*), \mu_F(s^{**})\} = \mu_F(s^*)$. The proof of Theorem 5.5 also shows the following Nguyen's theorem (Nguyen 1978; Klir and Yuan 1995):

**Theorem 5.6**. *If the supremum in* (5.48) *is attained for a pair* $(s_1^*, s_2^*)$, *then*

$$^{\alpha}f(F) = f(^{\alpha}F) \tag{5.51}$$

*otherwise:*

$$^{\alpha+}f(F) = f(^{\alpha+}F) \tag{5.52}$$

*where* $^{\alpha}F = \{s \in S : \mu_F(s) \geq \alpha\}$ *is the* $\alpha$*-cut of* $F$ *(Eq.* (3.27)*) and* $^{\alpha+}F = \{s \in S : \mu_F(s) > \alpha\}$ *is the strong* $\alpha$*-cut of* $F$.

By virtue of this theorem, the range of a fuzzy set $F$ is obtained by first mapping the α-cuts of $F$ into $Z$, and then by applying the decomposition theorem (Eq. (3.30)). In Figure 5.14, the starting point is now α-cut $^{\alpha}F$, which is mapped through $f$ onto $^{\alpha}R$ using a path exactly opposite to the one indicated in Theorem 5.5. When compared to Eq. (5.48), this method is computationally very efficient because it does not require inversion of function *f*; computational savings are especially evident when the entire membership function of *R* is required, as shown in the following examples. When the consonant random relation has an infinite number of focal sets, one can either:

- Perform simulations by following the algorithms given by Alvarez (2006).
- Include the actual random relations by using a random relation with a finite number of α-cuts (Tonon 2004, Hall 2004, Tonon 2008).

**Example 5.10.** Consider the 1-D consonant random set {([1, 1.5], 0.1), ([0, 2], 0.3), ([-1, 2], 0.2), ([-3, 3], 0.4)} on $S = \mathbb{R}$, and the function *f*: $\mathbb{R} \to Z = \mathbb{R}$, $a \mapsto a^2$. Suppose that one is interested in the membership value of $z$ = 3.5. By using Eq. (5.47), its range random set, $\left\{\left(R^i, \rho^i\right)\right\}$, is {([1, 2.25], 0.1),([0, 4], 0.3), ([0, 4], 0.2), ([0, 9], 0.4)}. Since $A^2$ and $A^3$ both map into [0, 4], the final range is {([1, 2.25], 0.1), ([0, 4], 0.5), ([0, 9], 0.4)}. Then, by using Eq. (3.24):

$$\mu_R(3.5) = Pla\left(\{3.5\}\right) = \rho\left(R_2\right) + \rho\left(R_3\right) = 0.5 + 0.4 = 0.9 \tag{5.53}$$

This is equivalent to using Nguyen's theorem because the α-cuts of *F* are: ${}^{0.4}F = A^4$; ${}^{0.6}F = A^3$; ${}^{0.9}F = A^2$; ${}^{1.0}F = A^1$. Once these sets have been mapped onto *Z* as above, the result is obtained by using Eq. (3.30):

$$\begin{aligned} \mu_R(3.5) &= \max_{\alpha}\left\{\min\left(\alpha, \chi_{{}^{\alpha}A}(3.5)\right)\right\} \\ &= \max\left\{\min(0.4,1), \min(0.6,1), \min(0.9,1), \min(1.0,0)\right\} = 0.9 \end{aligned} \tag{5.54}$$

Compare now with the direct application of Eq. (5.48). One has to calculate the inverse of *f*: $\pm\sqrt{3.5} \simeq \pm 1.871$, and then the superior of the membership functions for these values:

$$\begin{aligned} \mu_R(3.5) &= \sup\left\{\mu_F(-0.1871), \mu_F(0.1871)\right\} \\ &= \sup\left\{Pla\left(\{-0.1871\}\right), Pla\left(\{0.1871\}\right)\right\} = \sup\{0.4, 0.9\} = 0.9 \end{aligned} \tag{5.55}$$

**Example 5.11.** Let us consider a "consonant version" of the random set used in Example 5.9. Let us assume: $S_1 = S_2 = \{1, 2, 3\}$; $Z = \mathbb{N}$; $G$: $S_1 \times S_2 \longrightarrow \mathbb{N}$,

$(a_1, a_2) \mapsto \left[ INT(a_1 \cdot a_2)^2 - 1, INT(a_1 \cdot a_2)^2 + 1 \right]$. The consonant random set on $S_1 \times S_2$ is given as:

$\{(A^1 = \{1, 2, 3\} \text{ x } \{1, 2\}, m(A^1)= 0.6), \quad (A^2 = \{2\} \text{ x } \{1, 2\}, \ m(A^2) = 0.4)\}$.

One can immediately calculate the membership function of the normalized fuzzy relation on $S_1 \times S_2$ corresponding to the consonant random set:

$\mu(1,1)= m(A^1)= 0.6;$
$\mu(1,2)= m(A^1)= 0.6;$
$\mu(2,1) = m(A^1) + m(A^2) = 0.6 + 0.4 = 1;$
$\mu(2,2) = m(A^1) + m(A^2) = 0.6 + 0.4 = 1;$
$\mu(3,1) = m(A^1) = 0.6;$
$\mu(3,2) = m(A^1) = 0.6.$

The α-cuts coincide with the focal sets of the consonant random set: ${}^{1.0}F = A^2$, ${}^{0.6}F = A^1$. The fuzzy set induced by $f$ onto $Z$ is defined by its α-cuts (Eq. (5.51)):

$$\begin{aligned} {}^{1.0}R &= f({}^{1}F) = [3,5]\cup[15,17]; \\ {}^{0.6}R &= f({}^{0.6}F) = [0,2]\cup[3,5]\cup[3,5]\cup[15,17]\cup[8,10]\cup[35,37]= \\ &\qquad [0,2]\cup[3,5]\cup[8,10]\cup[15,17]\cup[35,37] \end{aligned} \tag{5.56}$$

Again, this result coincides with (but is much more efficient than) the one obtainable by applying (5.48). For example, for $z = 10$, we have:

$\mu_R(10) = \max\{\mu_F(3,1), \mu_F(1,3)\} = \max\{0.6,0\} = 0.6$; indeed, $z = 10 \in {}^{0.6}R$.

For $z = 5$:

$\mu_R(5) = \max\{\mu_F(1,2), \mu_F(2,1)\} = \max\{0.6,1\} = 1$; indeed, $z = 5 \in {}^{1.0}R$.

### 5.2.1.2 Consonant Random Cartesian Product

Consonant random Cartesian products were introduced in Section 4.1, Figure 4.3.

**Theorem 5.7**. *If $\mathcal{F}$ is a consonant random Cartesian product (Section* 4.3.5*) equivalent to a decomposable fuzzy relation $F = F_1 \times F_2 \times \ldots \times F_r$, by taking into account Eq. (4.115), Eq. (5.48) can be written as follows:*

$$\forall z, \mu_R(z) = \begin{cases} \sup\left\{\min\left\{\mu_F(s_1), \mu_F(s_2), \ldots, \mu_F(s_r)\right\} : (s_1, s_2, \ldots, s_r) \in f^{-1}(z)\right\} \\ 0 \text{ if } f^{-1}(z) = \varnothing \end{cases} \tag{5.57}$$

This is Zadeh's original extension principle (Zadeh 1975): among several extension principles put forward in fuzzy set theory, (5.57) is thus the only one that is consistent with random set theory. This conclusion is particularly important for establishing a framework for a coherent treatment of uncertainty in the analysis of engineering systems as propounded in this book. In Chapter 6, this issue will be furtherly discussed in the context of a theory of approximate reasoning.

Nguyen's theorem now specializes as follows.

**Theorem 5.8**. *If the supremum in Eq.* (5.57) *is attained for at least an r-tuple* $(s_1^*, s_2^*, \ldots, s_r^*)$, *then*

$$^{\alpha}f(F) = f\left(^{\alpha}F_1, ^{\alpha}F_2, \ldots, ^{\alpha}F_r\right) \tag{5.58}$$

*otherwise:*

$$^{\alpha+}f(F) = f\left(^{\alpha+}F_1, ^{\alpha+}F_2, \ldots, ^{\alpha+}F_r\right) \tag{5.59}$$

When $S_i = \mathbb{R}$, $^{\alpha}F_i$ are real intervals and $^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r$, an $r$-dimensional box in $\mathbb{R}^r$ with $2^r$ vertices. In this case, Eqs. (5.58) and (5.59) are interval analysis problems for which very powerful algorithms have been (and are being) developed (Moore 1966; Moore 1979; Alefeld and Herzberger 1983; Ratschek and Rokne 1984; Ratschek and Rokne 1988; Neumaier 1990; Corliss 1999; Jaulin, Kieffer et al. 2001; Hansen and Walster 2003; Berz 2007; Kreinovich, Berleant et al. 2007; Mulhanna and Mullen 2007; Nesterov 2007; Neumaier 2007). These algorithms significantly improve on the exponential number of function calls entailed by the Vertex Method (Theorem 5.9 below).

**Theorem 5.9** (Dong, Chiang et al. 1987; Dong and Shah 1987; Dong and Wong 1987). *If the maximum and minimum of f in Eq. (4.140) is not reached in the interior of* $^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r$ *nor at its edges, then*

$$^{\alpha+}f(F) = f\left(^{\alpha+}F_1, ^{\alpha+}F_2, \ldots, ^{\alpha+}F_r\right) = [\min\left\{f\left(c_j\right)\right\}, \max\left\{f\left(c_j\right)\right\}] \tag{5.60}$$

*where* $c_j$, $j=1,\ldots,2^r$ *are the vertices of* $^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r$.

**Theorem 5.10** (Tonon and Bernardini 1998). *If* $f(s_1,\ldots, s_r) \in C^1\left(\mathbb{R}^r\right)$ *and is strictly monotone with respect to each variable* $s_i$, *then*

$$\begin{aligned} &\exists! c_j{}^{*}: f(c_j{}^{*}) = \max\left\{f\left(c_j\right)\right\} \\ &\exists! c_j{}^{**}: f(c_j{}^{**}) = \min\left\{f\left(c_j\right)\right\} \end{aligned} \tag{5.61}$$

*where* $c_j$ *,* $j$=1,…,$2^r$ *are the vertices of* ${}^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r$ .

*Proof.* It stems from monotonicity assumption that either hypothesis a) or b) is true:

$$\begin{aligned} &\text{a)} \quad \left.\frac{\partial f}{\partial s_i}\right|_{(\mathbf{s})} > 0 \quad \forall \mathbf{s} \in {}^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r\,, \\ &\text{b)} \quad \left.\frac{\partial f}{\partial s_i}\right|_{(\mathbf{s})} < 0 \quad \forall \mathbf{s} \in {}^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r \end{aligned} \tag{5.62}$$

Let $\overline{\mathbf{s}}$ be a maximizer for $f$ in ${}^{\alpha}F$ . Then, if Eq. (5.62.a) is true, we have: $\overline{s}_i = \left({}^{\alpha}F_i\right)^U$ , otherwise $\overline{s}_i = \left({}^{\alpha}F_i\right)^L$. ■

Theorem 5.10 and its proof indicate that:

1) Vertex $c_j{}^{*}$ that maximizes $f$ in ${}^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r$ is the combination of the right (left, resp.) extremes of the variables with respect to which $f$ is increasing (decreasing, resp.).
2) Vertex $c_j{}^{**}$ that minimizes $f$ in ${}^{\alpha}F = {}^{\alpha}F_1 \times {}^{\alpha}F_2 \times \ldots \times {}^{\alpha}F_r$ is the combination of the left (right, resp.) extremes of the variables with respect to which $f$ is increasing (decreasing, resp.).
3) When a new α-level is chosen, $c_j{}^{*}$ and $c_j{}^{**}$ are given by the same combination of parameters as in points (1) and (2).
4) If (in points (1) and (2)) the increasing or decreasing variables are not known or difficult to determine, evaluate $f$ at all vertices $c_j$ , $j$=1,…,$2^r$, and determine the unique combinations that minimize and maximize $f$. Use point (3) for all other α-levels.

This algorithm allows one to identify the critical vertices either *a priori* (points (1) and (2)) or with just one application of the vertex method. Each α-cut is then evaluated by using just two function calls, whereas the vertex method requires $2^r$ function calls, which increases significantly with the number of variables (for $r$ = 10, $2^{10}$ = 1,064). Indeed, when several α-cuts must be computed, computational savings are significant: e.g., for 11 $\alpha$-cuts, 22 vs. 11,264 function calls are necessary. When the function $f$ is the output of a complex model (e.g., BEM, FEM, FDM, DEM) that takes one day to run, the above algorithm may make computations feasible.

Additionally, the above algorithm allows one to use existing numerical codes as "black boxes" that simply calculate function $f$ without need to modify them; these concepts are illustrated in the next example.

**Example 5.12.** Let us consider the non-linear elastic system shown in Figure 5.15 composed of a two-span linearly elastic beam, with Young modulus E, restrained by two rigid supports and a non-linear elastic spring whose stiffness monotonically increases with the displacements. The load intensities $P$ and $q$ are restricted by the fuzzy sets shown in Figure 5.16, and we want to determine the displacement of the supporting spring.

The displacement, $\delta_C$, is related to the external forces $P$ and $q$ and to the spring reaction $Y$ by means of a linear deterministic relation, whose coefficients depend on the elastic and geometrical properties of the beams ( $l_1 = 5.2$ m , $l_2 = 4.3$ m ; $E = 27$ GPa ; $J_1 = 0.7^3 \cdot 0.4 /12$ m$^4$ ; $J_2 = 0.4^3 \cdot 0.4 /12$ m$^4$):

$$\begin{aligned} \delta_C &= -D_1 \cdot P + D_2 \cdot q - D_3 \cdot Y = h(P,q) - D_3 \cdot Y = \\ &= -(0.02354 \text{ mm/kN})\cdot P + (0.9651 \text{ mm/(kN/m)})\cdot q \; - \\ &\quad - (0.5639 \text{ mm/kN})\cdot Y \quad (\text{mm}) \end{aligned} \tag{5.63}$$

where:

$$D_1 = \frac{l_1^2 \cdot l_2}{16 \cdot E \cdot J_1}; \quad D_2 = \frac{l_2^4}{8 \cdot E \cdot J_2} + \frac{l_1 \cdot l_2^3}{6 \cdot E \cdot J_1}; \quad D_3 = \frac{l_2^3}{3 \cdot E \cdot J_2} + \frac{l_1 \cdot l_2^2}{3 \cdot E \cdot J_1} \tag{5.64}$$

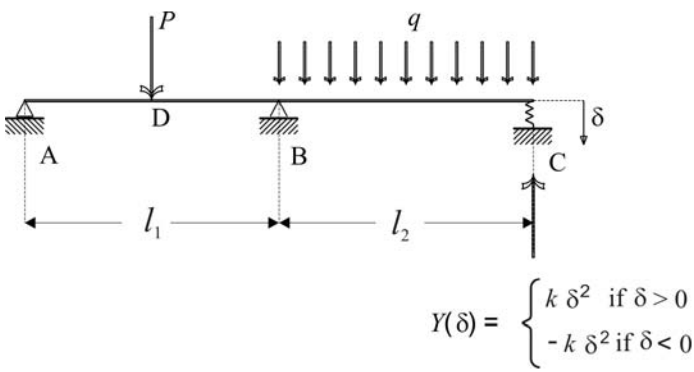



**Fig. 5.15** Example 5.12: A two-span linearly elastic beam restrained by two rigid supports and a non-linear elastic spring

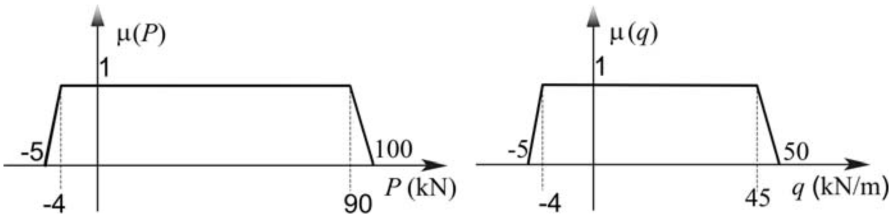



**Fig. 5.16** Example 5.12: Fuzzy sets of the concentrated load and of the uniformly distributed load q

The condition of compatibility between the deformations of the beam and the spring gives the solution, which is the sought function of *P* and *q*:

$$Y = \begin{cases} \left(b\cdot h + 1 - \sqrt{1 + 2\cdot b\cdot h}\right)/\left(b\cdot D_3\right) & \text{if } h \geq 0 \\ \left(b\cdot h - 1 + \sqrt{1 - 2\cdot b\cdot h}\right)/\left(b\cdot D_3\right) & \text{if } h < 0 \end{cases};\ b=2\,k\cdot D_3 \tag{5.65}$$

**Table 5.22** Interval analysis of the reaction *Y* for $\alpha = 0$

| Index *j* | *q* | *P* | *f*(*q*, *P*) | *Y* | Critical Indexes |
|---|---|---|---|---|---|
| 1 | - 5 | - 5 | -4.708 | | |
| 2 | 50 | - 5 | 48.375 | 82.900 | Maximizing |
| 3 | - 5 | 100 | - 7.180 | - 11.651 | Minimizing |
| 4 | 50 | 100 | 45.903 | | |

**Table 5.23** α-cuts of the reaction Y

| α-cut | Min(*Y*) | Max(*Y*) |
|---|---|---|
| 0 | - 11.651 | 82.900 |
| 0.1 | - 11.448 | 82.055 |
| 0.2 | - 11.245 | 81.210 |
| 0.3 | - 11.041 | 80.365 |
| 0.4 | - 10.838 | 79.520 |
| 0.5 | - 10.635 | 78.675 |
| 0.6 | - 10.432 | 77.830 |
| 0.7 | - 10.229 | 76.986 |
| 0.8 | - 10.027 | 76.141 |
| 0.9 | - 9.824 | 75.296 |
| 1 | -9.621 | 74.452 |

The expression for $Y$ is simple enough that one can take partial derivatives. Notice that $Y$ is a monotonically increasing function of $h$, which is decreasing with $P$ and increasing with $q$. Therefore, the maximizing interval extremes are the left extreme for $P$ and the right extreme for $q$; the minimizing interval extremes are the right extreme for $P$ and the left extreme for $q$.

For completeness, Table 5.22 gives the values of $h$ corresponding to the $2^2 = 4$ vertices for $\alpha = 0$ (with $k = 31.4$ kN / mm$^2$), which confirm the above result on the critical vertices. The fuzzy set of $Y$ reconstructed by its $\alpha$-cuts is displayed in Table 5.23.

### 5.2.2 Monotonicity of Operations on Random Relations

In Section 5.1, the concept of weak inclusion was introduced. The good news is that the inclusion is preserved by mappings; more precisely:

**Theorem 5.11**. *Let* $\mathcal{F}$ and $\overline{\mathcal{F}}$ *be two random relations defined on S with compatible sets of probability measures* $\Psi$ *and* $\overline{\Psi}$, *respectively, such that* $\Psi \subseteq \overline{\Psi}$. *Let f: S* $\to$ *Z, and let* $\mathcal{R}$ *and* $\overline{\mathcal{R}}$ *be the ranges of* $\mathcal{F}$ and $\overline{\mathcal{F}}$, *respectively, calculated using Eq.* (5.47). *Then*

$$\begin{aligned}&(i)\mathcal{F} \subseteq_{weak} \overline{\mathcal{F}} \Rightarrow \left[Bel_Z(U), Pla_Z(U)\right] \subseteq \left[\overline{Bel_Z}(U), \overline{Pla_Z}(U)\right], \forall U \subseteq Z\\&(ii)\Psi \subseteq \overline{\Psi} \Rightarrow \Psi_f \subseteq \overline{\Psi}_f\end{aligned} \tag{5.66}$$

*where* $\Psi_f$ *and* $\overline{\Psi}_f$ *are the sets of probability measures compatible with* $\mathcal{R}$ *and* $\overline{\mathcal{R}}$, *respectively; subscript Z denotes the underlying set where Belief and Plausibility are defined.*

*Proof*

(*i*) Since $B \subseteq C \Rightarrow f(B) \subseteq f(C)$:

$$\begin{aligned}Bel_Z(U) &\overset{Eq.(3.3)}{=} \sum_{R^i:R^i \subseteq U} \rho^i \overset{Eq.(5.47)}{=} \sum_{R^i:R^i \subseteq U} \sum_{A^j:R^i = f(A^j)} m^j\\&= \sum_{A^j:f(A^j)\subseteq U} m^j \overset{Eq.(3.3)}{=} Bel\left(f^{-1}(U)\right)\end{aligned} \tag{5.67}$$

From Eq. (5.1), $Bel\left(f^{-1}(U)\right) \geq \overline{Bel}\left(f^{-1}(U)\right)$. Likewise for the Plausibility.

(*ii*) Use (*i*) and apply Eq. (5.1) from right to left to $Bel_Z$ and $Pla_Z$. ■

Likewise, strong inclusion is preserved by mappings, and the proof of the following theorem also says how to determine the inclusion matrix:

**Theorem 5.12** (Dubois and Prade 1991). *Let* $\mathcal{F} \subseteq \overline{\mathcal{F}}$ *be two random relations defined on S, and let f: S* $\to$ *Z. Let* $\mathcal{R}$ *and* $\overline{\mathcal{R}}$ *be the ranges of* $\mathcal{F}$ and $\overline{\mathcal{F}}$ *, respectively, calculated using Eq.* (5.47). *Then*

$$\mathcal{F} \subseteq \overline{\mathcal{F}} \Rightarrow \mathcal{R} \subseteq \overline{\mathcal{R}} \tag{5.68}$$

*Proof.* (*i*) By point (*i*) in Eq. (5.4)), $\forall A^i \ \ \exists \, \overline{A}^j : A^i \subseteq \overline{A}^j$. By Eq. (5.47a), each focal set $R^k$ is the image of at least one focal set $A^i$. Let $A^i : R^k = f(A^i)$; since $A^i \subseteq \overline{A}^j \Rightarrow f\left(A^i\right) \subseteq f\left(\overline{A}^j\right)$, $\forall R^k \ \ \exists \, \overline{A}^j : R^k \subseteq f\left(\overline{A}^j\right)$.

(*ii*) Similar to (*i*).

Let us define the following inclusion matrix:

$$\omega_{kl} := \sum \{ w_{ij} : R^k = f\left(A^i\right), \overline{R}^l = f\left(\overline{A}^j\right), A^i \subseteq \overline{A}^j \} \ . \tag{5.69}$$

Recall that

$$A_i \supset \overline{A}_j \Rightarrow w_{ij} = 0 \, . \tag{5.70}$$

One has:

$$\begin{aligned}
\sum_{l:R^k \subseteq \overline{R}^l} \omega_{kl} &\overset{Eq.(5.4)}{=} \sum_{l:R^k \subseteq \overline{R}^l} \sum \{ w_{ij} : R^k = f\left(A^i\right), \overline{R}^l = f\left(\overline{A}^j\right), A^i \subseteq \overline{A}^j \} \\
&= \sum_{l:R^k \subseteq \overline{R}^l} \sum_{A^i:R^k = f\left(A^i\right)} \{ w_{ij} : R^k = f\left(A^i\right), \overline{R}^l = f\left(\overline{A}^j\right), A^i \subseteq \overline{A}^j \} \\
&= \sum_{A^i:R^k = f\left(A^i\right)} \sum_{l:R^k \subseteq \overline{R}^l} \{ w_{ij} : R^k = f\left(A^i\right), \overline{R}^l = f\left(\overline{A}^j\right), A^i \subseteq \overline{A}^j \} \\
&\overset{Eq.(5.4)}{=} \sum_{A^i:R^k = f\left(A^i\right)} \sum_{j:A^i \subseteq \overline{A}^j} w_{ij} = \sum_{A^i:R^k = f\left(A^i\right)} m\left(A^i\right) \overset{Eq.(5.47b)}{=} \rho\left(R^k\right)
\end{aligned} \tag{5.71}$$

Likewise for $\overline{\rho}\left(\overline{R}^k\right)$. ■

Since inclusion is a special case of weak inclusion, Theorem 5.11 specializes as follows:

**Theorem 5.13** (Dubois and Prade 1991). *Let $\mathcal{F} \subseteq \overline{\mathcal{F}}$ be two random relations defined on S, and let f: $S \rightarrow Z$. Let $\mathcal{R}$ and $\overline{\mathcal{R}}$ be the ranges of $\mathcal{F}$* and $\overline{\mathcal{F}}$, respectively, *calculated using Eq.* (5.47). *Then*

$$\left[Bel(U), Pla(U)\right] \subseteq \left[\overline{Bel}(U), \overline{Pla}(U)\right] \forall U \subseteq Z \tag{5.72}$$

In other words, one may find a random relation $\overline{\mathcal{F}}$ that includes another random relation $\mathcal{F}$ by using the algorithms in Section 5.1.3. By mapping $\overline{\mathcal{F}}$, one calculates probability bounds that contain the probability bounds that can be calculated by using $\mathcal{F}$. Major computational savings are achieved if computations performed on $\overline{\mathcal{F}}$ are much easier and/or faster and the including bounds are tight enough to make a decision.

**Example 5.13.** Consider again the random sets on $S = \mathbb{R}$ as in Example 5.3:

$\mathcal{F}$ = {([5, 8], 0.2), ([3, 7], 0.5), ([2, 4], 0.3)};
$\overline{\mathcal{F}}$ ={([5, 8], 0.200002), ([3, 8], 0.5), ([2, 8], 0.288888)},

where β = $10^{-6}$. Let *f*: $x \mapsto x \cdot e^{1/x}$. Focal sets are mapped by making use of Theorem 5.10:

$\mathcal{R}$ = {([6.10702, 9.06520], 0.2), ([4.18685, 8.07496], 0.5), ([3.29745, 5.13620], 0.3)},
$\overline{\mathcal{R}}$ = {([6.10701, 9.06520], 0.188888), ([4.18685, 8.07496], 0.5), ([3.29745, 5.136120], 0.300002)},

where outer rounding has been applied to the focal sets in order to ensure containment when using floating-point numbers. The probability bounds of $U$ = [3.2, 5.6] are:

$$Bel(U) = \rho(R^3) = 0.3;\ Pla(U) = \rho(R^2) + \rho(R^3) = 0.8$$

$$\overline{Bel}(U) = 0;\ \overline{Pla}(U) = \bar{\rho}(R^2) + \bar{\rho}(R^3) = 0.800002$$

## 5.3 Conclusions

The notion of weak inclusions between two random sets relies on the inclusions of the sets of compatible distributions, and it is equivalent to the inclusion of Belief and Plausibility bounds for all subsets in $S$. The notion of strong inclusion was introduced in order to: determine if two random sets are included one into the other by directly operating on the focal elements and probability assignements; find an including random set when a random set is assigned. By resorting to the reservoir-bathtub analogy, it has been shown that strong inclusion implies weak inclusion, but not vice versa.

Since computations with consonant random sets (and associated fuzzy sets) are very fast, four different strategies have been introduced to include a general random set into a consonant one. For ordered sets $S$, one may either choose to minimize the discrepancy in $F_{UPP}$ or $F_{LOW}$; the remaining two strategies are applicable if equal weight should be given to preserving $F_{UPP}$ or $F_{LOW}$, and if one wants to minimize the cardinality of the consonant inclusion, respectively.

When $\mathcal{F}_1$ and $\mathcal{F}_2$ are assigned on $S_1$ and $S_2$, respectively, one may determine two consonant inclusions of theirs: $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$. The random relation obtained from $\mathcal{F}_1$ and $\mathcal{F}_2$ under the hypothesis of random set independence is not included in the fuzzy Cartesian product on $S_1 \times S_2$ obtained from $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$, but it is included in the random relation on $S_1 \times S_2$ obtained from $\overline{\mathcal{F}}_1$ and $\overline{\mathcal{F}}_2$ under the hypothesis of random set independence. The later random relation, however, is not necessarily consonant, but there is a minimal outer approximation to it, and a weak inner approximation to it.

The information conveyed by random set or relation on $S$ may be extended to the range of a mapping defined on $S$. This mapping may be the mathematical model of an engineering system. Since the extension is valid for both single-valued and multi-valued mappings, the principle allows one to extend through a multi-valued mapping information given as a probability distribution on $S$, which is an unresolved problem in the theory of probability. The extension principle for random sets has two main specializations:

1) When the random relation on $S$ is consonant (fuzzy relation), its image is obtained by first mapping the $\alpha$-cuts of the fuzzy relation into $Z$, and then by applying the decomposition theorem This method is computationally very efficient because it does not require mapping inversion.

2) When the random relation on $S$ is a consonant random Cartesian product, the specialization of the principle coincides with Zadeh's original extension principle, which is the only one (among several extension principles put forward in fuzzy set theory) that is consistent with random set theory, and thus with a generalization of probability theory to accommodate imprecise information.

Both weak inclusion and strong inclusion are preserved by mappings (monotonicity), which has major practical consequences. Indeed, one may first find a random relation $\overline{\mathcal{F}}$ that includes another random relation $\mathcal{F}$ by using the algorithms described above. By mapping $\overline{\mathcal{F}}$, one may calculate probability bounds that contain the probability bounds that can be calculated by using $\mathcal{F}$. Major computational savings are achieved if computations performed on $\overline{\mathcal{F}}$ are much easier and/or faster and the including bounds are tight enough to make a decision.

# Chapter 6
# Approximate Reasoning

The theory of probability provides no straightforward answer to the problem of combining two probability measures on the same space, because it is not a problem of mathematics, but of judgment. As a consequence, random set theory too does not provide any prescriptive method for combining two random sets. In Section 6.1, some possible ways of combining or updating information on the same space are reviewed, distinguishing between models that stress the agreement among given bodies of information, and those which stress the conflict. We highlight and discuss their applicability limits (Section 6.2) and introduce the available answers within evidence theory (Section 6.3) and fuzzy set theory (Section 6.4). Analogies and extensions are underlined with both probabilistic procedures and fuzzy set operations.

Finally it will be shown in Section 6.5 that fuzzy logic gives powerful, simple and robust procedures in the field of pattern recognition, clustering and optimal choices in decision making.

## 6.1 The Basic Problem

The following is a basic problem in any theoretical model of uncertainty and in any consistent theory of information. Suppose:

(i) a certain level of information be known or assumed about a system (premise $A$) AND
(ii) further information be obtained independently (i.e. without utilizing previous information) (premise $B$) ;

how can we combine the two bodies of knowledge in order to achieve an updated, *a posteriori* description $C = A \circ B$ embracing all available information?

The solution to this problem is an essential one if for example we want to:

(i) construct rule-based expert systems, i.e. expert systems based on deductive logic;
(ii) extend classic deductive reasoning rules (Woods and Gabbay 2004) such as *modus ponens*, *modus tollens* and hypothetical syllogism to uncertain information;
(iii) to construct mathematical models of approximate reasoning.

The solutions to date pertaining to the classical set theory of uncertainty are examined in Section 6.1.1 and those pertaining to the probability theory in Section 6.1.2 and 6.1.3.

### *6.1.1 Combination and Updating within Set Theory*

The answer to the posed problem is quite simple if sought in the field classical set theory. See for example (Lipschutz 1964).

Consider the case in which two sources of information are available about the same variable (in, say, space $S$). One source (the *a priori* information) says that the variable is restricted by a set $A$, whereas another source says that the variable is restricted by a set $B$. Let $s \in S$, and let $\chi_A(s)$ , $\chi_B(s)$ be the characteristic functions (or the indicators $I_A$ , $I_B$) of $A$ and $B$ (i.e. $\chi_A(s) = I_A(s) = 1$ if $s \in A$, $\chi_A(s) = I_A(s) = 0$ if $s \notin A$).

#### 6.1.1.1 Intersection

If $A$ AND $B$ provide conditions which *necessarily* are to be satisfied by their combination, *a posteriori* uncertainty is described by their intersection (see Figure 6.1), or equivalent combination of their characteristic functions[N 6-1]:

$$\begin{aligned} &C = A \cap B \\ &\chi_C(s) = \min\left(\chi_A(s), \chi_B(s)\right) \end{aligned} \tag{6.1}$$

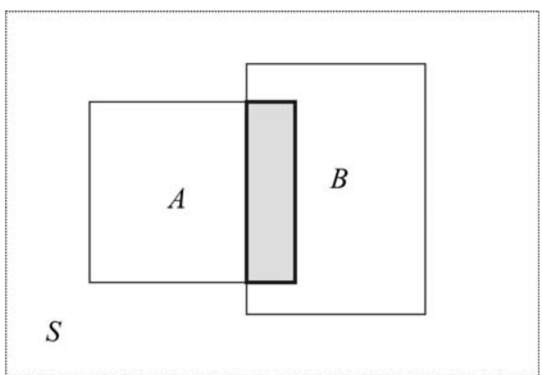



**Fig. 6.1** Combination by intersection

The combination of the two bodies of knowledge therefore reduces uncertainty, whenever $A$ and $B$ do not coincide, and this combination leads to a total loss of information when $A$, $B$ are totally conflicting (i.e. $A \cap B = \emptyset$). When the conflict is partial ($A \cap B \neq \emptyset$), the rule works very well, and decreases the uncertainty for the decision-maker; but this decrease could be unjustified and unrealistic if the sources of information are not reliable.

Intersection can be used to combine information on spaces of different orders. For example a subset of a two-dimensional space $S = S_1 \times S_2$ is a deterministic relation $R$, i.e. a pair of point-valued or more generally set valued relations $G$, $G^{-1}$ between points $s_1$ on $S_1$ (r. $s_2$ on $S_2$) and set $G(s_1)$ (r. $G^{-1}(s_2)$) on $S_2$ (r. $S_1$ ).

We consider here finite spaces with singletons ($s_1{}^1$, … $s_1{}^i$, …. $s_1{}^n$) and ($s_2{}^1$, … $s_2{}^j$, …. $s_2{}^m$) respectively. The relation $R$ is completely defined by the $n$ x $m$ matrix of the values of the characteristic function:

$$\chi_R(s_1^i, s_2^j) = \left\langle \begin{array}{l} 1 \quad \text{if } P_{1,2}\left(s_1^i, s_2^j\right) \in R; s_2^j \in G\left(s_1^i\right); s_1^i \in G^{-1}\left(s_2^j\right) \\ \qquad\qquad\qquad 0 \quad \text{otherwise} \end{array} \right. \tag{6.2}$$

Let $A$ be a subset on $S_1$ and let $\chi_A(s_1{}^i)$ be its characteristic function and assume $B = R$ on $S = S_1 \times S_2$ . The set $C = A \circ B$ results from the application of the classic *modus ponens* according to the scheme:

i) First premise: if $s_1 = s_1{}^i$, then $s_2 \in G\left(s_1{}^i\right), \forall i = 1,\dots,n$

ii) Second premise: $s_1 \in A$

iii) Conclusion: $s_2 \in C$

C is a subset of $S_2$ containing singletons $s_2^j$ such that $G^{-1}(s_2{}^j) \cap A \neq \emptyset$. Formally: let

$$\chi_A(s_1^i, s_2^j) = \chi_A(s_1^i) \tag{6.3}$$

be the *cylindrical extension* of $A$ on the product space $S = S_1 \times S_2$; on this space, the characteristic function of $C$ is given by

$$\chi_C(s_1^i, s_2^j) = \min\left(\chi_A(s_1^i, s_2^j), \chi_{B=R}(s_1^i, s_2^j)\right) \tag{6.4}$$

and hence, on the space $S_2$, the characteristic function of $C$ is given by its *shadow* or *projection*:

$$\chi_C(s_2^j) = \max_i\left(\chi_C(s_1^i, s_2^j)\right) \tag{6.5}$$

or, in an equivalent manner, that is more appropriate when an analytical expression is available for the mappings:

$$\chi_C(s_2^j) = \left\langle \begin{array}{l} \max\limits_{s_1^i \in G^{-1}\left(s_2^j\right)} \left(\chi_A(s_1^i)\right) \\ 0 \quad \text{if } G^{-1}\left(s_2^j\right) = \varnothing \end{array} \right. \tag{6.6}$$

Symmetrically, if $A$ is a subset on $S_2$ and $\chi_A(s_2{}^j)$ is its characteristic function:

$$\chi_C(s_1^i) = \max_j \min\left(\chi_A(s_1^i, s_2^j) = \chi_A(s_2^j), \chi_{B=R}(s_1^i, s_2^j)\right) \tag{6.7}$$

or, in an equivalent manner:

$$\chi_C(s_1^i) = \left\langle \begin{array}{l} \max\limits_{s_2^i \in G\ \left(s_1^i\right)} \left(\chi_A(s_2^i)\right) \\ 0 \quad \text{if } G\left(s_1^i\right) = \varnothing \end{array} \right. \tag{6.8}$$

For infinite spaces the max operator in Eq. (6.5) and (6.7) should be substituted by the sup operator.

**Example 6.1.** Let $n = 4$ and $m = 3$. Let the relation $R$ be defined by the point-valued function:

$$f(s_1{}^1) = s_2{}^2\,;\ \ f(s_1{}^2) = \varnothing\,;\ \ f(s_1{}^3) = s_2{}^2\,;\ \ f(s_1{}^4) = s_2{}^1$$

or by the matrix $\chi_R(s_1^i, s_2^j)$ in Table 6.1. The sums of rows and columns show that $G$ is a point-valued function while $G^{-1}$ is multi-valued.

**Table 6.1** Characteristic function $\chi_R(s_1^i, s_2^j)$

| | $s_2{}^1$ | $s_2{}^2$ | $s_2{}^3$ | **Tot** |
|---|---|---|---|---|
| $s_1{}^1$ | 0 | 1 | 0 | **1** |
| $s_1{}^2$ | 0 | 0 | 0 | **0** |
| $s_1{}^3$ | 0 | 1 | 0 | **1** |
| $s_1{}^4$ | 1 | 0 | 0 | **1** |
| **Tot** | **1** | **2** | **0** | **3** |

Let $A = \left\{ s_1^2, s_1^3 \right\}$, and hence $\chi_A(s_1) = (0,\ 1,\ 1,\ 0)$. Eq. (6.4) gives:

$$\chi_C\left(s_1, s_2\right) = \min\left( \begin{vmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{vmatrix}, \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} \right) = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

and finally Eq. (6.5):

$$\chi_C(s_2) = (\max(0,0,0,0) = 0,\ \ \max(0,0,1,0) = 1,\ \ \max(0,0,0,0) = 0))$$

For example in an equivalent manner Eq. (6.6) for $j = 2$ gives:

$$\chi_C(s_2^2) = \max_{s_1^i \in \left\{ s_1^1, s_1^3 \right\}} (\chi_A(s_1^i)) = \max\ (0,\ 1) = 1$$

If $A = \left\{ s_2^2, s_2^3 \right\}$, $\chi_A(s_2) = (0,1,1)$, Eq. (6.4) gives:

$$\chi_C\left(s_1, s_2\right) = \min\left( \begin{vmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix}, \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} \right) = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

and finally Eq. (6.7):

$$\chi_C(s_1) = (\max(0,1,0) = 1,\ \ \max(0,0,0) = 0,\ \ \max(0,1,0) = 1),\ \ \max(0,0,0) = 0)$$

The rule of intersection can be directly applied to combine 3 or more sets. For example let $R$ be a subset of a three-dimensional product space $S = S_1$ x $S_2$ x $Z$, defining a deterministic relation, i.e. a pair of point-valued or more generally set valued relations $G$, $G^{-1}$ between points $P$ on $S_1$ x $S_2$ and set $G(P)$ on $Z$ (r. between point $P$ on $Z$ and set $G^{-1}$ $(P)$ on $S_1$ x $S_2$).

Let $A$ and $B$ be two subsets, on $S_1$ and $S_2$ respectively, and let $\chi_A\left(s_1, s_2, z\right) = \chi_A\left(s_1\right); \chi_B\left(s_1, s_2, z\right) = \chi_B\left(s_2\right)$ be the cylindrical extension of their characteristic functions onto $S_1$ x $S_2$ x $Z$. The characteristic function of the subset $C$ on $Z$ is given by:

$$\chi_C(z) = \sup_{S_1 \times S_2} \min\left(\chi_A(s_1, s_2, z), \chi_B(s_1, s_2, z), \chi_R(s_1, s_2, z)\right) \tag{6.9}$$

or, in an equivalent manner:

$$\chi_C(z) = \left\langle \begin{array}{c} \sup\limits_{(s_1,s_2)\in G^{-1}(z)} \left(\min\left(\chi_A(s_1), \chi_B(s_2)\right)\right) \\ 0 \quad \text{if } G^{-1}(z) = \varnothing \end{array} \right. \tag{6.10}$$

#### 6.1.1.2 Union

If, on the other hand, $A$ AND $B$ are weaker conditions, i.e. they are just indications of *possibility* which exclude that the true solution *necessarily* belongs to the complementary sets, then the uncertainty of the combination of non-coincident sets grows. In fact, based on the second De Morgan's Law, we have (see Figure 6.2)[N 6-2]:

$$\begin{aligned} &C = \left(A^{\mathrm{C}} \cap B^{\mathrm{C}}\right)^{\mathrm{C}} = A \cup B \\ &\chi_C(s) = 1 - \min(1-\chi_A(s), 1-\chi_B(s)) = \max\left(\chi_A(s), \chi_B(s)\right) \end{aligned} \tag{6.11}$$

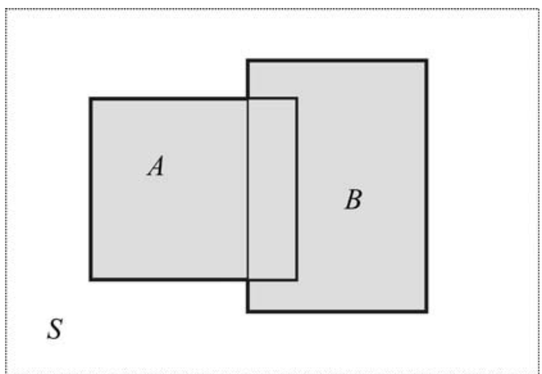



**Fig. 6.2** Combination by union

In this case, unlike the previous one, the resulting uncertainty for the decision-maker increases and the rule works with every pair of subsets (even if they are totally conflicting); it is hence strongly recommended when the sources of information are not entirely reliable.

#### 6.1.1.3 Convolutive Averaging

When a total or partial ordering is recognised or assigned on the space $S$, the decision-maker could employ a third rule based on the operation of averaging, well-suited to the natural logic of the human brain. For example, if a distance $d$ between points or subsets is defined in $S$, the decision-maker (for example an archer or marksman aiming at a target) could focus his attention on the set of points:

$$C(A, B) := \{P\in S \mid d(P_A, P) = d(P_B, B), P_A\in A, P_B\in B\} \tag{6.12}$$

**Example 6.2.** Let $S = \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9\}$, $A = \{P_1, P_2\}$, $B = \{P_7, P_8, P_9\}$ and assume $d(P_i, P_j) = |i\text{-}j|$. Hence $C = \{P_4, P_5\}$ because $d(P_1, P_4) = d(P_7, P_4) = 3$; $d(P_2, P_5) = d(P_8, P_5) = 3$ and $d(P_1, P_5) = d(P_9, P_5) = 4$.

Let $s_A$ and $s_B$ be elements of sets, on the space of the real numbers (for example intervals), $A$ and $B$, respectively. One obtains, according to the rules of interval arithmetic, the interval $C = (A + B) / 2$, or more generally on an $n$-dimensional Euclidean space (see Figure 6.3 in a two-dimensional space) the set defined, according to Eq. (6.10), by the characteristic function:

$$\chi_C(s) = \sup_{(s_A, s_B):\frac{s_A+s_B}{2}=s; s\in\mathbb{R}^n; s_B\in\mathbb{R}^n} \min\left(\chi_A(s_A), \chi_B(s_B)\right) \tag{6.13}$$

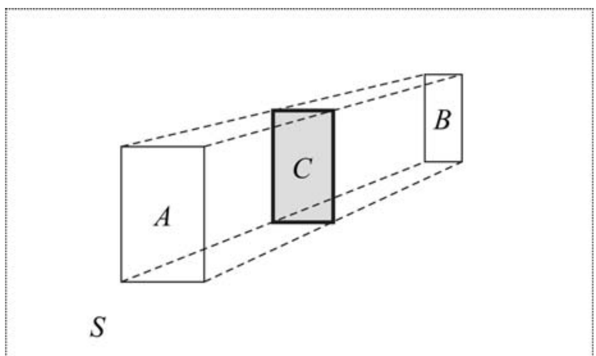



**Fig. 6.3** Combination by convolutive averaging

**Example 6.3.** Let A = [-1, 1], B = [7, 10]; the interval arithmetic rule for summing gives C = [7-1, 1+10]/2 = [3, 5.5]. Observe that Eq. (6.13) gives $\chi_C = 1$ for any point $s$ of $C$, $\chi_C = 0$ for any $s$ outside of the interval $C$.

This rule (known as *Convolutive Averaging* (or c-*Averaging*), considering the analogy with the definition of "expectation" or "mean") works with every pair of subsets and drastically reduces the uncertainty presented to the decision-maker; in the case of partial or total conflict between $A$ and $B$, the rule hides the conflict.

#### 6.1.1.4 Discussion

The 3 basic rules presented above satisfy the requirements of:

Commutativity: $C(A, B) = C(B, A)$;
Associativity: $C(A, C(B, D)) = C(C(A,B), D)$
Idempotence: $C(A, A) = A$.

In the following, we will distinguish between combining and updating.

Updating entails the presence of an existing piece of information, which must be changed in the light of newly acquired information, and order may or may not matter. For example, knowledge is the result of a process of updating, rather than combining, and the order in which information has been received and processed by the brain may make a difference to one's knowledge at a given time $t$.

The ability to optimally and quickly combine pieces of information acquired successively is a fundamental property of the human brain that has evolved over the millennia, thus assuring a definite advantage to mankind over all other species. Likewise, at a higher level, a civilization's knowledge nearly always depends on its history. Therefore, we argue that temporal indifference may or may not be a desirable attribute in the epistemic interpretation of a basic theory modeling logical and psychological degrees of partial belief of a person or intentional system.

Commutativity is justified considering that when combining two pieces of information $A$ and $B$, $A$ and $B$ must be treated symmetrically, i.e. combining $A$ with $B$ must yield the same result as combining $B$ with $A$, because there is no information on which of the two pieces of information was prior. Similarly, Associativity is reasonable if temporal indifference is necessary in the treatment of the combination.

On the other hand, the third requirement (Idempotence) does not appear to be well justified: indeed, our confidence in $A$ grows if the same information is observed repeatedly, and this increase in confidence justifies the rules of statistical combination and subjective/objective probability theory.

### *6.1.2 Statistical Combination and Updating*

Let $n_i$ be the number of observations of singleton $s^i$ in a finite space of events, let $n$ be the total number of observations, and let $\chi_j$ be the characteristic function of the singleton $s^j$ observed at the $j$-th observation. Assuming that our confidence grows linearly with the number of repetitions of events, one estimates confidence through the relative frequency:

$$fr\left(s^i\right) = \frac{n_i}{n} = \sum_{j=1}^{n} \frac{\chi_j\left(s = s^i\right)}{n} \tag{6.14}$$

It is interesting to observe that relative frequencies *fr* are obtained mixing (p-averaging) characteristic functions; moreover, in the case of conflicting events, Case 2 (union) is applied, thus completely preserving the conflict for the attention of the decision-maker.

On the other hand, when the expectation of a random variable or the mean value of a list of numerical values are evaluated, Case 3 (c-averaging) is applied, and the conflict disappears.

Eq. (6.14) can be generalized to the case of observations of not necessarily disjoint events $A^j$, i.e. to the statistical collection of sets or focal elements of a random set:

$$fr\left(s^i\right) = \sum_{j=1}^{n} \frac{\chi_{A^j}\left(s^i\right)}{n} \tag{6.15}$$

The relative frequencies of the singletons given by this equation can be compared with the *white* distributions defined in Section 3.2.3.1.

Therefore, Case 2 is a natural candidate for combining probabilistic assignments both in the case of the frequentist and subjectivist view of probability.

Let $\{(A^i, m_1)\}$ and $\{(A^i, m_2)\}$ be two random sets on the same set $S$ with $m_1$ and $m_2$ two assigned relative frequencies of events (focal elements) $A^i$:

$$m_1\left(A^i\right) = n_{1,i} / n_1 \quad ; \quad m_2\left(A^i\right) = n_{2,i} / n_2 \tag{6.16}$$

The statistical combination of the random sets is given by:

$$m_{12}\left(A^i\right) = \frac{n_{1,i} + n_{2,i}}{n_1 + n_2} \tag{6.17}$$

For an infinite number of realisations, the combination of two random sets is obtained by a simple averaging operation:

$$m_{12}\left(A^i\right) = \frac{m_1\left(A^i\right) + m_2\left(A^i\right)}{2} \tag{6.18}$$

which treats the two random sets in a symmetric fashion: i.e. the operation is commutative.

### 6.1.3 Bayesian Combining and Updating in Probability Theory

Within the context of probabilistic modeling of uncertainty, updating follows on from the Bayesian techniques, i.e. from the axiomatic definition of conditioned probability of two events $A$ and $B$ in a probability space:

$$P(A/B) = P(A \cap B)/P(B) = P(B/A)P(A)/P(B) \tag{6.19}$$

or the equivalent relation between conditional, joint and marginal probabilities on joint spaces (see Chapter 2). A simple example can clarify how these techniques operate.

**Example 6.4.** Let us consider the population composed of different batches of concrete used for structural purposes. For this population of objects, conditional probability distributions are assumed to be known:

a) between a variable relative to mechanical resistance $r \in R$ (e.g. the cubic characteristic compressive strength after 28 days) and a variable relative to the mixture composition $c \in C$ (e.g. the water/cement ratio, in weight);

b) between the resistance $r$ and the result $s \in S$ from a non-destructive test (e.g. a sclerometric test).

For the sake of simplicity, let:

$$R = \{ r^1, r^2, r^3, r^4 \}\ ;\ C = \{c^1, c^2, c^3, c^4\}\ ;\ S = \{ s^1, s^2, s^3, s^4 \}$$

and assume the conditional probabilities shown in Table 6.2 for the relation between c and r, and Table 6.3 for the relation between $s$ and $r$.

**Table 6.2** Conditional probabilities $P(c/r)$

| | $P(c/r^1)$ | $P(c/r^2)$ | $P(c/r^3)$ | $P(c/r^4)$ |
|---|---|---|---|---|
| $c^1$ | 0.72 | 0.12 | 0.04 | 0 |
| $c^2$ | 0.24 | 0.60 | 0.24 | 0.04 |
| $c^3$ | 0.04 | 0.24 | 0.60 | 0.24 |
| $c^4$ | 0 | 0.04 | 0.12 | 0.72 |
| | **1.00** | **1.00** | **1.00** | **1.00** |

**Table 6.3** Conditional probabilities $P(s/r)$

| | $P(s/r^1)$ | $P(s/r^2)$ | $P(r/r^3)$ | $P(s/r^4)$ |
|---|---|---|---|---|
| $s^1$ | 0.80 | 0.20 | 0 | 0 |
| $s^2$ | 0.2 | 0.60 | 0.20 | 0 |
| $s^3$ | 0 | 0.20 | 0.60 | 0.20 |
| $s^4$ | 0 | 0 | 0.20 | 0.80 |
| | **1.00** | **1.00** | **1.00** | **1.00** |

Suppose marginal distribution $P(r)$ is specified or assumed *a priori*: for example, lacking any specific information, let us assume (according to the Principle of Indifference) uniform marginal distribution for the variable $r$, i.e. $P(r^i) = 0.25$, for $i = 1, 2, 3, 4$ .

The joint probability distributions $P(s^j, r^i) = P(s^j / r^i)\, P(r^i)$, $P(c^k, r^i) = P(c^k / r^i)\, P(r^i)$ and the marginals $P(s^j) = \Sigma_i P(s^j, r^i)$, $P(c^k) = \Sigma_i P(c^k, r^i)$ can be calculated as shown in Table 6.4.

**Table 6.4** Joint probability distributions $P(s^j, r^i)$, $P(c^k, r^i)$ and the marginals $P(s^j)$, $P(c^k)$

| $P(c,r)$ | $r^1$ | $r^2$ | $r^3$ | $r^4$ | $P(c)$ | $P(c,r)$ | $r^1$ | $r^2$ | $r^3$ | $r^4$ | $P(s)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $c^1$ | 0.18 | 0.03 | 0.01 | 0 | 0.22 | $s^1$ | 0.2 | 0.05 | 0 | 0 | 0.25 |
| $c^2$ | 0.06 | 0.15 | 0.06 | 0.01 | 0.28 | $s^2$ | 0.05 | 0.15 | 0.05 | 0 | 0.25 |
| $c^3$ | 0.01 | 0.06 | 0.15 | 0.06 | 0.28 | $s^3$ | 0 | 0.05 | 0.15 | 0.05 | 0.25 |
| $c^4$ | 0 | 0.01 | 0.03 | 0.18 | 0.22 | $s^4$ | 0 | 0 | 0.05 | 0.2 | 0.25 |
| | **0.25** | **0.25** | **0.25** | **0.25** | **1** | | **0.25** | **0.25** | **0.25** | **0.25** | **1** |

**Table 6.5** Joint probability distributions $P(s^j, r^i, c^k = c^2)$ and updated conditionals $P(s^j, r^i / c^k = c^2)$, $P(r^i / c^k = c^2, s^j = s^3)$

| $P(r,s,c^2)$ | $r^1$ | $r^2$ | $r^3$ | $r^4$ | $P(s)$ |
|---|---|---|---|---|---|
| $s^1$ | 0.048 | 0.03 | 0 | 0 | 0.078 |
| $s^2$ | 0.012 | 0.09 | 0.012 | 0 | 0.114 |
| $s^3$ | 0 | 0.03 | 0.036 | 0.002 | 0.068 |
| $s^4$ | 0 | 0 | 0.012 | 0.008 | 0.02 |
| | **0.06** | **0.15** | **0.06** | **0.01** | **0.28** |

| $P(r,s/c^2)$ | $r^1$ | $r^2$ | $r^3$ | $r^4$ | $P(s/c^2)$ |
|---|---|---|---|---|---|
| $s^1$ | 0.171429 | 0.107143 | 0 | 0 | 0.278571429 |
| $s^2$ | 0.042857 | 0.321429 | 0.04286 | 0 | 0.407142857 |
| $s^3$ | 0 | 0.107143 | 0.12857 | 0.00714 | 0.242857143 |
| $s^4$ | 0 | 0 | 0.04286 | 0.02857 | 0.071428571 |
| | **0.214286** | **0.535714** | **0.21429** | **0.03571** | **0.28** |

| | $r^1$ | $r^2$ | $r^3$ | $r^4$ | |
|---|---|---|---|---|---|
| $P(r/s^3,c^2)$ | **0** | **0.441176** | **0.52941** | **0.02941** | **1** |

In their turn, these distributions $P(r, c)$ and $P(r, s)$ can be seen as marginal distributions on the plane ($C$ x $R$) and ($S$ x $R$), respectively, of a joint probability distribution on the space ($C$ x $R$ x $S$) defined by the equations:

$$P\,(r, s, c) = P\,(r, s)\, P(c/r) = P\,(r, c)\, P(s/r) = P(r, s)\, P(r, c) \,/\, P(r)$$

Let us now suppose that we obtain new deterministic information about the exact value of one of the variables, e.g. $c = c^*$. We immediately conclude that:

$$P\,(r, s\,/c = c^*) = P(r, s, c^*) \,/\, P(c^*)$$

and moreover, the updated distributions for $r$ and $s$ are the new marginal ones.

Suppose that we now obtain the deterministic information that $c = c_2$ ; the marginal distributions, relative to $r$ and $s$ respectively, are modified as indicated by the sums of rows and columns of the matrix $P(r, s\,/\,c^2)$ in Table 6.5.

If we get further information $s = s^3$ , we obtain: $P(r\,/s^3, c^2) = P(r, s^3\,/\,c^2) \,/\, P(s^3\,/\,c^2)$, as shown finally in Table 6.5, last row.

Bayes Theorem clearly derives from the application of Case 1 (intersection). Combining a probabilistic distribution $m_1(s^i)$ for singletons $s^i$ and a deterministic event $B$ ($m_2(B) = 1$), the resulting updated distribution $m_{12}$ is given by the ratio of the probability of intersections to a normalization factor $K$ equal to $m_1(B)$ (supposed positive):

$$m_{12}\left(A^i\right) = P\left(A^i \middle| B\right) = \frac{m_1\left(A^i \cap B\right)}{m_1\left(B\right)} \tag{6.20}$$

Of course the normalisation allows to obtain:

$$m_{12}\left(B\right) = P\left(B \mid B\right) = \frac{\sum_{s^i \in B} m_1\left(s^i\right)}{m_1\left(B\right)} = 1$$

When the normalization factor $K$ (i.e. the prior probability of the observed event) is much smaller than 1, posterior probabilities increase dramatically.

Eq. (6.20) can be recursively applied for updating, when two or more successive deterministic events $B^1$, $B^2$,… are given or observed. It is easy to check that the resulting distribution is independent from the order (temporal indifference, as previously discussed in § 6.1.1.4) and it is equivalent to evaluating conditional probabilities with respect to the event $B = B^1 \cap B^2 \ldots\ldots$ This result is coherent with the meaning of Case 1: when all the successively obtained sets are reliable, attention should be restricted to their intersection.

Therefore, in a probabilistic setting, there is no difference between combining and updating.

The extension of a probabilistic measure through a point-valued deterministic mapping, $f$, can be considered as a particular application of the Bayes theorem: in both cases, we have to combine probabilistic *a priori* information with deterministic information.

In fact, let a joint probability distribution $P(s) = P(s_1, s_2 \ldots s_n)$ of the variables $s_1, s_2, \ldots s_n$ be given on the space defined by the Cartesian product $S = S_1 \times S_2 \times \ldots S_n$. The mapping $z = f(s)$, whose range is $Z$, *extends* to the space $S \times Z$ the joint probability distribution:

$$P(s,z) = \left\langle \begin{array}{ll} P(s) & \text{if } s \in f^{-1}(z) \\ 0 & \text{if } f^{-1}(z) = \varnothing \end{array} \right. \tag{6.21}$$

and then the marginal probability on $Z$ reads:

$$P(z) = \sum_S P(s,z) \tag{6.22}$$

On a finite space, let $\chi_R(s_i, z_j)$ be the characteristic function of the equivalent (to the pair of the point-valued mapping $f$ and the multi-valued mapping $G^{-1}=f^{-1}$) deterministic relation

$$R \subseteq S \times Z \mid \forall s^i : \sum_{z^j \in Z} \chi(s^i, z^j) \leq 1 \tag{6.23}$$

Eqs. (6.21) and (6.22) can be given by the following equation:

$$P(z^j) = \sum_{s^i \in S} P(s^i) \cdot \chi_R(s^i, z^j) \tag{6.24}$$

or, equivalently and underlining the analogy with Eq. (6.5) when sup operator is substituted by summation (see Note 6.1):

$$P(z^j) = \sum_{s^i \in S} \min\left(P(s^i), \chi_R(s^i, z^j)\right) \tag{6.25}$$

**Example 6.5.** Let us again consider the updated probability distribution of the concrete strength $r$ (on which the safety of a structure depends) obtained in Example 6.4 and the point-valued relation between $r$ and a finite list of values of a measure $z$ of the safety of the structure. Let relation $f$ be given by the matrix of the joint characteristic function $\chi(r^i, z^j)$ displayed in Table 6.6.

**Table 6.6** Deterministic relation between $r$ and $y$ in Example 6.5

| | $\chi(r^i, z^1)$ | $\chi(r^i, z^2)$ | $\chi(r^i, z^3)$ | $\chi(r^i, z^4)$ | **Tot** |
|---|---|---|---|---|---|
| $r^1$ | 1 | 0 | 0 | 0 | **1** |
| $r^2$ | 0 | 1 | 0 | 0 | **1** |
| $r^3$ | 0 | 1 | 0 | 0 | **1** |
| $r^4$ | 0 | 0 | 0 | 1 | **1** |
| **Tot** | **1** | **2** | **0** | **1** | |

The totals of rows demonstrate that the monotonically increasing $f$ is single-valued (while on the contrary the totals of the columns show that $f^{-1}$ is multi-valued. Eqs. (6.21) and (6.22) give the matrix of joint probabilities $P(r^i, z^j)$ and marginals $P(z^j)$ respectively in Table 6.7.

**Table 6.7** Joint distribution and marginals in Example 6.5

| | $P(r^i, z^1)$ | $P(r^i, z^2)$ | $P(r^i, z^3)$ | $P(r^i, z^4)$ | $P(r^i)$ |
|---|---|---|---|---|---|
| $r^1$ | 0 | 0 | 0 | 0 | 0 |
| $r^2$ | 0 | 0.44118 | 0 | 0 | 0.44118 |
| $r^3$ | 0 | 0.52941 | 0 | 0 | 0.52941 |
| $r^4$ | 0 | 0 | 0 | 0.02941 | 0.02941 |
| $P(z^j)$ | **0** | **0.97059** | **0** | **0.02941** | **1** |

## 6.2 Limits Entailed by the Probabilistic Solution

The above discussion on the probabilistic solution and Bayes theorem, together with the results shown in the previous chapters, highlights some limitations and possible extensions of the Bayesian approach.

### 6.2.1 Set-Valued Mapping

Let the function extending the probability measures from $S$ to $Z$ be given through a set-valued mapping $G$ (or a general deterministic relation $R$ without the constraint given in Eq. (6.23)) to the power set of $Z$ ($\mathcal{P}(Z)$, $G$ : $S \rightarrow \mathcal{P}\{Z\}$; $G$ associates every $s \in S$ with a subset $A=G(s) \subseteq Z$.

In this case, a (generally non-consonant) random set measure is induced on Z

$$\Im = \left\{ \left( A^i = G\left(s^i\right), m(A^i) = P(s^i) \right) \right\} \tag{6.26}$$

and it is consequently possible to calculate the interval [$Bel$ ($B$), $Pla$($B$)] containing $P(B)$ for each $B \subseteq Z$ (see Section 3.2.3).

### 6.2.2 Variables Linked by a Joint Random Relation

Let the variables ($s_1$, $s_2$..) be linked by a joint random relation $\Im = \left\{ A^i, m(A^i) \right\}$ (Chapter 4, § 4.1). The extension principle (Section 5.2.1), allows the corresponding random set measure on Z $\Re = \left\{ R^i = G(A^i), m(R^i) = \rho^i = m(A^i) \right\}$ to be determined and then again, for each $B \subseteq Y$, the interval [$Bel$ ($B$), $Pl$($B$)] containing $P(B)$ can be calculated.

For finite spaces let $R \subseteq S$ x $Z$ be a deterministic relation and let $\chi_R$ be its characteristic function. The characteristic function of the focal elements $R^i$ can be computed by Eq. (6.9):

$$\forall i : \chi_{R^i}\left(z\right) = \sup_{S}\ \min\left(\chi_{A^i}\left(s\right), \chi_R\left(s, z\right)\right) \tag{6.27}$$

Some particular cases are of interest, summarizing the results given in Section 5.2.1.

Suppose that the random relation $\Im$ is consonant. Hence $\Im$ is completely defined by its contour function (plausibility of the singletons $s$) $\mu(s)$ or the associated possibility distribution $\pi(s) = \mu(s)$ (see Eq. (3.31)). Hence the random set induced on $Z$ is also consonant and its possibility distribution $\pi(z) = \mu(z)$ is defined by Eq. (5.48), or equivalently:

$$\mu(z) = \sup_{S} \min\left(\mu(s), \chi_R(s,z)\right) \tag{6.28}$$

Eq. (6.28) can be derived in a form quite similar to the Bayesian procedure: we can write the *cylindrical extension* of $\mu(s)$ on the space $S$ x $Z$:

$$\mu(s,z) = \mu(s) \tag{6.29}$$

The combination in this space is expressed by the rule:

$$\mu_C(s,z) = \min\left(\mu(s,z), \chi_R(s,z)\right) \tag{6.30}$$

and then $\mu(z)$ is nothing else but the *projection* of $\mu_C(s, z)$ on $Z$ (analogous to the Bayesian case of marginal distribution on $Z$):

$$\mu(z) = \sup_{S} \mu_C(s,z) \tag{6.31}$$

Suppose further that the random relation on $S$ is a fuzzy Cartesian random product (see § 4.3.5) (i.e. the focal elements are a nested sequence of Cartesian products of intervals, each of which pertains to one variable only): the variables are non-interactive (see § 5.2.1.2) and their extension is given by Eq. (5.57). Again on finite spaces this equation can be equivalently given by the following:

$$\mu(z) = \sup_{S_1 \times S_2 \times .. \times S_n} \min\left(\mu_{F_1}(s_1), \mu_{F_2}(s_2), ..., \mu_{F_n}(s_n), \chi_R(s_1, s_2, ..., s_n, z)\right) \tag{6.32}$$

### *6.2.3 Conditioning a Random Set to an Event B*

Let *Bel* be a belief function corresponding to a random set defined on $S$, and let $\Psi$ be the set of all probability functions consistent with *Bel*. As in Section 3.2.5, conditional belief and plausibility functions (valid only if $Bel(B) > 0$) are given by:

$$\begin{aligned} Bel(A \mid B) &= \inf_{P \in \Psi} P(A \mid B) \\ Pla(A \mid B) &= \sup_{P \in \Psi} P(A \mid B) \end{aligned} \tag{6.33}$$

Moreover, in finite spaces the search can be restricted to the finite set *EXT* of the extremes of the convex set $\Psi$.

In the same way that conditional probability functions are probability functions, conditional belief functions are belief functions and can be calculated by Eq. (3.35), which also gives the corresponding plausibility functions. Hence conditioning a random set to an event $B$ gives a conditional posterior random set.

However two limits of this formulation must be underlined.

Firstly, as discussed at the end of Example 3.12, a closed rule directly giving the probabilistic assignment of the posterior random set from the probabilistic assignment of the prior random set cannot be given. Mobius inversion (Eq. (3.8)) of the conditional belief function must be applied.

Secondly, when updating belief function recursively applying Eq. (3.35), given or observed two or more successive deterministic events $B^1$, $B^2$,… the order in which the updating events are observed does matter. In other words, imprecision entails the loss of commutativity. It seems that temporal indifference (as previously discussed in § 6.1.1.4) could be not respected and the final result is not equivalent to conditioning the prior belief function with respect to the event $B = B^1 \cap B^2$.....

**Example 6.6.** (Modified after (Fagin and Halpern 1991). Let $R = \{r^1, r^2, r^3, r^4\}$ be a finite space for the characteristic strength of a concrete, measured according to the prior information by the random set

$$\left\{\left(A^1 = r^1, m^1 = \frac{1}{4}\right), \left(A^2 = r^2, m^2 = \frac{1}{4}\right), \left(A^3 = \left\{r^3, r^4\right\}, m^3 = \frac{1}{2}\right)\right\}$$

The focal elements are disjoint sets and hence:

$Bel(\{r^1\}) = Pla(\{r^1\}) = m^1 = 0.25$
$Bel(\{r^2\}) = Pla(\{r^2\}) = m^2 = 0.25$
$Bel(\{r^3, r^4\}) = Pla(\{r^3, r^4\}) = m^3 = 0.5$

while probabilities of the singletons $r^3$ , $r^4$ are bounded by:

$Bel(\{r^3\}) = Bel(\{r^4\}) = 0$ ; $Pla(\{r^3\}) = Pla(\{r^4\}) = 0.5$

Further information provides evidence that event $\{r^4\}$ is impossible: i.e. event $B_1 = \{r^1, r^2, r^3\}$, with $P(B_1 | B_1) = 1$ is the updated space of possible events. Notice that $Bel(B_1) = m^1 + m^2 = 0.5 > 0$, while $Pla(B_1) = m^1 + m^2 + m^3 = 0.5$.

Let us evaluate, through Eq. (3.34), the updated bounds of the probability of the event $A = \{r^1\}$. Taking into account that $A^C = \{r^2, r^3, r^4\}$, $A \cap B_1 = \{r^1\}$, $A^C \cap B_1 = \{r^2, r^3\}$:

$$Bel\left(\left\{r^1\right\} / B^1\right) = \frac{\left(Bel\left(\left\{r^1\right\}\right) = 0.25\right)}{0.25 + \left(Pla\left(\left\{r^2, r^3\right\}\right) = 0.25 + 0.5\right)} = 0.25;$$

$$Pla\left(\left\{r^1\right\} / B^1\right) = \frac{\left(Pla\left(\left\{r^1\right\}\right) = 0.25\right)}{0.25 + \left(Bel\left(\left\{r^2, r^3\right\}\right) = 0.25\right)} = 0.5$$

Further information again provides evidence that event $\{r^3\}$ is impossible: i.e. event $B^2 = \{r^1, r^2\}$, with $P(B^2| B^1, B^2) = 1$ is the final updated space of possible events.

Observe that $Bel\ (B^2 /B^1) = (Bel(\{r^1, r^2\}) = 0.5)/(0.5 + (Pla(\{r^3\}) = 0.5)) = 0.5 > 0$, while $Pla(B^2 /B^1) = 1$.

Let us evaluate again, through Eq. (3.35), the updated bounds of the probability of the event $A = \{r^1\}$. Taking into account that $A^C = \{r^2, r^3, r^4\}$, $A \cap B^2 = \{r^1\}$, $A^C \cap B^2 = \{r^2\}$:

$$Bel\left(\left\{r^1\right\}/B^1, B^2\right) = \frac{\left(Bel\left(\left\{r^1\right\}/B^1\right) = 0.25\right)}{0.25 + \left(Pla\left(\left\{r^2\right\}/B^1\right) = 0.25 + 0.5\right)} = 0.333;$$

$$Pla\left(\left\{r^1\right\}/B^1, B^2\right) = \frac{\left(Pla\left(\left\{r^1\right\}/B^1\right) = 0.25\right)}{0.25 + \left(Bel\left(\left\{r^2\right\}/B^1\right) = 0.25\right)} = 0.667$$

Changing the order of the observed events gives very different results. In fact, observing $B^2 = \{r^1, r^2\}$: $Bel\ (B^2) = Pla\ (B^2) = P\ (B^2) = 0.5$, and hence Bayes formula directly gives $P(\{r^1\}/ B^2) = P(\{r^1\})/ P(B^2) = 0.25/0.5 = 0.5$. Of course observing now $B^1 = \{r^1, r^2, r^3\}$ does not change the probabilities, because $B^2$ is included in $B^1$. This conclusion can be checked through Eq. (3.35):

$$Bel\left(\left\{r^1\right\}/B^1, B^2\right) = \frac{\left(Bel\left(\left\{r^1\right\}/B^2\right) = 0.5\right)}{0.5 + \left(Pla\left(\left\{r^2\right\}/B^2\right) = 0.5 + 0.5\right)} = 0.5;$$

$$Pla\left(\left\{r^1\right\}/B^1, B^2\right) = \frac{\left(Pla\left(\left\{r^1\right\}/B^2\right) = 0.5\right)}{0.5 + \left(Bel\left(\left\{r^2\right\}/B^2\right) = 0.5\right)} = 0.5$$

These results seem paradoxical, because Eq. (6.33) shows that the final result can be obtained applying Bayes' rule of conditioning to the set *EXT* of the extreme distribution of the random set, and this commutative rule directly gives the final result conditioning to the intersection of the observed events.

Clearly in this example the extreme distributions coincide with the two selectors:

$P^1{}_{EXT} = (0.25, 0.25, 0.5, 0)$, $P^1{}_{EXT} = (0.25, 0.25, 0, 0.5)$. Moreover $B^1 \cap B^2 = B^2 = \{r^1, r^2\}$, and hence

$$Bel\left(\left\{r^1\right\}/B^1, B^2\right) = \min\left(\frac{P^1_{EXT}\left(r^1\right) = 0.25}{P^1_{EXT}\left(B^2\right) = 0.5}, \frac{P^2_{EXT}\left(r^1\right) = 0.25}{P^2_{EXT}\left(B^2\right) = 0.5}\right) = 0.5;$$

$$Pla\left(\left\{r^1\right\}/B^1, B^2\right) = \max\left(\frac{P^1_{EXT}\left(r^1\right) = 0.25}{P^1_{EXT}\left(B^2\right) = 0.5}, \frac{P^2_{EXT}\left(r^1\right) = 0.25}{P^2_{EXT}\left(B^2\right) = 0.5}\right) = 0.5$$

The problem originates from Eq. (3.35): it gives exact results at any step of conditioning, but the set $\Psi$ of probability distributions compatible with the conditional belief function (whose probabilistic assignment can be calculated through the Mobius inversion) not necessarily coincides (but generally only includes) the set of conditional probability distributions derived through the Bayes' rule. This conclusion confirms and enlarges the criticism about any uncertain model based on upper/lower probabilities discussed in Section 3.3. Some other numerical examples are reported in (Walley 2000), demonstrating incorrect results obtained conditioning Choquet capacities of different order.

In Example 6.6, by using Mobius inversion it is easy to calculate that the random set obtained by conditioning to $B^1$, is the following:

$$\left\{\left(A^1 = r^1, m^1 = \frac{1}{4}\right), \left(A^2 = r^2, m^2 = \frac{1}{4}\right), \left(A^3 = \left\{r^1, r^3\right\}, m^3 = \frac{1}{4}\right), \left(A^4 = \left\{r^2, r^3\right\}, m^4 = \frac{1}{4}\right)\right\}$$

The extreme distributions coincide with the four selectors:

$P^1_{EXT/B1} = (0.25, 0.25, 0.5, 0)$, $P^2_{EXT/B1} = (0.5, 0.5, 0, 0)$, $P^3_{EXT/B1} = (0.5, 0.25, 0.25, 0)$, $P^4_{EXT/B1} = (0.25, 0.5, 0.25, 0)$.

Conditioning to $B^2$, Bayes' rule gives for example:

$$Bel\left(\left\{r^1\right\} / B^1, B^2\right) = \min\left(\begin{array}{c} \dfrac{P^1_{EXT/B^1}\left(r^1\right) = 0.25}{P^1_{EXT/B^1}\left(B^2\right) = 0.5}, \dfrac{P^2_{EXT/B^1}\left(r^1\right) = 0.5}{P^2_{EXT/B^1}\left(B^2\right) = 1}, \\ \dfrac{P^3_{EXT/B^1}\left(r^1\right) = 0.5}{P^3_{EXT/B^1}\left(B^2\right) = 0.75}, \dfrac{P^4_{EXT/B^1}\left(r^1\right) = 0.25}{P^4_{EXT/B^1}\left(B^2\right) = 0.75} \end{array}\right) =$$
$$= \min\left(0.5, 0.5, 0.667, 0.333\right) = 0.333$$

### *6.2.4 Not Deterministic Mapping*

The situation is rather more complex when function $G$ or $f$ is not deterministic, because the problem goes well beyond Bayesian formulation. Now we are facing the combination of two kinds of information, both being uncertain and derived from two distinct bodies of evidence: one relative to the independent variables and the other to the function which relates independent variables to a dependent variable.

More generally, in probabilistic terms, we have to combine information which yields two distinct probability distributions on the same space. This topic leads us to the core of the problems from which both Shafer's Evidence Theory and Zadeh's Fuzzy Set Theory originated. Both theories may be seen as attempts to solve this problem.

The critical point about the limits of Bayes'rule for updating is well described in this quotation in the initial chapters of Shafer's book:

*"In the Bayesian theory, the task of telling how our degrees of belief ought to change as new evidence is obtained falls to Bayes' rule of conditioning: we represent the new evidence as a proposition and condition our prior Bayesian belief function on that proposition. Here we find no obvious symmetry in the treatment of new and old evidence. And more importantly, we find that the assimilation of new evidence depends on an astonishing assumption: we must assume that the exact and full effect of that new evidence is to establish a single proposition with certainty. In contrast to Dempster's rule of combination, which can accommodate new evidence that justifies only partial beliefs, Bayes' rule of conditioning* requires *that the new evidence be expressible as a certainty*" (Shafer 1976, pp. 25-26).

On the another hand, alternative models of uncertainty (particularly fuzzy models) have been explicitly conceived as antitheses to the probabilistic paradigm and to classical Boolean logic and Set Theory.

The theory of fuzzy sets embodies Zadeh's original idea of "vagueness" (Zadeh 1965) i.e. "the lack of precise or sharp distinction or boundaries" (Klir 1995)[N 6-3].

In the following Sections 6.3 and 6.4 the above-mentioned attempts will be reviewed, underlining their connections to the theory of random sets. In fact in Evidence Theory the formal descriptions of the available uncertain information to be combined together is given through belief functions or equivalent probabilistic assignment of random sets, while the information conveyed by fuzzy sets are equivalently described by associated consonant random sets or by possibility distributions of the singletons, as discussed in Section 3.2.4.

### *6.2.5 Probability Kinematics and Nets of Italian Flags*

An independent but converging criticism about the limits of Bayes rule was expressed by Jeffrey, in his book on the *Logic of Decisions* (Jeffrey 1983), the first edition of which dated to 1965 (the same year as Zadeh's first paper on fuzzy sets and 2 years before Dempster's paper on upper/ lower probabilities).

"*One day Bayesian robots may be built; but at present there are not such creatures, and in particular human beings are not de facto Bayesians. Bayesian decision theory provides a set of norms for human decision making; but it is far from being a true description of our behavior*"(Jeffrey 1983, pp. 166-167)

Jeffrey notes that conditionalization performed via Bayes' rule repeatedly applied to events $B^1,\ldots, B^k$ is independent of the order but irreversible. There is no observed event or proposition $B^2$, such that for an event $A$:

$$P(A \mid B^1, B^2) = P(A)$$

except the trivial case $P(B^1) = 1$, in which case the agent was sure of the truth of the event $B^1$ even before the first observation.

Consider now an agent observing an event B in a probability space ($S$, $\mathcal{S}$, $P$) with prior positive probability and attributing to this event a posterior probability $a < 1$. Jeffrey suggests that the probabilities of any other measurable event $A$ could be updated, when supposing that the conditional probabilities $P(A \mid B)$ and $P(A \mid B^C)$ do not change. In fact the theorem of total probability (see Eq. (2.7) gives:

$$P_{POSTERIOR}(A \mid B) = a \cdot P(A \mid B) + (1\text{-}a) \cdot P(A \mid B^C) \tag{6.34}$$

**Example 6.7.** (Modified after (Jeffrey 1983), the *mudrunner* example). A concrete mix is supposed to perform much better in a humid environment than in a dry environment.

The judgment of the engineer about the probability of good performance of a particular structure built using this concrete mix should be updated by the information that the probability of $B$ = "local dry environment" is high (and hence probability of $B^C$ = "local humid environment" is low). However the forecast should have no effect on the proposition that the concrete should have a good performance in the humid environment (high conditional probability of the state $A$ of good performance) and on the contrary bad performance in the dry environment.

For example, let $P(A \mid B) = 0.6$, $P(A \mid B^C) = 0.99$ and $a = 0.8$. Eq. (6.34) gives:

$$P_{POSTERIOR}(A \mid B) = 0.8 \text{ x } 0.6 + (1 - 0.8) \text{ x } 0.99 = 0.678$$

When no information is available about the environment the Principle of Indifference (see Section 1.3) could suggest $a = 0.5$ and hence $P_{POSTERIOR}(A \mid B) = (.8 + 0.99) \text{ x } 0.5 = 0.895$.

Clearly any conclusion obtained via Eq. (6.34) is reversible: i.e. mistakes can be erased by successive more reliable information. However the final conclusion is order dependent, because strictly related to the last observation of event $B$ and its final probability.

Jeffrey's rule (Jeffrey 1983) can be extended to a collection of observed events, when precise probabilities can be attached to the σ-algebra generated by the collection (see section 2.1). Suppose for example a prior $P$ is available, and that one makes an observation $Ob$ of $k$ mutually exclusive events $B^1,\ldots, B^k$ (measurable with respect to $P$) with probability $a_i$, $i = 1,\ldots,k$, with $a_1+\ldots+ a_i =1$. The probability of an event $A$ (again measurable with respect to $P$) is denoted as $P(A \mid Ob)$, and is defined as:

$$P(A|\,Ob) := a_1 \cdot P(A|\,B^1) + \ldots + a_k \cdot P(A|\,B^k) \tag{6.35}$$

Now, if either $A$ or any of the sets $B^i$ is non-measurable with respect to $P$, i.e. if we start from a random set, then one can define a lower envelope as done before for the belief function. Indicating with μ the extension of $P$ such that the $A$, $B^1$, …, $B^k$ are measurable with respect to μ:

$$\mu_{LOW}(A|Ob) = \inf\{\ \mu\ (A|Ob)\} \tag{6.36}$$

Although Fagin and Halpern (Fagin and Halpern 1991) conjecture that $\mu_{LOW}(A|Ob)$ is a belief function, this has not yet been proved. Moreover, no closed-form solution such as Eq. (3.35) is available for its calculation.

In an independent manner Blockley and his research team in Bristol have developed a model to evaluate the final state of truth of a proposition (for example about the safety of an engineering system) depending on a net of different levels of compound propositions (Cui and Blockley 1991). For every proposition the state of truth is measured by an *interval probability* and the mechanism transferring the measure from a lower level to a higher one is again based on the theorem of total probability (Hall, Blockley and Davis 1998). A proposition $E$ on the space (the *universe of discourse*) defines a partition ($E$, $E^C$) of $S$. The interval probability:

$$\mathrm{IP}(E) = [l(E), u(E)] \tag{6.37}$$

gives an *Italian flag* for the proposition, i.e. a partition of the interval [0, 1] in the sub-intervals *green* $[0, l(E)]$, *white* $[l(E), u(E)]$ and *red* $[u(E), 1]$.

It is equivalent to a random set on the power set of the set $\{E, E^C\}$, that is the set of the focal elements $\{E, E^C, S\}$, with the probabilistic assignment (derived for example by incomplete information given by experts on the true-value of the proposition $E$):

$$m(E) = l(E)\ ;\ \ m(E^C) = 1 - u(E)\ \ ;\quad m(S) = u(E) - l(E) \tag{6.38}$$

All results obtained through IP can therefore be obtained by using random sets and on the other hand random sets theory can provide insight into the rules of IP (Bernardini 2000).

## 6.3 Combination of Random Sets

### 6.3.1 Evidence Theory: Dempster's Rule of Combination

Bayes theorem can be considered as a particular case of a more general rule of combination suggested by Dempster (Dempster 1967). Let $\mathcal{F}_1 = \left\{A_1^i, m_1\left(A_1^i\right)\right\}$ and $\mathcal{F}_2 = \left\{A_2^i, m_2\left(A_2^i\right)\right\}$ be two random sets on the same space $S$. Dempster's rule defines the combined random set as:

$$\mathcal{F}_{12} = \left( C^{i,j} = A_1^i \cap A_2^j ; m_{12}\left(C^{i,j}\right) = \frac{m_1\left(A_1^i\right) \times m_2\left(A_2^j\right)}{K} \right) ;$$
$$K = \sum_{C^{i,j} \neq \varnothing} m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) = 1 - \sum_{C^{i,j} = \varnothing} m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) \qquad (6.39)$$

Hence the result is a third random set, whose focal elements (according to Eq. (6.1)) are the intersections of the initial focal elements and whose probability assignment is obtained as the product of the corresponding probabilities (following the hypothesis of independence by which they were initially evaluated), normalized in order to take into account that a part of the initial evidence (non null products of probabilities) may focus on empty intersections.

Of course if some focal elements $C^{i,j}$ are coincident, their probabilistic assignments $m_{12}(C^{i,j})$ can be added together.

**Example 6.8.** Let us return to Example 6.4 and suppose that both sclerometric and mix composition tests, performed independently (i.e. in a non-joint way), yield evidence about concrete resistance $r$ in the form of the following (non-consonant) focal elements and probability assignments:

$A$ : $m(A^1 = \{r^1, r^2\}) = 0.2$ ; $m(A^2 = \{r^1, r^2, r^3\}) = 0.7$ ; $m(A^3 = \{r^2, r^3\}) = 0.1$
$B$ : $m(B^1 = \{r^1\}) = 0.05$ ; $m(B^2 = \{r^1, r^2\}) = 0.1$ ; $m(B^3 = \{r^2, r^3, r^4\}) = 0.85$

What conclusions can be drawn about the probability $P(C)$ of any subset $C \subseteq R$ ? Or what conclusions can be drawn, at least, as far as the interval [$Bel$ ($C$), $Pla$($C$)] containing $P(C)$ is concerned?

Shafer proposed to utilize the rule summarized in Eq. (6.39), obtaining in this numerical example:

$$\begin{aligned}
m(A^1 \cap B^1) &= m(\{r^1\}) &&= 0.2 \text{ x } 0.05 &&= 0.01 \\
m(A^1 \cap B^2) &= m(\{r^1, r^2\}) &&= 0.2 \text{ x } 0.1 &&= 0.02 \\
m(A^1 \cap B^3) &= m(\{r^2\}) &&= 0.2 \text{ x } 0.85 &&= 0.17 \\
m(A^2 \cap B^1) &= m(\{r^1\}) &&= 0.7 \text{ x } 0.05 &&= 0.035 \\
m(A^2 \cap B^2) &= m(\{r^1, r^2\}) &&= 0.7 \text{ x } 0.1 &&= 0.07 \\
m(A^2 \cap B^3) &= m(\{r^2, r^3\}) &&= 0.7 \text{ x } 0.85 &&= 0.595
\end{aligned}$$

$$m(A^3 \cap B^1) = m(\varnothing) = 0.1 \text{ x } 0.05 = 0.005$$
$$m(A^3 \cap B^2) = m(\{r^2\}) = 0.1 \text{ x } 0.1 = 0.01$$
$$m(A^3 \cap B^3) = m(\{r^2, r^3\}) = 0.1 \text{ x } 0.85 = 0.085$$

and then : $K = 1 - 0.005 = 0.995$

$$m(\ C^1 = \{r^1\}) = (0.01 + 0.035) / 0.995 = 0.0452$$
$$m(\ C^2 = \{r^1, r^2\}) = (0.02 + 0.07) / 0.995 = 0.0905$$
$$m(\ C^3 = \{r^2\}) = (0.17 + 0.01) / 0.995 = 0.1810$$
$$m(\ C^4 = \{r^2, r^3\}) = (0.595 + 0.085) / 0.995 = 0.6832$$

**Fig. 6.4** Plausibility and Belief of the singletons $\{r^i\}$

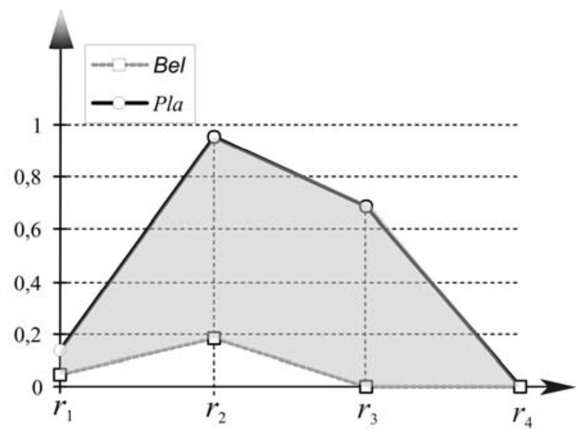



Figure 6.4 and Figure 6.5 portray the corresponding intervals in which the probabilities of the singletons and the cumulative probabilities are included.

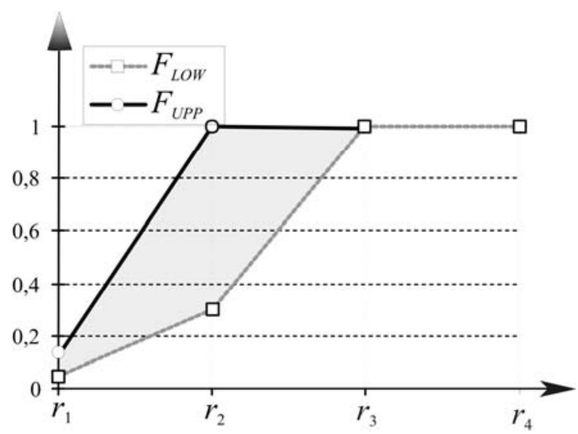



**Fig. 6.5** Upper and lower cumulative distribution functions

Dempster's rule coincides exactly with Bayes' rule when the first body of evidence is probabilistic (the focal elements are all singletons and hence their intersections are empty) and the information conveyed by the second body of evidence is deterministic (an event $B$ has been observed with complete certainty). In this case Eq.(6.39) gives:

$$\begin{aligned}
&\mathcal{F}_{12} = \left\{ C^i = s^i \cap B \; ; m_{12}\left(C^i\right) = \frac{\left(m_1 = P\left(s^i\right)\right)\cdot\left(m_2\left(B\right) = 1\right)}{K} \right\} \; ; \\
&K = \sum_{C^i \neq \varnothing} P\left(s^i\right)\cdot 1 = P\left(B\right) \\
&Bel\left(A/B\right) = Pla\left(A/B\right) = \sum_{C^i \subseteq A} m_{12}\left(C^i\right) = \sum_{C^i \cap A \neq \varnothing} m_{12}\left(C^i\right) = \\
&= \frac{P\left(A/B\right)}{P\left(B\right)}
\end{aligned} \tag{6.40}$$

**Example 6.9.** We can verify this fact in Example 6.5, by assuming that the initial body of evidence is given by the joint probability distribution $p(c, r, s)$. The focal elements are nothing else but the 64 singletons with probability assignment $m(A^{ijk} = \{(r^i, s^j, c^k)\}) = p\,(r^i, s^j, c^k)$ ; while deterministic information is defined by the unique focal element:

$$m\,(B = \{r^1, r^2, r^3, r^4\} \times \{s^3\} \times \{c^2\}) = 1$$

The non-empty intersections are just the four focal elements, with joint probability (Table 6.5, row 4):

$$m\,(C^i = \{(r^i, s^3, c^2)\} \cap B = \{(r^i, s^3, c^2)\}) = p(r^i, s^3, c^2)$$

The empty intersections are the remaining 60 focal elements, whose probability summation is obtained as 1 minus the sum of the probabilities of the previous four focal elements. Therefore it yields simply:

$$K = 1 - (1 - 0.0680) = 0.0680$$

Thus, after *a posteriori* information, the probability assignment of the 4 focal elements assumes the values $m(C^i) = \mathrm{p}\,(r^i / s^3, c^2)$ of Table 6.8, exactly coinciding with those of Table 6.5, last row.

**Table 6.8** Joint distribution and marginals in Example 6.9

| | $m\,(C^i) = p(r/s^3, c^2)$ |
|---|---|
| $r^1$ | 0 |
| $r^2$ | 0.030 / 0.068 = 0.441 |
| $r^3$ | 0.036 / 0.068 = 0.530 |
| $r^4$ | 0.002 / 0.068 = 0.029 |
| **Total** | **0.068 / 0.068 = 1.000** |

However, when Dempster' rule is applied to combine a body of evidence measured by a random set and an observed deterministic event $B$, the resulting belief functions do not coincide with the exact results given by Eq. (3.35). In this case Eq.(6.39) gives:

$$
\begin{aligned}
&\mathcal{F}_{12} = \left\{ C^i = A^i \cap B\ ; m_{12}\left(C^i\right) = \frac{m_1\left(A^i\right)\cdot 1}{K} \right\}\ ; \\
&K = \sum_{C^i = A^i \cap B \neq \varnothing} m_1\left(A^i\right)\cdot 1 = Pla\left(B\right) \\
&Pla\left(A/B\right) = \sum_{C^i \cap A \neq \varnothing} m_{12}\left(C^i\right) = \sum_{A^i \cap B \cap A \neq \varnothing} m_{12}\left(C^i\right) = \sum_{A^i \cap \left(A \cap B\right) \neq \varnothing} \frac{m_1\left(A^i\right)}{K} = \\
&= \frac{Pla\left(A \cap B\right)}{Pla\left(B\right)}
\end{aligned}
\tag{6.41}
$$

On the other hand Dempster's rule is commutative when successively conditioning a random set to a list of observed deterministic events $B^1$, $B^2$, … and the final result can be again obtained conditioning the random sets to their intersection $B^1 \cap B^2 \cap \ldots$ .

**Example 6.10.** Let us consider again Example 6.6 and calculate the random set conditional to $B^1 = \{ r^1, r^2, r^3\}$.

$$
\begin{aligned}
&A^1 \cap B^1 = \{r^1\}\ ;\ m(A^1) \text{ x } m(B^1) = 0.25 \text{ x } 1 = 0.25 \\
&A^2 \cap B^1 = \{r^2\}\ ;\ m(A^2) \text{ x } m(B^1) = 0.25 \text{ x } 1 = 0.25 \\
&A^3 \cap B^1 = \{r^3\}\ ;\ m(A^2) \text{ x } m(B^1) = 0.5 \text{ x } 1 = 0.5
\end{aligned}
$$

Hence $K = 1$ and the obtained random set is a probability distribution on the singletons of the event $B^1$.

Conditioning the prior random set to $B^2 = \{ r^1, r^2 \}$ gives:

$$
\begin{aligned}
&A^1 \cap B^2 = \{r^1\}\ ;\ m(A^1) \text{ x } m(B^2) = 0.25 \text{ x } 1 = 0.25 \\
&A^2 \cap B^2 = \{r^2\}\ ;\ m(A^2) \text{ x } m(B^2) = 0.25 \text{ x } 1 = 0.25 \\
&A^3 \cap B^2 = \varnothing\ ;\ m(A^2) \text{ x } m(B^2) = 0.5 \text{ x } 1 = 0.5
\end{aligned}
$$

Hence $K = 0.5$ and the obtained random set is again a probability distribution (the white distributions on the singletons of the event $B^2$.

Of course the same results could be obtained conditioning to $B^2$ the above calculated random sets conditional to $B^1$, because $B^2$ is included by $B^1$.

Comparison with Example 6.6 shows that in this particular case the final results (in both cases obtained conditioning to $B^2$) coincide. However the reduction of uncertainty obtained through Dempster's rule is generally much stronger: for example, in the case here considered, conditioning to $B^1$ the exact Eq. (3.33) gives $P(r^1) = P(r^2)$=[0.25, 0.5] while Dempster's rule gives $P(r^1) = P(r^2)$=0.25.

The hypothesis of independence of the combined bodies of evidence implicit in Dempster's rule is a further constraint that is generally unjustified.

More generally, Dempster's rule is commutative when combining a list $\mathcal{F}_1, \mathcal{F}_2, \ldots \mathcal{F}_k$ of random sets on the same space S. The property derives directly from the commutativity of set intersection and product of numbers. The final result is:

$$\mathcal{F}_{12\ldots k} = \left( \begin{array}{l} C^{i,j,\ldots,m} = A_1^i \cap A_2^j \ldots A_k^m \,;\, m_{12\ldots k}\left(C^{i,j,\ldots,m}\right) = \\ \dfrac{m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) \times \ldots \times m_k\left(A_k^m\right)}{K} \end{array} \right) ;$$

$$K = \sum_{C^{i,j,\ldots,m} \neq \varnothing} m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) \times \ldots \times m_k\left(A_k^m\right) = \tag{6.42}$$

$$= 1 - \sum_{C^{i,j,\ldots,m} = \varnothing} m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) \times \ldots \times m_k\left(A_k^m\right)$$

As a comment about Dempster's rule, it can be observed that its first proposal dates back to the 18th century and precisely to Lambert's speculation about the Theory of Chances. In an independent way, it was propounded by Dempster in 1967 and since then posed as a basis for Shafer's Theory of Evidence.

However this rule has not been free of criticism deriving especially from the observation that independence of bodies of knowledge does not necessarily imply stochastic independence.

Some paradoxes have been worked out by applying Dempster's rule, as reported in the next section, and alternative solutions have been proposed, even if frequently less simple and effective.

### 6.3.2 *A Critical Discussion of Dempster's Rule: Yager's Rule of Combination*

Dempster's rule maximizes the "points in common" between the two sources of information because it is based on the rule of intersection (Case 1 in Section 6.1.1.1; AND operator), and therefore it breaks down when all the intersections of the assigned focal elements are empty. As a consequence, applications in which intersections are nearly empty ($K \ll 1$) are outside its intended use, and it is no surprise that conclusions could appear questionable or paradoxical.

**Example 6.11.** Let us consider the combination of two independent diagnoses for neurological symptoms in a patient proposed by Zadeh in 1984 (Zadeh 1984):

$\mathcal{F}_1 = \{(A^1 = \{\text{meningitis}\},\ m_1(A^1) = 0.99),\ (A^2 = \{\text{brain tumor}\}, m_1(A^2) = 0.01)\ )\}$
$\mathcal{F}_2 = \{(B^1 = \{\text{concussion}\},\ m_2(B^1) = 0.99),\ (B^2 = \{\text{brain tumor}\}, \mathrm{m}_2(B^2) = 0.01)\ )\}$

Dempster's rule combines the diagnosis giving:

$$\mathcal{F}_{12} = \left\{ C^{2,2} = A^2 \cap B^2 = \{\text{brain tumor}\}; m_{12}\left(C^{2,2}\right) = \frac{0.01 \times 0.01}{K} \right\}$$

$K = 1 - 0.99\text{x}0.99 - 2\ \text{x}\ (0.99\ \text{x}\ 0.01) = 0.01\ \text{x}\ 0.01$

Therefore we obtain the paradoxical certain conclusion: $\mathrm{m}_{12}(\{\text{brain tumor}\}) = Bel(\{\text{brain tumor}\}) = Pla(\{\text{brain tumor}\}) = 1$.

Yager (Yager 1987) modified Dempster's rule, by observing that the probabilities of the empty intersections should be used to increase the probabilistic assignment of the space $S$, instead of normalizing the probabilities of the non-empty intersections:

$$\mathcal{F}_{12} = \left\{ \begin{array}{l} C^{i,j} = A_1^i \cap A_2^j, m_{12}\left(C^{i,j}\right) = m_1\left(A_1^i\right) \times m_2\left(A_2^j\right); \\ S\ , m_{12}(S) = m_1(S) \times m_2(S) + \sum\limits_{C^{i,j} = \varnothing} m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) \end{array} \right\} \tag{6.43}$$

**Example 6.12.** Let us reconsider the above-mentioned example:

$$\mathcal{F}_{12} = \left( \begin{array}{l} C^{2,2} = A^2 \cap B^2 = \{\text{brain tumor}\}; m_{12}\left(C^{2,2}\right) = 0.01 \times 0.01 = 1 \cdot 10^{-4} \\ S = \{\text{meningitis,brain tumor, concussion}\}; m_{12}(S) = 1 - 1 \cdot 10^{-4} \end{array} \right)$$

Hence the much more uncertain forecasting is obtained:

$Bel(\{\text{brain tumor}\}) = 10^{-4} < P(\{\text{brain tumor}\}) < Pla(\{\text{brain tumor}\}) = 1$
$Bel(\{\text{meningitis}\}) = 0 < P(\{\text{meningitis}\}) < Pla(\{\text{meningitis}\}) = 1 - 10^{-4} = 0.9999$
$Bel(\{\text{concussion}\}) = 0 < P(\{\text{concussion}\}) < Pla(\{\text{concussion}\}) = 1 - 10^{-4} = 0.9999$

Dubois and Prade (Dubois and Prade 1987 ) observed that the combination of strongly conflicting information requires some hypothesis to explain why the conflict occurs (e.g. the sources of information are not reliable; the considered reference set is not exhaustive; the sources do not speak about the same event). If the sources of information are not reliable, Yager's rule (totally conflicting information is unreliable) or some application of Case 2 in Section 6.1.1.2 (totally conflicting information support the union of the disjoint subsets) could be more adequate.

A reasonable trade-off could be to assume that the sources of information are reliable when they are not in conflict (i.e. when a non-empty intersection exists): this is Case 1 in Section 6.1.1.1. Likewise, it is a reasonable trade-off to assume that one of the sources of information is right when they are totally conflicting (i.e. when the intersection is empty): this is Case 2 in Section 6.1.1.2. Hence:

$$\mathcal{F}_{12} = \left\{ \begin{array}{l} C^{i,j} = \left\langle \begin{array}{ll} A_1^i \cap A_2^j & \text{if } A_1^i \cap A_2^j \neq \varnothing \\ A_1^i \cup A_2^j & \text{if } A_1^i \cap A_2^j = \varnothing \end{array} \right. \\ m_{12}\left(C^{i,j}\right) = m_1\left(A_1^i\right) \times m_2\left(A_2^j\right) \end{array} \right\} \tag{6.44}$$

**Example 6.13.** Let us consider again the above mentioned example and combine the diagnoses according to the above suggested criterion:

$$\mathcal{F}_{12} = \left\{ \begin{array}{l} \left(C^{2,2} = A^2 \cap B^2 = \{\text{brain tumor}\}, m_{12}\left(C^{2,2}\right) = 0.01 \times 0.01 = 1 \cdot 10^{-4}\right), \\ \left(C^{1,2} = \{\text{meningitis,brain tumor}\}, m_{12}\left(C^{1,2}\right) = 0.99 \times 0.01 = 0.99 \cdot 10^{-2}\right), \\ \left(C^{2,3} = \{\text{brain tumor, concussion}\}, m_{12}\left(C^{2,3}\right) = 0.99 \times 0.01 = 0.99 \cdot 10^{-2}\right) \end{array} \right\}$$

Therefore:

$$\begin{array}{lll} Bel(\{\text{brain tumor}\}) = 10^{-4} & < P(\{\text{brain tumor}\}) & < Pla(\{\text{brain tumor}\}) = 0.0199 \\ Bel(\{\text{meningitis}\}) = 0 & < P(\{\text{meningitis}\}) & < Pla(\{\text{meningitis}\}) = 0.9900 \\ Bel(\{\text{concussion}\}) = 0 & < P(\{\text{concussion}\}) & < Pla(\{\text{concussion}\}) = 0.9900 \end{array}$$

The highest probabilistic assignment to the union {meningitis, concussion} strongly reduces the plausibility of {brain tumor} with respect to both Dempster's and Yager's rule.

## 6.4 Fuzzy Logic and Fuzzy Composition Rule

### *6.4.1 Introduction*

Dempster's rule holds even in the case of consonant structures, i.e. when random sets or random relations can be completely defined by means of membership functions (possibility distributions) of fuzzy sets or fuzzy relations. However, the result of the combination of consonant structures is generally a non-consonant structure.

**Example 6.14.** With reference to Example 6.8, let us consider the combination of the following consonant structures of data relative to the concrete resistance *r*, derived from information about mix composition and sclerometric tests respectively:

$A$ : $m(A^1 = \{r^1, r^2, r^3, r^4\}) = 0.4$ ; $m(A^2 = \{r^2, r^3\}) = 0.6$
$B$ : $m(B^1 = \{r^2, r^3, r^4\}) = 0.2$ ; $m(B^2 = \{r^3, r^4\}) = 0.8$

They correspond to fuzzy sets whose membership functions (possibility distributions) are displayed in Figure 6.6.

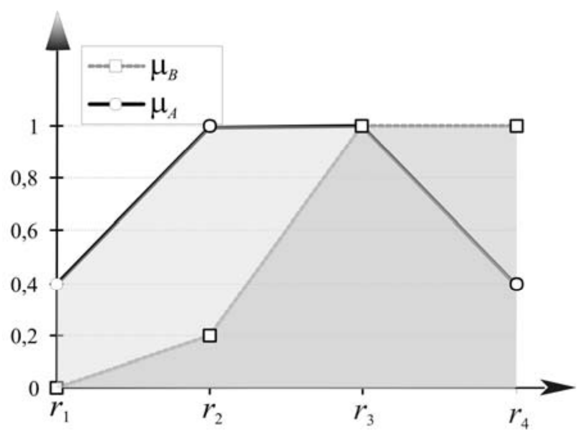



**Fig. 6.6** Membership functions of consonant random sets *A* and *B*

By using Dempster's rule, the combination is defined by the following focal elements and relative probability assignments, which clearly do not represent a consonant structure; in this case, it is simply $K = 1$.

$$A^1 \cap B^1 = B^1, \quad m(A^1) \text{ x } m(B^1) = 0.4 \text{ x } 0.2 = 0.08$$
$$A^1 \cap B^2 = B^2, \quad m(A^1) \text{ x } m(B^2) = 0.4 \text{ x } 0.8 = 0.32$$
$$A^2 \cap B^1 = A^2, \quad m(A^2) \text{ x } m(B^1) = 0.6 \text{ x } 0.2 = 0.12$$
$$A^2 \cap B^2 = \{r^3\}, \quad m(A^2) \text{ x } m(B^2) = 0.6 \text{ x } 0.8 = 0.48$$

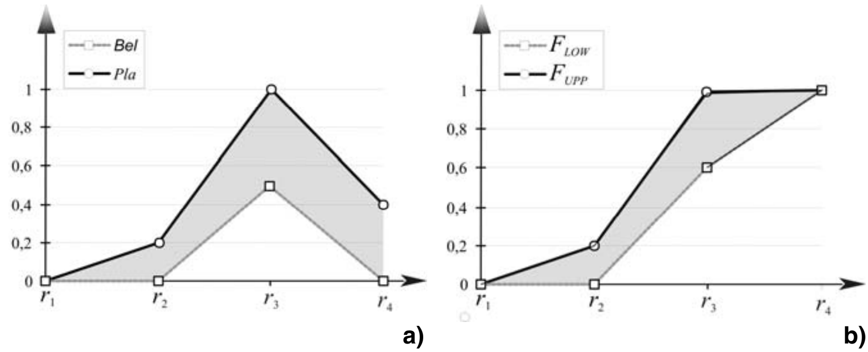



**Fig. 6.7** Example 6.14: **a)** Plausibility and Belief of the singletons $r^i$; **b)** Upper and lower cumulative distribution functions

In Figure 6.7a and b the values of Plausibility and Belief of the single values and, respectively, the upper and lower bounds of cumulative distributions are plotted.

Certainly, a solution is preferable that maintains consonance and allows the development of a theory of approximate reasoning completely within fuzzy set theory; this is even more valuable if one remembers that any non-consonant structure can be enclosed in an approximate consonant structure.

### *6.4.2 Fuzzy Extension of Set Operations*

From a historical point of view, the problem of the combination of uncertain information conveyed by fuzzy sets has been developed by Zadeh since the origins of fuzzy set theory. This has been performed in an autonomous way with respect to Bayesian procedures or procedures developed within evidence theory, as pointed out in Section 3.2.4.

Within fuzzy logic, the fundamental problem of combining information pertaining to two distinct bodies of evidence has been frequently interpreted as the application of the union and complementation operations as extended to fuzzy sets, thus extending Eq. (6.1) and (6.11) when $A$ and $B$ are two fuzzy sets. The rules of classical logic can be naturally extended to fuzzy sets when the membership function of a fuzzy set is substituted for the characteristic function of a classical crisp set.

These operations, when applied to crisp sets, can be performed through different operators applied to their characteristic functions (see Notes N 6-1 and N 6-2), obtaining the same results. On the contrary when applied to the membership functions of fuzzy sets (even when they are chosen in a dual way, i.e. respecting De Morgan's Laws) they give different results. Hence important difficulties arise over their interpretation.

For this reason, fuzzy logic has sometimes been charged with incoherence, explicitly by scholars educated within the probabilistic paradigm.

A vast array of *t*-norms and *t*-conorms have been proposed for modeling the AND and OR operators (e.g., see (Zimmermann 1991), (Klir and Yuan, 1995)). However, in the following, reference is maintained to the so-called *standard operators* (min for intersection and max for union) employed for the basic approaches defined in Section 6.1.1. The operators are coherent with De Morgan's Laws when the rule for the complementation of fuzzy sets is the direct extension of the rule for the characteristic functions of complementary crisp sets:

$$\mu_{A^C}(s) = 1\text{-}\mu_A(s) \tag{6.45}$$

Indeed, only these operators are coherent within the general theory of uncertainty described in this book, which unifies probability and fuzziness through the concept of non consonant and consonant random sets. As discussed in Section 5.2.1, this choice of the operators for the dual operations of union/intersection of fuzzy sets is the only one compatible with the extension principle of random sets when applied to consonant random sets.

Hence the extension for the 3 cases discussed in Section 6.1.1 follows immediately.

**Case 1:** The AND operator leads to the intersection $C$ of sets $A$ and $B$. The natural extension of Eq. (6.1) to fuzzy sets is obtained by substituting its membership functions (or the associated possibility distributions) for the characteristic functions:

$$\mu_C(s) = \min\left(\mu_A(s), \mu_B(s)\right) \tag{6.46}$$

Again this combination leads to a total loss of information when $A$, $B$ are totally conflicting (i.e. the fuzzy intersection $A \cap B = \varnothing$). When the conflict is partial ($A \cap B \neq \varnothing$), the rule decreases the uncertainty for the decision-maker; but again this decrease could be unjustified and unrealistic if the sources of information are not very reliable.

**Case 2:** The OR operator leads to the union $C$ of the fuzzy sets $A$ and $B$ . Therefore Eq. (6.11) becomes:

$$\mu_C(s) = \max\left(\mu_A(s), \mu_B(s)\right) \tag{6.47}$$

The resulting uncertainty for the decision-maker increases, but the rule works with every pair of subsets (even if they are totally conflicting).

**Case 3:** CONVOLUTIVE AVERAGING (c-Averaging): a total or partial ordering is recognized or assigned in the space *S*. In a Euclidean space, let $x_A$ and $x_B$ be elements of the fuzzy sets *A* and *B*, respectively. One obtains, for fuzzy sets, the fuzzy extension of Eq. (6.13), in a coherent manner with the extension principle for consonant not-interactive random sets. Considering for example the one-dimensional real space *S*:

$$\mu_C(s) = \sup_{s=\frac{s_A+s_B}{2}} \min\left(\mu_A(s_A), \mu_B(s_B)\right) \tag{6.48}$$

### 6.4.3 Fuzzy Composition Rule

It seems more natural to derive the composition rule for consonant structures as an "extension of the extension principle", i.e. as an extension of the rule of composition for information constrained by consonant structures (information A) with deterministic information (information B) conveyed by a point-valued or a set-valued function.

When information B is expressed by means of a consonant structure (a fuzzy relation, i.e. a fuzzy subset of the space *S* x *Z*, the natural extension of both Eq. (6.4) and Eq. (6.30) is given again by the intersection operator, i.e. by the following *fuzzy composition rule*:

$$\mu_C(s,z) = \min\left(\mu_A(s,z), \mu_B(s,z)\right) \tag{6.49}$$

The marginal fuzzy set on Z can finally be evaluated with the max (Eq. (6.5) or sup (Eq. (6.31) operator.

**Example 6.15.** Let us suppose that information about a concrete mix is given by means of a fuzzy set whose values $\mu_A(c)$ are listed in Table 6.9.

**Table 6.9** Fuzzy set measuring the mix composition *c* in Example 6.15

| $\mu_A(c^1)$ | $\mu_A(c^2)$ | $\mu_A(c^3)$ | $\mu_A(c^4)$ |
|---|---|---|---|
| 0.1 | 1 | 0.4 | 0 |

The relation between concrete resistance *r* and mix composition *c* is given by the fuzzy relation whose values $\mu_B(r, c)$ can be found in Table 6.10, demonstrating a strong monotone relation between the variables (membership equal to 1 on the main diagonal of the matrix, with decreasing, nearly symmetric membership out of the diagonal cells).

By combining the two pieces of information, $A$ and $B$, we obtain the consonant structure (fuzzy set) defined by the values $\mu_C(r) = \max \min(\mu_A(c), \mu_B(r, c))$, as shown in Table 6.11.

**Table 6.10** Fuzzy relation between concrete resistance $r$ and mix composition $c$ in Example 6.15

| | $\mu_B(r, c^1)$ | $\mu_B(r, c^2)$ | $\mu_B(r, c^3)$ | $\mu_B(r, c^4)$ |
|---|---|---|---|---|
| $r^1$ | 1 | 0.5 | 0 | 0 |
| $r^2$ | 0.4 | 1 | 0.5 | 0.1 |
| $r^3$ | 0.1 | 0.4 | 1 | 0.6 |
| $r^4$ | 0 | 0 | 0.4 | 1 |

**Table 6.11** Marginal fuzzy set measuring concrete resistance $r$ in Example 6.15

| | $\mu(r^i)$ |
|---|---|
| $r^1$ | max( min(1, 0.1), min (0.5,1), min(0,0.4), min (0,0)) = 0.5 |
| $r^2$ | 1 |
| $r^3$ | 0.4 |
| $r^4$ | 0.4 |

In the case of two fuzzy sets that induce the possibility distributions $\pi_A(s) = \mu_A(s)$ and $\pi_B(s) = \mu_B(s)$, gained in an independent way, on the same space $S$, the fuzzy composition rule is simply expressed by the equation:

$$\mu_C(s) = \min\left(\mu_A(s), \mu_B(s)\right), \tag{6.50}$$

which coincides with the standard definition of intersection $C = A \cap B$ of two fuzzy sets $A$ and $B$ (Eq. (6.46)) and therefore it is the extension of Eq. (6.1). Moreover, with the standard definition of complementation the extension of (6.11) is as follows

$$\mu_C(s) = 1 - \min\left(1 - \mu_A(s), 1 - \mu_B(s)\right) = \max\left(\mu_A(s), \mu_B(s)\right) \tag{6.51}$$

which coincides with the standard definition of union $C = A \cap B$ of two fuzzy sets $A$ and $B$ (Eq. (6.47)) and therefore it is the extension of Eq. (6.11).

**Example 6.16.** With reference to Example 6.14 previously solved by means of Dempster's rule, equation (6.50) gives the result depicted in Figure 6.8.

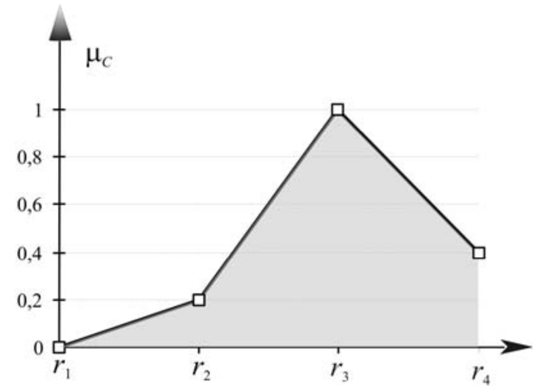



**Fig. 6.8** Membership function of the combination of consonant random sets plotted in Figure 6.6, obtained by means of the fuzzy composition rule. The resulting fuzzy set is the intersection of the fuzzy sets displayed in Figure 6.6

This combination has the following consonant structure:

$m(C^1 = \{r^2, r^3, r^4\}) = 0.2$ ;
$m(C^2 = \{r^3, r^4\}) = 0.2$ ;
$m(C^3 = \{r^3\}) = 0.6$

whose probabilistic content is shown in Figure 6.9 and is therefore different from that of Figure 6.7. Although upper/lower cumulative distribution functions (Figure 6.7b and Figure 6.9b) and contour functions (plausibility of the singletons in Figure 6.7a and Figure 6.9a) coincide, the belief of the singleton $\{r^3\}$ increases from 0.48 in Figure 6.7a to 0.6 in Figure 6.9a: a considerable reduction of the uncertainty about its probability.

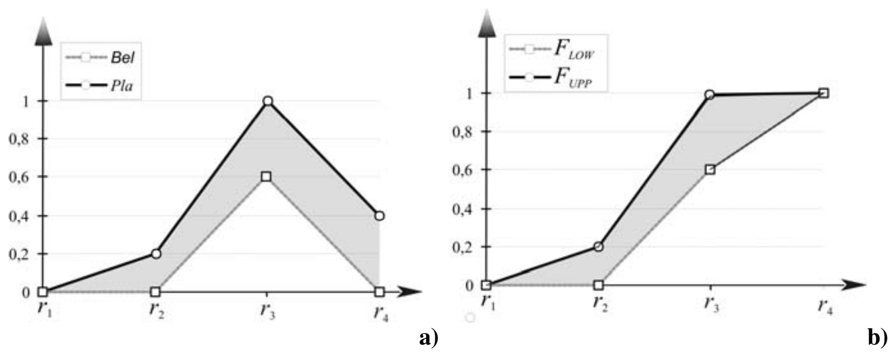



**Fig. 6.9** Example 6.16**: a)** Plausibility and Belief of the singletons $r^i$; **b)** Upper and lower cumulative distribution functions

It is worth noting that both the result obtained by means of Dempster's rule and the result yielded by the fuzzy composition have to be considered reasonable representations of the evidence combined on the basis of the available information, but not actual frequentistic forecasts of objectively observable occurrences, even when the basic data are such, for example separately given by random sets.

The fuzzy composition rule can also be expressed through a rule for combining consonant random sets.

Let $\mathcal{F}_1 = \left\{A_1^i, m_1\left(A_1^i\right)\right\}$ and $\mathcal{F}_2 = \left\{A_2^i, m_2\left(A_2^i\right)\right\}$ be two consonant random sets on the same finite set $S$ and $F_1$, $F_2$ their associated *normal* fuzzy sets, whose membership functions take values in the common finite set $\{\alpha^1 < \ldots \alpha^{ij} \ldots < \alpha^m < \alpha^{m+1} = 0\}$. As explained in Section 3.2.4, the connection between the two representations is obtained through the common finite set of $\alpha$–cuts ${}^{\alpha}F$ (not necessarily distinct for each $k$) as follows:

$$\mathcal{F}_k = \left\{A_k^{ij} = {}^{\alpha^{ij}}F_k, m_k\left(A_k^{ij}\right) = \alpha^{ij} - \alpha^{ij+1}\right\}, k = 1, 2 \tag{6.52}$$

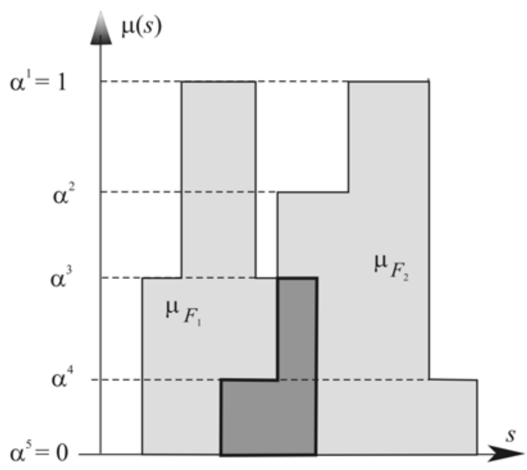



**Fig. 6.10** Intersection of two normal fuzzy sets with a finite number of α-cuts

The rule is as follows (see Figure 6.10)

$$\mathcal{F}_{12} = \left\{\begin{array}{l} \text{for } ij = 1, 2, \ldots, m: \quad C^{ij} = {}^{\alpha^{ij}}F_1 \cap {}^{\alpha^{ij}}F_2, \\ m_{12}\left(C^{ij}\right) = m_1\left({}^{\alpha^{ij}}F_1\right) == m_1\left({}^{\alpha^{ij}}F_1\right) = \alpha^{ij} - \alpha^{ij+1}; \end{array}\right\} \tag{6.53}$$

Hence the focal elements of the combined random set are again, as in Dempster's rule, obtained by intersection, but the intersection is performed only by the $\alpha$–cuts of the same level $\alpha$. Moreover the probabilistic assignment of the intersected focal elements is preserved[N 6-4].

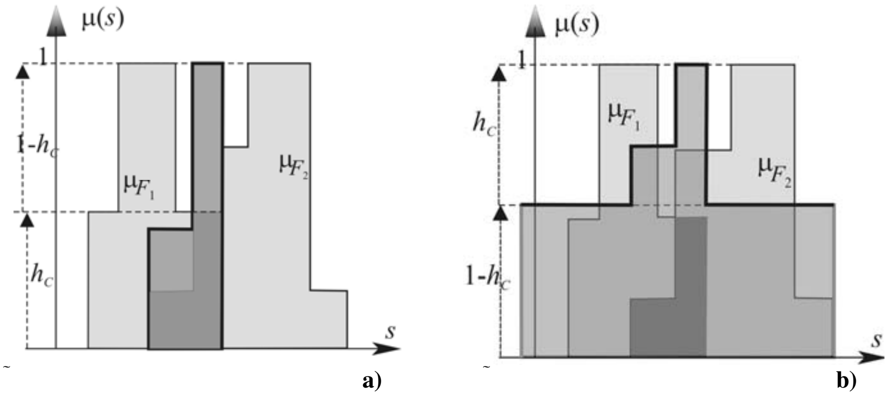



**Fig. 6.11 a)** – Dempster's normalization of a fuzzy intersection. **b)** – Yager's normalization of a fuzzy intersection

It can be observed that the rule in Eq. (6.53) uses Case 1 (intersection); therefore, if $A_1^m \cap A_2^m = \varnothing$, intersections are void $\forall\, ij$ and hence the rule does not work. Moreover, if $A_1^1 \cap A_2^1 = \varnothing$, the resulting fuzzy set $C$ is *sub-normal***,** i.e. its *height* $h(C) < 1$, and $K$=1-$h(C)$ is the probability assignment of the empty set $\varnothing$.

The problem of a normalization of the resulting fuzzy set appears in exactly same manner as discussed in the application of Dempster's rule. Hence alternative rules can again be used for normalization, following the Bayes/Dempster/Shafer criterion (Figure 6.11a)) or the modification proposed by Yager according to Eq. (6.43) (Figure 6.11b)).

## 6.5 Fuzzy Approximate Reasoning

### *6.5.1 Introduction*

Although the theory of random sets and the connected interpretation of a fuzzy set as a consonant random set can be very useful in applications (as an extension of the classical probability theory to the case of incomplete or set-valued data), the most important applications of fuzzy set theory have been based, since its foundation in the 1960s, on the powerful extension of the classical logic and of the classical set theory, to give a numerical description of vague and qualitative information, and also to combine in a very simple and expressive manner uncertain information independently given on the same space.

The main problem is to make decisions, on the basis of the forecast of future events, to design a new system or to operate the control of an existing system, in situations of uncertainty, i.e. when two or more alternatives are possible, and a list of objectives or constraints (safety, financial cost, serviceability) is defined.

The basic instruments are firstly the rule of fuzzy composition discussed in Section 6.4, to extend the classical rules of inference, and secondly the definition of an optimal choice when many, generally contrasting, objectives or constraints should be taken into account. The first instrument can be used to develop fuzzy rule based expert systems and fuzzy on line controller of dynamic systems (this may be the most popular application of the theory). The second instrument suggests powerful, simple and robust procedures in the fields of pattern recognition, clustering and multi-objective optimization.

Many thousands of papers and books and many hundreds of alternative procedures have been written or proposed in the last 40 years. So in the following sections just some introductory ideas are given, trying to clarify the most relevant conceptual aspects, without discussing or classifying the different algorithms and operators proposed to extend the classical logic and set theory.

It is important here to underline once again that in these applications, the obtained results cannot be interpreted in the sense of expected frequencies of objective phenomena, even in the case when the data to be combined derived from statistics, although incomplete or set-valued, of objective phenomena.

In comparison with fuzzy logic, Dempster's rule (Section 6.3.1) has very rarely been used in real applications, despite its more extended validity to consonant and non-consonant data; perhaps the reason depends on the computational difficulties arising from the necessity to operate on each focal element, while fuzzy logic allows a point-valued representation.

### 6.5.2 Inference from Conditional Fuzzy Propositions

The implementation of expert systems and online computer-aided system controllers require the development of quick procedures of automatic decision reproducing in any manner the capacity of the human brain to recognize, in a largely uncertain environment, the most relevant information, to compare objects, to evaluate rules of general (but not absolutely universal) validity to be used in the approximate reasoning. When applied to very complex systems the required techniques should be simple and computationally robust, to directly evaluate the main structures and regularity of the data, without passing through an accurate analysis of any particular bit of information.

From this point of view the theory of fuzzy sets seems a particularly powerful instrument, because it enables one to demonstrate, in a very condensed manner, the informative content subtended by a population of individually distinct objects or measures.

As an example a fuzzy relation can summarize a vague or qualitative monotonically increasing dependency between two variables ($x$, $y$), expressed in a linguistic manner by the propositions:

IF $x$ is SMALL (a fuzzy set $A_1$ on $X$) THEN
y is MEDIUM (a fuzzy set $B_1$ on $Y$)

..............

IF $x$ is LARGE (a fuzzy set $A_j$ on $X$) THEN
$y$ is VERY LARGE (a fuzzy set $B_j$ on $Y$)

..............

and numerically:

$$\mu_R(x, y) = \max_j \min\left(\mu_{A_j}(x), \mu_{B_j}(y)\right) \tag{6.54}$$

This rule can be justified observing that the min operator combines, according to the fuzzy composition rule (6.49), the cylindrical extensions of $\mu_{A_j}(x)$ and $\mu_{B_j}(y)$ from $X$ to $X$ x $Y$ and from $Y$ to $Y$ x $X$ respectively, while the max operator gives the standard union of the obtained $R_j$ fuzzy relations on $X$ x $Y$ = $Y$ x $X$ (according to (6.51)).

The rule of fuzzy composition (symbol ° in the following) discussed in Section 6.4.3 builds up the basis for a model of approximate reasoning, extending the rules that in classical logic are given to infer the “truth value” of a dependent proposition. This is performed by combining one or more propositions of universal validity (a deterministic relation) and the evidence of a particular property.

1. The extended *Modus Ponens*:

Premise 1: $R \subseteq X \text{ x } Y$
Premise 2: $A^* \subseteq X$

Then:

$$\mu_{B^*=A^*\circ R}(y) = \sup_{x\in X} \min\left(\mu_{A^*}(x), \mu_R(x,y)\right) \tag{6.55}$$

2. The extended *Modus Tollens*:

Premise 1: $R \subseteq X \text{ x } Y$
Premise 2: $B^* \subseteq Y$

Then:

$$\mu_{A^*=R\circ B^*}(x) = \sup_{y\in Y} \min\left(\mu_{B^*}(y), \mu_R(x,y)\right) \tag{6.56}$$

3. The extended *Hypothetical syllogism*:

Premise 1: $R_1 \subseteq X \text{ x } Y$
Premise 2: $R_2 \subseteq Y \text{ x } Z$

Then:

$$\mu_{R=R_1\circ R_2}(x,z) = \sup_{y\in Y} \min\left(\mu_{R_1}(x,y), \mu_{R_2}(y,z)\right) \tag{6.57}$$

**Example 6.17.** Let $X = Y = \{1, 2, 3, 4\}$, and assume that on $X$ and $Y$ the linguistic judgments SMALL and LARGE correspond to the fuzzy subsets (μ / x means that μ is membership of x to a set):

SMALL $= S =$ $\{1/1,\ 0.9/2,\ 0.1/3,\ 0/4\}$
LARGE $= L = S^C = \{0/1,\ 0.1/2,\ 0.9/3,\ 1/4\}$

In X x Y a relation is defined as follows:

$R_1 \cup R_2 =$ (IF $x$ IS $S$ THEN $y$ is $L$) $\cup$ (IF $x$ IS $L$ THEN $y$ is $S$)

Numerical results for $R_1$ and $R_2$ are displayed in Table 6.12 a) and b) respectively, while Table 6.13 displays the relation $R = R_1 \cup R_2$.

**Table 6.12 a)** Relation $R_1$ ;

| | | y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| | | $\mu_{B1}$ | 0 | 0.1 | 0.9 | 1 |
| x | $\mu_{A1}$ | $\mu_{R1}$ | | | | |
| 1 | 1 | | 0 | 0.1 | 0.9 | 1 |
| 2 | 0.9 | | 0 | 0.1 | 0.9 | 0.9 |
| 3 | 0.1 | | 0 | 0.1 | 0.1 | 0.1 |
| 4 | 0 | | 0 | 0 | 0 | 0 |

**b)** Relation $R_2$

| | | y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| | | $\mu_{B2}$ | 1 | 0.9 | 0.1 | 0 |
| x | $\mu_{A2}$ | $\mu_{R2}$ | | | | |
| 1 | 0 | | 0 | 0 | 0 | 0 |
| 2 | 0.1 | | 0.1 | 0.1 | 0.1 | 0 |
| 3 | 0.9 | | 0.9 | 0.9 | 0.1 | 0 |
| 4 | 1 | | 1 | 0.9 | 0.1 | 0 |

**Table 6.13** Relation $R = R_1 \cup R_2$

| | y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| x | $\mu_R$ | | | | |
| 1 | | 0 | 0.1 | 0.9 | 1 |
| 2 | | 0.1 | 0.1 | 0.9 | 0.9 |
| 3 | | 0.9 | 0.9 | 0.1 | 0.1 |
| 4 | | 1 | 0.9 | 0.1 | 0 |

**Table 6.14** Modus Ponens

| | | y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| x | $\mu_{A*}$ | $\min(\mu_R, \mu_{A*})$ | | | | |
| 1 | 0 | | 0 | 0 | 0 | 0 |
| 2 | 0.1 | | 0.1 | 0.1 | 0.9 | 0.9 |
| 3 | 0.9 | | 0.5 | 0.5 | 0.1 | 0.1 |
| 4 | 1 | | 0 | 0 | 0 | 0 |
| | | $\mu_{B*}$= sup | **0.5** | **0.5** | **0.9** | **0.9** |

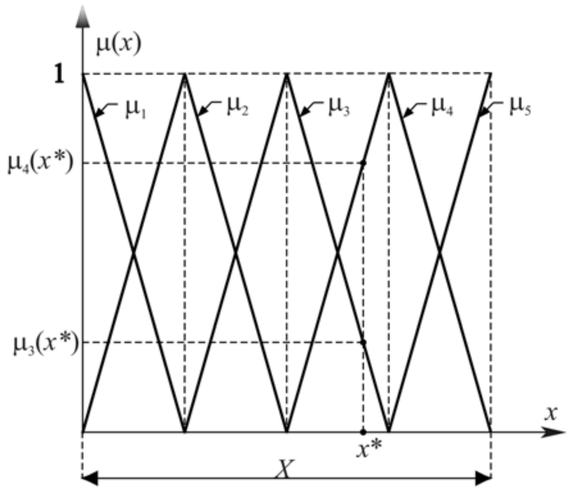
Let's suppose to obtain the fuzzy measure on $X$: $A^* = \{0 / 1,\ 1 / 2,\ 0.5 / 3,\ 0 /4 \}$. Then the rule of the Modus Ponens works as shown in Table 6.14. The result is the sub-normal fuzzy subset of $Y$ : $B^* = \{0.5 / 1,\ 0.5 / 2,\ 0.9 / 3,\ 0.9 /4 \}$.

### *6.5.3 Pattern Recognition and Clustering*

Classical pattern recognition is generally based on the subdivision of a space $X$, where some variables $s$ assume values, in a standard partition: i.e. a group of disjoint subsets $B^j$ ( $j$=1,2..$c$) whose union is $X$; if the actual pattern $x^*$ is observed, it can be classified according to the standard subsets or patterns. Formally its membership to the standard patterns is given by

$$S(x^*, B^j) = \chi_j\left(x^*\right) \qquad (6.58)$$

indicating with $\chi_j(x)$ the characteristic function on $X$ of the subset $B^j$ (standard pattern $j$).



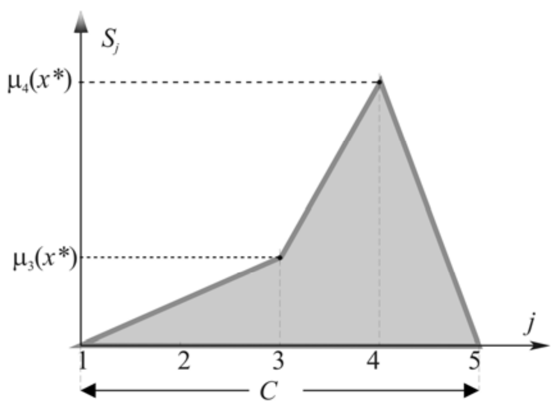




**Fig. 6.12** Pattern recognition of a crisp measure $x^*$: **a)** fuzzy pseudo-partition; **b)** fuzzy classifier

A group of $c$ fuzzy relations $R_j$ ($j$ = 1,2..$c$) on the same space $X$ can be assumed to define a list of $c$ standard patterns; the patterns are frequently given as a fuzzy pseudo-partition of $X$ (Klir and Yuan 1995, p. 359), if the following condition is respected:

$$\forall x \in X: \quad \sum_j \mu_j(X) = 1 \tag{6.59}$$

If the actual pattern $x^*$ is crisp, i.e. observed without uncertainty, it can be classified according to the standards by means of a *fuzzy* classifier, a fuzzy subset of the set $C = \{1, 2, ...c\}$ whose membership function is given by the extension of Eq. (6.58):

$$S(x^*, R_j) = \mu_j(x^*) \tag{6.60}$$

If a crisp classification is required, the pattern (or patterns) $j^*$ with maximum of S can be selected by the decision-maker (Klir and Yuan 1995, pp. 367-369).

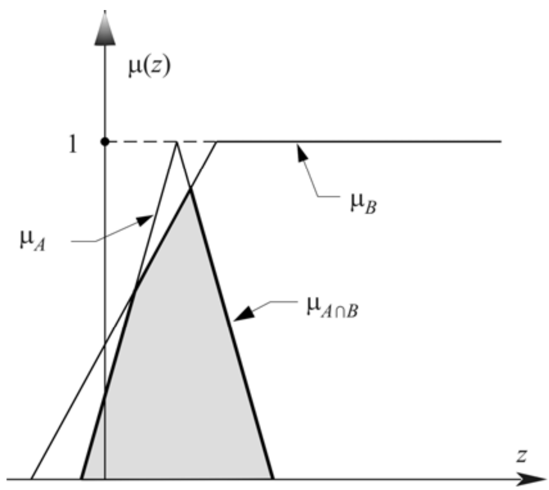
On the contrary, if the actual observed pattern is uncertain, i.e. measured by a random or fuzzy relation $\Re$ on $X$, a criterion of comparison of actual and standard pattern is required. The comparison of fuzzy set can be based on the idea of *degree of sub-sethood* (DoS) of fuzzy subsets of the space $X$:

$$\text{DoS}(A, B) = |A \cap B| / |A| \tag{6.61}$$

where the *cardinality* $|A|$ of a fuzzy subset $A$ on a finite space $X$ is the extension of the cardinality of a crisp set, measured by the sum of the membership values over $X$.

In another context the operator for the comparison can be considered a *filtering* of $A$ through the *filter* $B$ (Bignoli 1991). As an example in structural engineering, the fuzzy set $B$ on the space of a limit state function $z$ could define the class of safe structures, extending the deterministic definition generally assumed in Codes ( $z > 0$ ) to a more reasonable transition from safety to unsafety, as shown in Figure 6.13.

**Fig. 6.13** Filtering the fuzzy measure $A$ of the safety margin $z$ through the fuzzy filter $B$ (defining safety)




It is possible to observe that, assuming the standard max and min operators (dual with respect to the complementation rule given by (6.45) for union and intersection, the Law of the excluded middle is not respected: if $A$ is a not trivial (i.e. not classic) fuzzy subset of the universal set $X$ (Figure 6.14)

$$A \cup A^{C} \subset X \tag{6.62}$$

and therefore, for any pair of not trivial fuzzy sets:

$$\mathrm{DoS}\ (A, B) + \mathrm{DoS}\ (A^{C}, B) < 1 \tag{6.63}$$

That is: the judgment on safety and unsafety cannot be derived from one another, as in probability theory. In some applications this asymmetry is not reasonable or is computationally heavy; so it can be eliminated assuming the following symmetric operator (normalized from 0 to 1; (Bignoli 1991)) in the comparison of fuzzy sets:

$$\mathrm{K}(A, B) = (1 + \mathrm{DoS}\ (A, B) - \mathrm{DoS}\ (A^{c}, B)\ )\ /\ 2 \tag{6.64}$$

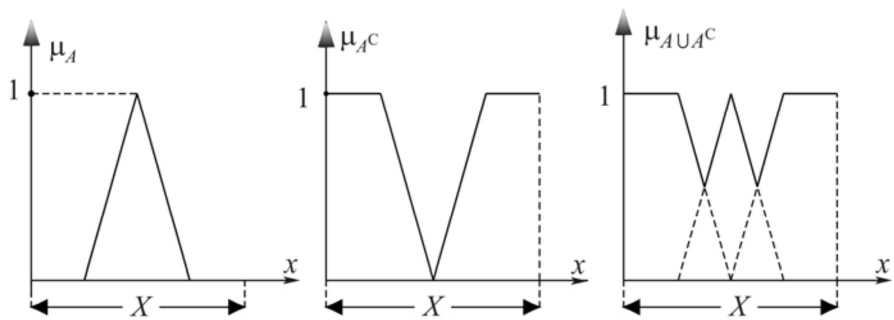



**Fig. 6.14** Standard complementation and union violate the Law of the excluded middle

The same operators (6.61) or (6.64) can be used to compare a fuzzy set or relation A (the actual observed pattern) with a list of fuzzy sets or relations (the standard patterns $B^1$, $B^2$, ... $B^c$) of the universal set $S$, by assuming the values of DoS or K as membership of the fuzzy classifier.

**Example 6.18.** The fuzzy sets measuring SMALL and LARGE in Example 6.17 give a fuzzy partition of the space $X = \{1, 2, 3, 4\}$. Many applications of fuzzy pattern recognition are based on a fuzzy partition, for example of the space $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ by three patterns corresponding to the linguistic judgment SMALL , MEDIUM and LARGE:

$$\begin{aligned}
\text{SMALL} &= S = \{1/0,\ 0.8/1,\ 0.6/2,\ 0.4/3, 0.2/4, 0/5\} \\
\text{MEDIUM} &= M = \{1/0,\ 0.2/1,\ 0.4/2,\ 0.6/3, 0.8/4, 1/5, \\
&\qquad 0.8/6,\ 0.6/7,\ 0.4/8,\ 0.2/9,\ 0/10\} \\
\text{LARGE} &= L = \{0/5,\ 0.2/6,\ 0.4/7,\ 0.6/8, 0.8/9, 1/10\}
\end{aligned}$$

Having observed the fuzzy pattern $A^* = \{0.6/5,\ 1/6,\ 0.4/7,\ 0.4/3, 0.2/4\}$, with cardinality $|A^*| = 2$, and using DoS operator, the fuzzy classifier is the fuzzy set:

$C_{DoS}(A^*) = \{(0/2)$ / SMALL, $((0.6+0.8+0.4)/2 = 0.9)$ / MEDIUM, $((0.2+0.4)/2 = 0.3)$ / LARGE$\}$

Usink K operator SoD should also be calculated for the complementary fuzzy set $A^{*C} = \{1/0,\ 1/1,\ 1/2,\ 1/3, 1/4, 0.4/5,\ 0/6,\ 0.6/7,\ 1/8,\ 1/9,\ 1/10\}$, with $|A^{*C}| = 9$.

The fuzzy classifier is the fuzzy set:

$C_K(A^*) = \{((1+0-0.333)/2 = 0.333)$ / SMALL, $((1+0.9-0.4)/2 = 0.75)$ / MEDIUM, $((1+0.3-0.311)/2 = 0.494)$ / LARGE$\}$

With both operators the preferred deterministic classification is MEDIUM.

Other applications refine the classification through 5 patterns, assuming:

$$\begin{aligned}
\text{VERY SMALL} &= VS = \text{SMALL}^2 = \\
&= \{1/0,\ 0.64/1,\ 0.36/2, 0.16/3, 0.04/4, 0/5\} \\
\text{VERY LARGE} &= VL = \text{LARGE}^2 = \\
&= \{0/5,\ 0.04/6,\ 0.16/7,\ 0.36/8, 0.64/9, 1/10\}
\end{aligned}$$

But in this case the patterns are not a fuzzy partition of $X$.

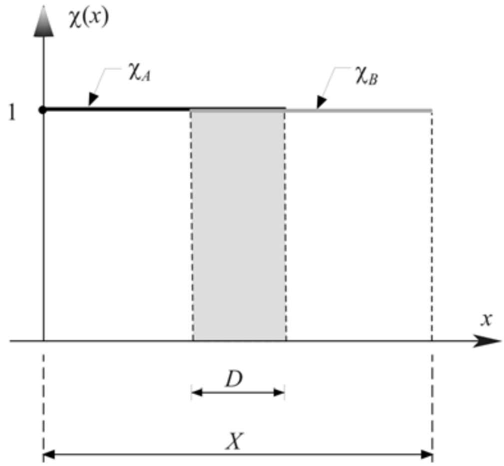
*Fuzzy clustering* can be considered the dual problem to the pattern recognition. Here a list of crisp or fuzzy data $(x_1, x_2, \ldots x_q)$ is given and a fuzzy partition is required, according to a criterion of optimality of the number $c$ of clusters, or, if $c$ is selected, of the membership functions of the clusters (Klir and Yuan 1995; pp. 358-365).

### 6.5.4 Fuzzy Model of Multi-objective Decision Making

In the space of the variables $X$ any objective or any constraint gives a restriction to the values to be chosen by the decision-maker. If the restriction is given by two classical subsets $A$ and $B$ (Figure 6.15), their intersection gives the range of the values **s** satisfying both the objectives and constraints, i.e. it gives the subset $D$ of the decisions through the characteristic function:

$$\chi_D(x) = \chi_{A\cap B}(x) = \min\left(\chi_A(x), \chi_B(x)\right) \tag{6.65}$$

**Fig. 6.15** Classical Decision set




All the decisions in $D$ are completely equivalent for the decision-maker, so he or she could be like the donkey described by Giovanni Buridano, which died of hunger having two lots of hay at exactly the same distance to the left and to the right.

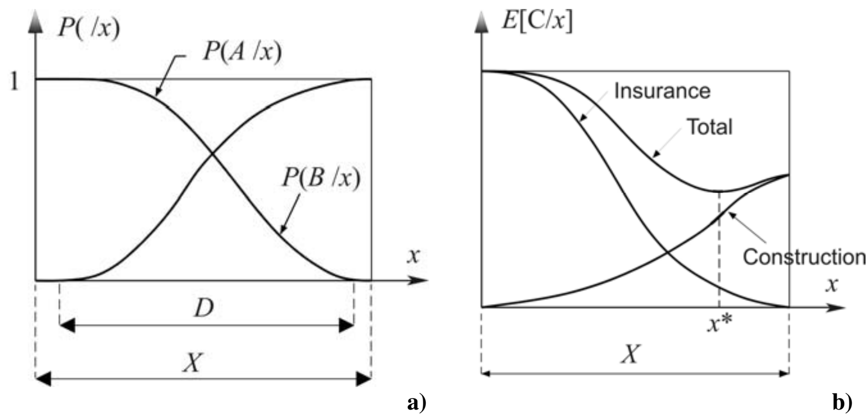



**Fig. 6.16** a) Probabilistic Decision Set ; b) crisp decision combining the objectives

Alternatively the restriction could be according to a probabilistic model, i.e. through the conditional probabilities to satisfy the objectives (or constraint) A and B with respect to any choice $x \in X$.

Even in this case we can derive a range $D$ of the choices with positive probability of satisfying both $A$ and $B$ (Figure 6.16a), but not a univocal criterion for a crisp choice; for this purpose the multi-objective problem should be reduced to a mono-objective problem, selecting a priority or combining the objectives.

For example let $A$ be the cost of construction of a structure to be minimized and $B$ the safety with respect to a limit-state to be maximized; we could make the choice to minimize the expectation of the total cost in the life of the structure: cost of construction plus cost of the insurance of the expected damages due to exceeding the limit-state (Figure 6.16b).

If the restrictions are given by two fuzzy sets or relations, the decision subset is the fuzzy set or relation $D$, with membership function given by the extension of Eq. (6.65):

$$\mu(x) = \min(\mu_A(x), \mu_B(x)) \tag{6.66}$$

i.e. $D$ is the intersection of the fuzzy sets $A$ and $B$, when the standard operator min is selected for intersection. Moreover it is quite reasonable for the decision-maker to select the crisp decision $x^*$ such that (Figure 6.17a):

$$\mu(x^*) = \max_{x \in X} \min(\mu_A(x), \mu_B(x)) \tag{6.67}$$

According to De Morgan's Law, with the standard definition of union and complementation, such a decision also minimizes the membership of the union of the complementary fuzzy sets (Figure 6.17b):

$$\mu(x^*) = \min_{x \in X} \max(1 - \mu_A(x), 1 - \mu_B(x)) \tag{6.68}$$

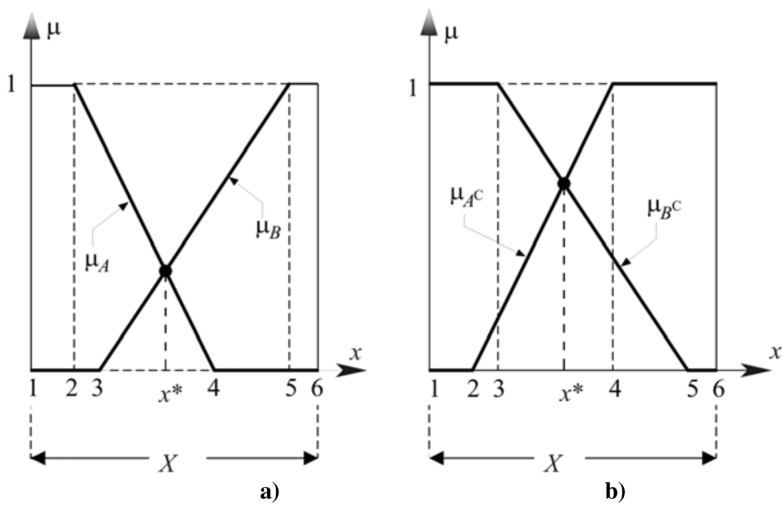



**Fig. 6.17** Optimal decision in the decision space: **a)** maximizing the advantage ; **b)** minimizing the disadvantage

A geometric interpretation of the solution can be given in the *space of the membership functions* (Figure 6.18a) or of their complements to 1

(Figure 6.18b), i.e. the multi-dimensional Cartesian product of their intervals of variation [0, 1] (the unit square with two objectives or constraints as in Figure 6.18). In these spaces the point (1, 1,..1) and point (0, 0, ..0) are the *ideal points* respectively, and the point corresponding to $x^*$ should be the point closest to the ideal points.

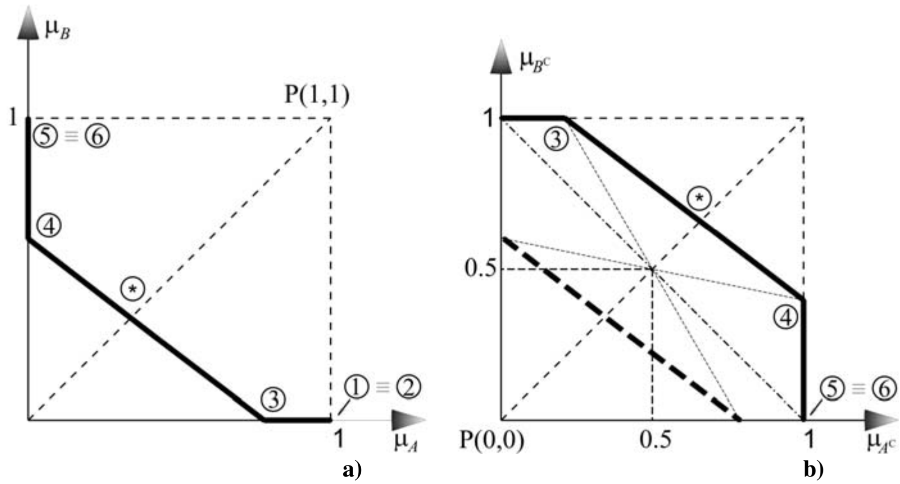



**Fig. 6.18** Optimal decision in the membership space: **a)** maximizing the advantage; **b)** minimizing the disadvantage

Of course a metric should be chosen in these spaces to measure the distance, and there is no reason for preferring a Euclidean metric to any other. A large class of possible metrics is the $l_p$ class, defined by the formula:

$$\mathrm{d}\left(\mathrm{P}_1,\mathrm{P}_2\right)=\left(\sum_j |\,\mu_j\left(\mathrm{P}_1\right)-\mu_j\left(\mathrm{P}_2\right)|^p\right)^{\frac{1}{p}} \tag{6.69}$$

The Euclidean metric belongs to this class with $p = 2$. Equations (6.67) and (6.68) can be derived by measuring the distances in the membership spaces through a metric $l_\infty$, i.e. if the distance between points $P_1$ ($\mu_{11}$, $\mu_{21}$,.. $\mu_{j1}$...) and $P_2$ ($\mu_{12}$, $\mu_{22}$,.. $\mu_{j2}$...) is defined by:

$$\mathrm{d}\left(P_1,P_2\right)=\max_j |\,\mu_j\left(P_1\right)-\mu_j\left(P_2\right)| \tag{6.70}$$

Using this metric, the points at equal distance from the ideal point are on the perimeter of a square (with two objectives; a multidimensional square with more objectives) centered on the ideal point, with each side orthogonal to one axis (Figure 6.19). As a comparison, in the same Figure the points at equal distance from the ideal point are displayed for $p = 1$ and $p = 2$ (Euclidean metric).

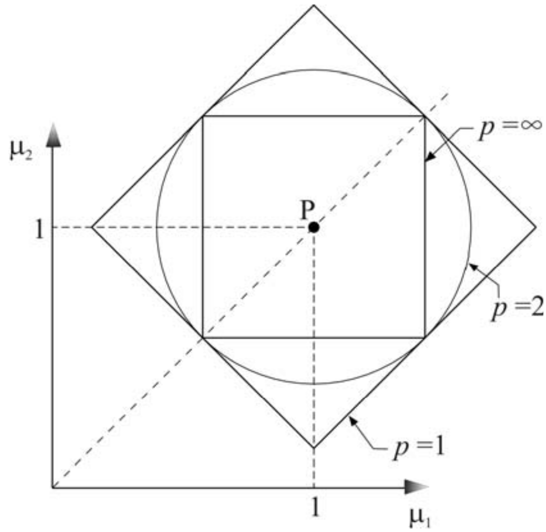



**Fig. 6.19** Points at equal distance from the ideal point $P(1, 1)$ according to different $l_p$ metrics

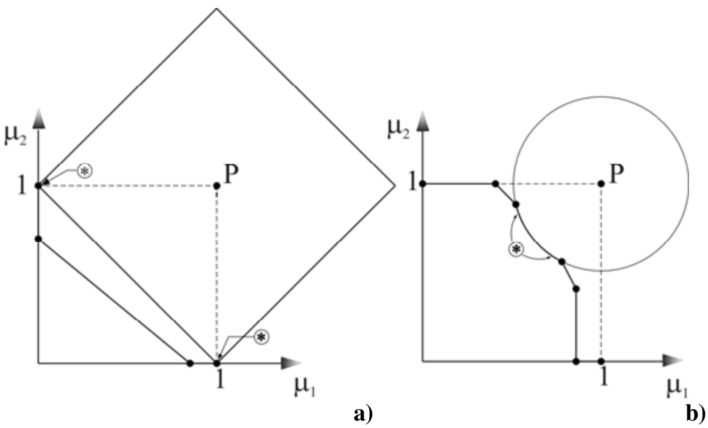


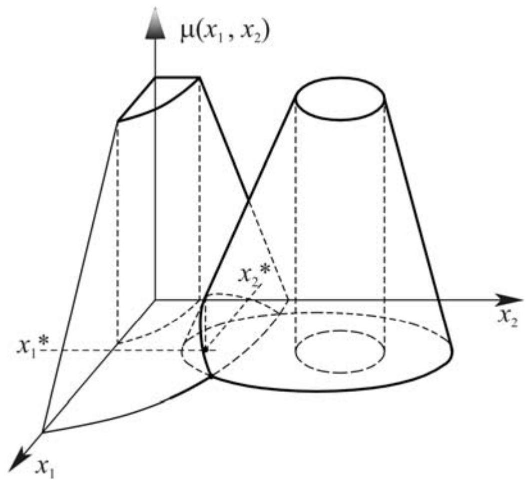

**Fig. 6.20** Not unique decision with: **a)** $p = 1$ ; **b)** $p = 2$

It is possible to demonstrate (Tonon and Bernardini 1999) that with the choice $p = \infty$, if the functions $\mu_j$ are strictly monotone continuous functions (increasing or decreasing) on $X$, the optimal decision is unique. If we use different metrics the optimal decision could be not unique (Figure 6.20), even in the case of strictly monotone membership functions.

The above solution can be extended to the case of many variables $x = (x_1, x_2, \ldots x_n)$ and more than two objectives or constraints, when each objective or constraint is measured by a fuzzy relation $R_j$ with joint membership function

$\mu_j(x)$, $j = 1, 2, \ldots m$ (see Figure 6.21 for $n = 2$ and $m = 2$, and Figure 6.22 for $n = 1$, $m = 3$).

**Fig. 6.21** Optimal decision on the space $X_1$ x $X_2$ as a compromise between one objective and one constraint




When $m - n > 1$ the optimal solution is generally determined by some *active* objectives or constraints, while the others are satisfied with a membership greater than the membership obtained for the active ones (Figure 6.22). As a general procedure, we must search for the maximum value of membership, α, such that all the objectives or constraints are satisfied at least with membership α . That is:

$$\begin{aligned}&\textit{Maximize}\ \alpha \\ &\textit{Subject to}: \quad \mu_j(x) - \alpha \geq 0\ ;\ j = 1, 2, \ldots m \\ &\qquad\qquad\quad \alpha \in (0, 1] \quad ; \quad x \in X\end{aligned} \tag{6.71}$$

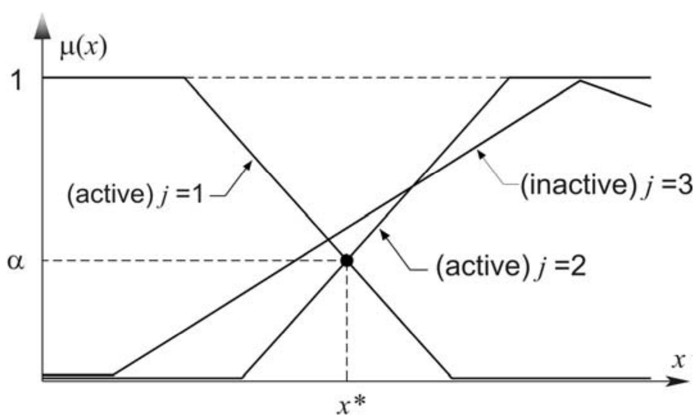



**Fig. 6.22** Optimal decision on $X_1$ with $m = 3$

When the membership function $\mu_j(x)$, $j = 1, 2, \ldots m$ are linear functions of the variables $(x_1, x_2, \ldots x_n)$ for $0 < \alpha < 1$ (as shown in Fig. 6.22), formulae (6.71) define a standard linear programming problem with respect to the variables $x_1, x_2, \ldots x_n, \alpha$.

## 6.6 Conclusions

When information is affected by both randomness and imprecision, random sets allow the whole spectrum of uncertainty experienced in data collection to be taken into account. In this case, imprecision leads to upper and lower bounds on the probability of an event of interest. This result is particularly useful in the reliability evaluation of engineering systems. Indeed, imprecision on input variables has strong repercussions on the prediction of a system's behavior, so that probabilistic analyses that ignore imprecision are meaningless, especially when very low probabilities of failure are calculated or required.

Three alternative basic rules have been identified for the combination of imprecise data. The subjective choice of the decision-maker must depend on the reliability of the available information and the aim of the analysis.

In the application of the “Intersection” rules, attention should be given to the normalization of the obtained probabilistic assignment, especially when strongly conflicting sources of information should be combined. Yager's or Dubois and Prade's modifications of Dempster's rule appear to be reasonable depending on the reliability of the sources.

When imprecision affects the available information, a clear distinction must be made between combining two pieces of information and updating one piece of information with another, because the rules for combining information are different from the rules for updating information. Accordingly, it is necessary to distinguish between belief functions as generalized probabilities and belief functions as representations of evidence.

## Notes

**N 6-1.** The choice of operators to combine characteristic functions of sets to obtain the characteristic function of their intersection or union is not unique. For example, for intersection the equivalent product rule $\chi_C(s) = \chi_A(s) \cdot \chi_B(s)$ could be used. The class of equivalent operators to be used to obtain the characteristic function of the intersection from the characteristic function of the combined sets is known as *triangular norms* (*or t-norms*). See for example (Klir and Yuan 1995).

This question is discussed in Section 6.4 with reference to the extended operations in fuzzy logic. The choice of the min for intersection (and max for union) is justified in this discussion, mainly.

**N 6-2.** The choice of operator giving the characteristic function of the complementary set from the characteristic function of a set is not unique. Here the rule $\chi_{A^c}(s) = 1 - \chi_A(s)$ is assumed. De Morgan's Law and a rule for complementation define a class of dual operators for union, in the class known as *triangular conorms* (*t-conorms*). For example assuming the product rule suggested in N 6-1 as the operator of intersection one obtains:

$$\chi_C(s) = 1 - (1 - \chi_A(s)) \cdot (1 - \chi_B(s)) = \chi_A + \chi_B(s) - \chi_A(s) \cdot \chi_B(s)$$

**N 6-3.** These alternative models of uncertainty have demonstrated their usefulness especially in the field of control device design and the formalization of approximate reasoning for the development of expert systems and decision-making procedures (performed either by computer-aided human brains or totally autonomous machines (Artificial Intelligence)).

However, in recent years, even the most inveterate supporters of these alternative models have expressed a willingness to compare their results with the relevant conclusions obtained using probabilistic methods, particularly when their conclusions and recommendations to a decision-maker need to be justified in terms of reliability, robustness and cost/benefit comparisons. This comparison seems to be imperative in the field of Structural Engineering (structural safety, seismic vulnerability and risk assessment of buildings and infrastructures).

An interesting example of this wider cultural climate is the recent identification of theoretical problems (proposed as Challenge Problems to the international scientific community (Oberkamp, Helton et al. 2001), which can only be solved with great difficulty using classical probabilistic methods. These problems involve uncertainty propagation through mathematical models in a decision-support context, when the basic variables are differently measured by intervals, multiple intervals, precise probability distributions or a probability distribution with imprecise parameters. Moreover in some cases some parameters are given by $n$ independent, equally credible, sources of information, highlighting an additional topic: how can one combine different, perhaps independent, sources of uncertain information in the same variable? The same problem is at the core of fuzzy set theory (Fagin 1999; Fagin 2002).

Interesting reviews of the state-of-the-art on the subject have been presented by Genest and Zidek (Genest and Zidek 1986), Dubois and Prade (Dubois and Prade 1988), Levefre et al. (Lefevre, Colot et al. 2002), and Sentz and Ferson (Sentz and Ferson 2002).

**N 6-4.** In a very interesting paper, Yager (1991) attempts to overcome the same problem: he starts by observing that the application of Dempster's rule of two consonant random sets (or belief structures) gives a generally non consonant random set; i. e. this operation is not closed on the set of consonant belief structures.

In order to obtain a closed operation, Yager proposed again the fuzzy composition, i.e. the fuzzy intersection by means of the min operator applied to the membership functions of the two associated fuzzy sets (or, alternatively, their union by the max operator). His justification for this operator as an alternative to Dempster's rule is very interesting because it provides a deep insight into the relationship between the information described by probabilitiy and fuzziness, respectively.

Firstly, he notices that two fuzzy sets (hence two consonant random sets) can assume a *commensurate representation*: i.e they can be described by two sets of nested α-cuts (focal elements of the equivalent random sets) taken at the same α-levels. Since the α-levels are the same for the two fuzzy sets, the probabilistic assignements for the two associated consonant random sets will be identical (recall that probabilistic assignements are obtained as differences between consecutive α-values; see Eq. (6.52) and (6.53) in this book).

Demspter's rule of combination applied to these commensurate random sets assumes stochastic independence for the joint probabilistic assignements of intersections of any pair of α-cuts, whose marginals are the two identical probabilistic assignements indicated above.

On the other hand, Yager's rule of fuzzy composition (see Eq. (6.50) is justified by assuming perfect correlation: in the matrix of the joint distribution, the main diagonal contains the marginal values, and all off-diagonal terms are equal to zero (*synonyminity* property according toYager).

This property suggests that the information conveyed by fuzzy sets is completely described by the nested ordered family of their α-cuts, while their probabilistic assignement is invariant (and hence completely correlated when distinct fuzzy sets are considered).

# References


Alefeld, G., Herzberger, J.: Introduction to Interval Computations. Academic Press, New York (1983)

Algorithmic Solutions. LEDA (2007), `http://www.algorithmic-solutions.com/enleda.htm`

Alvarez, D.A.: On the calculation of the bounds of probability of events using infinite random sets. International Journal of Approximate Reasoning 43(3), 241–267 (2006)

Ang, A.H.-S., Tang, W.H.: Probability Concepts in Engineering Planning and Design, vol. 1. John Wiley & Sons, New York (1975)

Babuska, I., Strouboulis, T.: The Finite Element Method and its Reliability. Oxford Press, Oxford (2001)

Barton, N.: The shear strength of rock and rock joints. Int. J. Rock Mech. Min. Sci. & Geomech. Abstr. 13, 255–279 (1976)

Ben-Haim, Y.: Uncertainty, probability and information-gaps. Reliability Engineering & System Safety 85, 249–266 (2004)

Ben-Haim, Y., Elishakoff, I.: Convex Models of Uncertainty in Applied Mechanics. Elsevier Science Publishers, Amsterdam (1990)

Bernardini, A.: Valutazioni di frequenze attese di eventi da misure sfuocate di vulnerabilità sismica. 7° Convegno ANIdIS L'ingegneria sismica in Italia, Siena, Collegio degli Ingegneri della Toscana (1995) (in Italian)

Bernardini, A.: What are the Random and Fuzzy Sets and How to Use them for Uncertainty Modelling in Engineering Systems. In: Elishakoff, I. (ed.) Whys and Hows in Uncertainty Modelling. CISM Courses and Lectures, vol. 338, pp. 63–125. Springer, Wien (1999)

Bernardini, A.: Upper and lower probabilities of events from Random sets and Fuzzy sets. In: Bignoli, A.J. (ed.) Seguridad en Ingenieria. Academia Nacional de Ingenieria, Buenos Aires (2000)

Bernardini, A.: Random sets of the expected seismic damage to building types. In: 9th ICOSSAR, Rome (2005)

Berz, M.: Taylor Models (2007), `http://www.bt.pa.msu.edu/pub/`

Bignoli, A.J.: Teoria Elemental de los Conjuntos Borrosos. Academia Nacional de Ingenieria, Buenos Aires (1991) (in Spanish)

Bowles, J.E.: Foundation Analysis and Design, 5th edn. McGraw Hill, New York (1996)

Burrill, C.W.: Measure, Integration, and Probability. McGraw-Hill, New York (1972)

Campos, L.M.D., Huete, J.F., Moral, S.: Probability Intervals: a tool for uncertainty reasoning. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2(2), 167–196 (1994)

Cariolaro, G., Pierobon, G.: Teoria della Probabilità e dei Processi Aleatori, vol. 1. Pàtron Editore, Bologna (1992)

CEN, Eurocode 6- Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures. European Committe for Standardization. EN 1996-1-1 (2005)

Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. Mathematical Social Sciences 17, 263–283 (1989)

Chatillon, G.: The balloon rules for a rough estimate of the correlation coefficient. The American Statistician 38(1), 58–60 (1984)

Chaudhuri, B.B., Bhattacharya, A.: On correlation between two fuzzy sets. Fuzzy Sets and Systems 118, 447–456 (2001)

Choquet, G.: Theory of capacities. Annales de L'Institut Fourier 5, 131–295 (1954)

Clarkson, K.L.: Safe and effective determinant evaluation. In: Proc. 33rd Annual IEEE Symposium on Foundations of Computer Science (1992)

Computer Methods in Applied Mechanics and Engineering. Special issue on Adaptive Modeling and Simulation 195(4-6), 205–480 (2006)

Corliss, G.F.: SCAN-98 collected bibliography. In: Csendes, T. (ed.) Developments in Reliable Computing. Kluwer, Dordrecht (1999)

Couso, I., Moral, S., Walley, P.: Examples of independence for imprecise probabilities. In: ISIPTA 1999 - Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications, Zwijnaarde, Belgium (1999)

Cui, W., Blockley, D.I.: Interval probability theory for evidential support. Int. J. Intelligent Systems 5, 183–192 (1990)

Dantsin, E., Kreinovich, V., Wolpert, A., Xiang, G.: Population variance under interval uncertainty: a new algorithm. Reliable Computing 12, 273–280 (2006)

de Berg, M., Cheong, O., van Kreveld, M., Overmars (eds.): Computational Geometry: Algorithms and Applications. Springer, Berlin (2000)

De Luca, A., Termini, S.: A definition of a non-probabilistic entropy in the setting of fuzzy set theory. Information and Control 20, 301–312 (1972)

Delgado, M., Moral, S.: A definition of inclusion for evidences. Fuzzy Mathematics (Huazhong, China) 6(4), 81–87 (1987)

Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38(2), 325–339 (1967)

Diamond, P.: Least squares fitting of compact set-valued data. J. Math. Anal. Appl. 147, 531–544 (1990)

Ditlevsen, O.: Narrow reliability bounds for structural systems. J. Structural Mechanics, ASCE 7, 453–472 (1979)

Dong, W.M., Wong, F.S.: Fuzzy weighted averages and implementation of the extension principle. Fuzzy Sets and Systems 21, 183–199 (1987)
Dong, W.M., Shah, H.C.: Vertex method for computing functions of fuzzy variables. Fuzzy Sets and Systems 24, 65–78 (1987)
Dong, W.M., Chiang, W.L., Sha, H.C.: Fuzzy information processing in seismic hazard analysis and decision making. Soil Dynamics and Earthquake Engineering 6(4), 220–226 (1987)
Dow, J.O.: A Unified Approach to the Finite Element Method and Error Analysis Procedures. Academic Press, New York (1998)
Dubois, D., Prade, H.: On several representations of an uncertain body of evidence. In: Gupta, M.M., Sanchez, E. (eds.) Fuzzy Information and Decision Processes, pp. 167–181. North-Holland, Amsterdam (1982)
Dubois, D., Prade, H.: A set-theoretic view of belief functions. Logical operations and approximations by fuzzy sets. Int. J. of General Systems 12, 193–226 (1986)
Dubois, D., Prade, H.: Representation and combination of uncertainty with belief functions and possibility measures. Comput. Intell. 4, 244–264 (1988)
Dubois, D., Prade, H.: Possibility Theory: an Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
Dubois, D., Prade, H.: Fuzzy sets, probability and measurement. European Journal of Operational Research 40, 135–154 (1989)
Dubois, D., Prade, H.: Consonant approximations of Belief Functions. Int. J. Approximate Reasoning 4, 491–499 (1990)
Dubois, D., Prade, H.: Random sets and fuzzy interval analysis. Fuzzy Sets and Systems 42, 87–101 (1991)
Eskesen, S.D., Tengborg, P., Kampmann, J., Veicherts, T.H.: Guidelines for tunnelling risk management. Tunnelling and Underground Space Technology 19, 217–237 (2004)
Fagin, R.: Combining fuzzy information from multiple systems. J. Computer and System Sciences 58, 83–99 (1999)
Fagin, R.: Combining fuzzy information: an overview. SIGMOD Record 31(2), 109–118 (2002)
Fagin, R., Halpern, J.Y.: A new approach to updating beliefs. In: Bonissone, P.P., Herion, M., Kanal, L.N., Lemmer, J. (eds.) Uncertainty in Artificial Intelligence, vol. VI, pp. 347–374. Elsevier, Amsterdam (1991)
Feng, Y., Hu, L., Shu, H.: The variance and covariance of fuzzy random variables and their applications. Fuzzy Sets and Systems 120, 487–497 (2001)
Ferson, S., Kreinovich, V.: Modeling correlation and dependence among interval. In: NSF Workshop on Reliable Engineering Computing, Savannah, GA (2006)
Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing Probability Boxes and Dempster-Shafer Structures. SAND REPORT 2002-4015. Albuquerque and Livermore, Sandia National Laboratories: 1-143 (2003)
Ferson, S., Ginzburg, L., Wu, B.: Computing variance for interval data is NP-hard. ACM SIGACT News 33(2), 108–118 (2002)

Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., Aviles, M.: Exact bounds on finite populations of interval data. Reliable Computing 11(3), 207–233 (2005)

Ferson, S., Ginzburg, L., Kreinovich, V., Aviles, M.: Exact bounds on sample variance of interval data. In: 2002 SIAM Workshop on Validated Computing. Toronto, Canada (2002)

Ferson, S., Nelsen, R.B., Hajagos, J., Berleant, D.J., Zhang, J., Tucker, W.T., Ginzburg, L.V., Oberkampf, W.L.: Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis. SAND REPORT 2004-3072, Albuquerque and Livermore, Sandia National Laboratories (2004)

Fetz, T.: Sets of joint probability measures generated by weightedmarginal focal sets. In: de Cooman, G., Fine, T., Seidenfeld, T. (eds.) ISIPTA 2001, Proceedings of the Second Symposium on Imprecise Probabilities and Their Applications. Shaker Publ. BV, Maastricht (2001)

Fetz, T.: Mengen von gemeinsamen Wahrscheinlichkeitsmaβen erzeugt von zufalligen Mengen, Innsbruck. University of Innsbruck, Ph.D. Dissertation (2003)

Fetz, T.: Multi-parameter models: rules and computational methods for combining uncertainties. In: Fellin, W., Lessmann, H., Vieider, R., Oberguggenberger, M. (eds.) Analyzing Uncertainty in Civil Engineering, pp. 73–100. Springer, Berlin (2004)

Fetz, T., Oberguggenberger, M.: Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. Reliability Engineering & System Safety 85(1-3), 73–87 (2004)

Genest, C., Zidek, J.V.: Combining probability distributions: a critique and an annotated bibliography. Statistical Science 1(1), 114–148 (1986)

Gil, M.A., Lubiano, M.A., Montenegro, M., Lopez, M.T.: Least squares fitting of an affine function and strength of association for interval-valued data. Metrika 56, 97–111 (2002)

Goodman, J.E., O'Rourke, J. (eds.): Handbook of Discrete and Computational Geometry. Chapman & Hall, Boca Raton (2004)

Grunthal, G.: European Macroseismic Scale 1998. Cahiers du Centre Européen de Géodynamique et de Séismologie 15, 1–97 (1998)

Grötschel, M., Lovász, L., Schrijver, A.: Geometric Algorithms and Combinatorial Optimization. Springer, Berlin (1988)

Hall, J.W., Blockley, D.I., Davis, J.P.: Uncertain inference using interval probability theory. Int. J. of Approximate Reasoning 19, 247–264 (1998)

Hall, J.W., Lawry, J.: Generation, combination and extension of random set approximations to coherent lower and upper probabilities. Reliability Engineering & System Safety 85(1-3), 89–101 (2004)

Halmos, P.R.: Measure Theory. Van Nostran, New York (1950)

Hammer, R., Hocks, M., Kulisch, U., Ratz, D.: C++ Toolbox for Verified Computing. Springer, Berlin (1995)

Hansen, E., Walster, G.W.: Global Optimization Using Interval Analysis. Marcel Dekker, New York (2003)

Hasofer, A.M., Lind, N.C.: Exact and invariant second-moment code format. J. Engineering Mechanics, ASCE 100, 111–121 (1974)
Horst, R., Pardalos, P.M., Thoai, N.V.: Introduction to Global Optimization, 2nd edn. Kluwer, Dortrecht (2000)
Hudson, J.A., Harrison, J.P.: Engineering Rock Mechanics - An Introduction to the Principles. Pergamon, Oxford (1997)
Hunter, J.K., Nachtergaele, B.: Applied Analysis. World Scientific, Singapore (2001)
ISRM. Suggested methods for the quantitative description of rock masses. Int. J. Rock Mechanics & Mining Sciences 15, 319–368 (1978)
Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied Interval Analysis. Springer, London (2001)
Jeffrey, R.C.: The Logic of decision. University of Chicago Press, Chicago (1983)
Jeleva, M.: Demand for Insurance, Imprecise Probabilities and Ambiguity Aversion. In: ISIPTA 1999 - Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications, Zwijnaarde, Belgium (1999)
Joslyn, C., Klir, G.: Minimal information loss in possibilistic approximations of random sets. In: IEEE Int. Conf. Fuzzy Systems, San Diego, CA (1992)
Karmarkar, N.: A new polynomial-time algorithm for linear programming. Combinatorica 4, 373–395 (1984)
Kearfott, R.B.: Rigorous Global Search: Continuous Problems. Kluwer Academic Publishers, Dordrecht (1996)
Kinkaid, D., Cheney, W.: Numerical Analysis. Thomson Learning, Pacific Grove (2002)
Klir, G.J.: Principles of Uncertainty: What are they? Why do we need them? Fuzzy Sets and Systems 74, 15–31 (1995)
Klir, G.J.: Uncertainty and Information. Foundations of generalized Information Theory. Wiley & Sons, Inc., Hoboken (2005)
Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic -Theory and Applications. Prentice Hall, Upper Saddle River (1995)
Kolmogorov, A.N.: Foundations of the Theory of Probability. Chelsea Pub. Co., New York (1956)
Kolmogorov, A.N., Fomin, S.V.: Introductory Real Analysis. Translated and edited by Richard A. Silverman. Dover Publications, New York (1975)
Kreinovich, V., Berleant, D., Joan-Arinyo, R., Koshelev, M.: Interval Computations (2007), `http://www.cs.utep.edu/interval-comp` (retrieved June 27, 2009)
Kruse, R.: On the variance of random sets. J. Math. Anal. and Applications 122(2), 469–473 (1987)
Lefevre, E., Colot, O., Vannoorenberghe, P.: Belief function combination and conflict management. Information Fusion 3, 149–162 (2002)
LINDO: LINGO User's guide. Chicago, LINDO Systems Inc. (1999)
Liu, S.-T., Kao, C.: Fuzzy measures for correlation coefficient of fuzzy numbers. Fuzzy Sets and Systems 128, 267–275 (2002)

Lipschutz, S.: Schaum's outline of theory and problems of set theory and related topics. McGraw-Hill, New York (1964)

Loève, M.: Probability theory I, 4th edn. Springer, New York (1977)

Loève, M.: Probability theory II, 4th edn. Springer, New York (1994)

Matheron, G.: Random Sets and Integral Geometry. Wiley & Sons, Inc., New York (1975)

Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM Journal on Optimization 2, 575–601 (1992)

Meyer, K.D., Kruse, R.: On calculating the covariance in the presence of vague data. In: Janko, W.J., Roubens, M., Zimmermann, H.-J. (eds.) Progress in Fuzzy Sets and Systems. Kluwer, Dordrecht (1990)

Miranda, E., Couso, I., Gil, P.: Study of the probabilistic information of a random set. In: Third International Symposium on Imprecise Probabilities and Their Applications, Lugano, Switzerland, Carleton Scientific (2003)

Miranda, E., Couso, I., Gil, P.: Upper probabilities and selectors of random sets. In: Grzegorzewski, P., Hryniewicz, O., Gil, M.A. (eds.) Soft Methods in Probability, Statistics and Data Analysis, pp. 126–133. Physica-Verlag, Heidelberg (2002)

Molchanov, I.: Theory of Random Sets. Springer, Dordrecht (2005)

Moore, R.E.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)

Moore, R.E.: Methods and Applications of Interval Analysis. SIAM, Philadelphia (1979)

Mulhanna, R.L., Mullen, R.L.: Reliable Engineering Computing (2007), `http://www.gtsav.gatech.edu/rec/`

Murthy, C.A., Pal, S.K., Dutta Majumder, D.: Correlation between two fuzzy membership functions. Fuzzy Sets and Systems 17, 23–38 (1985)

Nemirovsky, A., Yudin, N.: Interior-Point Polynomial Methods in Convex Programming. SIAM, Philadelphia (1994)

Nesterov, V.M.: Reliable Computing (2007), `http://www.springer.com/math/cse/journal/11155` (retrieved June 27, 2009)

Neumaier, A.: Interval Methods for Systems of Equations. Cambridge University Press, Cambridge (1990)

Neumaier, A.: Global Optimization (2007), `http://www.mat.univie.ac.at/~neum/glopt.html` (retrieved June 27, 2009)

Nguyen, H.T.: A note on the extension principle for fuzzy sets. J. Math. Anal. and Applications 64, 369–380 (1978)

Nguyen, H.T.: Some mathematical tools for linguistic probabilities. Fuzzy Sets and Systems 2, 53–65 (1979)

Nguyen, H.T.: An Introduction to Random Sets. CRC Press, Hoboken (2006)

Oberkamp, W.L., Helton, J.C., Joslyn, C.A., Wojtkiewicz, S.F., Ferson, S.: Challenge Problems: Uncertainty in System Response given uncertain Parameters (draft: November 29, 2001), `http://www.sandia.gov/epistemic/`

Oberkampf, W.L., Trucano, T.G., Hirsch, C.: Verification, Validation, and Predictive Capability in Computational Engineering and Physics. SAND REPORT 2003-3769, Albuquerque and Livermore, Sandia National Laboratories (2003)

Oden, J.T., Babuska, I., Nobile, F., Feng, Y., Tempone, R.: Theory and methodology for estimation and control of errors due to modeling, approximation, and uncertainty. Comput. Methods Appl. Engrg. 194, 195–204 (2005)

O'Rourke, J.: Computational Geometry in C. Cambridge University Press, Cambridge (1998)

Papoulis, A., Pilai, S.U.: Probability, Random Variables, and Stochastic Processes. McGraw-Hill, Boston (2002)

Parès, N., Bonet, J., Huerta, A., Peraire, J.: The computation of bounds for linear-functional outputs of weak solutions to the two-dimensional elasticity equations. Comput. Methods Appl. Mech. Engrg. 195, 406–429 (2006)

Pion, S.: CGAL Open Source Project (2007), `http://www.cgal.org/` (retrieved June 27, 2009)

Ratschek, H., Rokne, J.: Computer Methods for the Range of Functions. Halsted Press, Chichester (1984)

Ratschek, H., Rokne, J.: New Computer Methods for Global Optimization. Ellis Horwood, Chichester (1988)

Reagan, R., Mosteller, F., Youtz, C.: Quantitative meanings of verbal probability expressions. J. Applied Psychology 74, 433–442 (1989)

Ribacchi, R.: Recenti orientamenti nella progettazione statica delle gallerie. In: Proc. XVIII Convegno Nazionale di Geotecnica, Rimini, May 11-13, vol. 2, pp. 37–94. Associazione Geotecnica Italiana, Roma (1993)

Robbins, H.E.: On the measure of random set, I. Annals of Mathematical Statistics 15, 70–74 (1944)

Robbins, H.E.: On the measure of random set, II. Annals of Mathematical Statistics 16, 342–347 (1945)

Sack, J.-R., Urrutia, J. (eds.): Handbook of Computational Geometry. Elsevier, Amsterdam (2000)

Sauer-Budge, A.M., Bonet, J., Huerta, A., Peraire, J.: Computing bounds for linear functionals of exact weak solutions to Poisson's equation. SIAM Journal on Numerical Analysis 42, 1610–1630 (2004)

Schilling, M.F.: Some remarks on quick estimation of the correlation coefficient. The American Statistician 38(4), 330–331 (1984)

Sentz, K., Ferson, S.: Combination of Evidence in Dempster-Shafer Theory. Sandia National Laboratories Report SAND2002-0835, Albuquerque, New Mexico (2002)

Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)

The Geometry Center. Qhull (2007), `http://www.qhull.org/` (retrieved June 27, 2009)

Tonon, F.: Using Random Set Theory to propagate epistemic uncertainty through a mechanical system. Reliability Engineering & System Safety 85, 169–181 (2004)

Tonon, F.: Some properties of a Random Set approximation to upper and lower distribution functions. Int. J. Approximate Reasoning 48, 174–184 (2008)

Tonon, F., Bernardini, A.: A Random Set approach to the optimization of uncertain structures. Computers and Structures 68, 583–600 (1998)

Tonon, F., Bernardini, A.: Multiobjective optimization of uncertain structures through Fuzzy Set and Random Set theory. Computer-Aided Civil and Infrastructure Engineering 14, 119–140 (1999)

Tonon, F., Pettit, C.: Structural reliability application of correlation in random-set valued variables. In: Proc. 9th ASCE Joint Specialty Conf. on Probabilistic Mechanics and Structural Reliability, PMC 2004, Albuquerque, NM, ASCE (2004)

Tonon, F., Pettit, C.: Toward a definition and understanding of correlation for variables constrained by random relations. In: Proc. 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference and 7th AIAA Non-Deterministic Approaches Forum, Austin, TX (2005)

Tonon, F., Pettit, C.: Toward a definition and understanding of correlation for variables constrained by random relations. In: Proc. ICOSSAR 2005, Universita' La Sapienza, Rome, Italy (2005)

Tonon, F., Chen, S.: Inclusion properties for random relations under the hypotheses of stochastic independence and non-interactivity. Int. J. of General Systems 34(5), 615–624 (2005)

Tonon, F., Bernardini, A., et al.: Determination of parameters range in Rock Engineering by means of Random Set Theory. Reliability Engineering & System Safety 70(3), 241–261 (2000)

Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Walley, P.: Towards a unified theory of imprecise probabilities. Int. J. Approximate Reasoning 24, 125–148 (2000)

Warmus, M.: Calculus of approximations. Bull. Acad. Pol. Sci., Cl. III 4, 253–259 (1956)

Wasserman, L.A., Kadane, J.B.: Bayes' Theorem for Choquet Capacities. The Annals of Statistics 18(3), 1328–1339 (1990)

Woods, J., Gabbay, D.M. (eds.): Handbook of the history of logic. Elsevier – North Holland, Amsterdam (2004)

Vicig, P.: Epistemic independence for imprecise probabilities. In: ISIPTA 1999 - Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications, Zwijnaarde, Belgium (1999)

Vick, S.G.: Considerations for Estimating Structural Response Probabilities in Dam Safety Risk Analysis. Final Technical Report, Order no. 99DG81029. Denver (CO), U.S. Bureau of Reclamation, Technical Service Center (1999)

Xiang, G.: Fast algorithm for computing the upper endpoint of sample variance for interval data: case of sufficiently accurate measurements. Reliable Computing 12, 59–64 (2006)

Xuan, Z.C., Parés, N., Peraire, J.: Computing upper and lower bounds for the J-integral in two-dimensional linear elasticity. Computer Meth. in Appl. Mech. and Engng. 195, 430–443 (2006)

Yager, R.R.: The entailment principle for Dempster-Shafer granules. Int. J. Intelligent Systems 1, 247–262 (1986)

Yager, R.R.: On the Dempster-Shafe framework and new combination rules. Information Sciences 41, 93–137 (1987)

Yager, R.R.: On probabilities induced by multi-valued mappings. Fuzzy Sets and Systems 42, 301–314 (1991)

Yager, R.R.: Aggregating fuzzy sets represented by belief structures. Journal of Intelligent and Fuzzy Systems 1, 215–224 (1993)

Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)

Zadeh, L.A.: The concept of a linguistic variable, and its applications to approximate reasoning. Information Sciences 8, 199–249 (1975)

Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1(1), 3–28 (1978)

Zadeh, L.A.: A mathematical theory of evidence (book review). AI Magazine 5, 81–83 (1984)

Zimmermann, H.-J.: Fuzzy Set Theory and Its Applications, 2nd edn. Kluwer, Dordrecht (1991)

# Subject Index