Probabilistic Analysis of Dempster-Shafer Theory. Part One

Kramosil, Ivan
1997

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Probabilistic Analysis of Dempster–Shafer Theory
Part One

Ivan Kramosil

Technical report No. 716

September, 1997

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+4202) 6605 3940 fax: (+4202) 8585789
e-mail: kramosil@uivt.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Probabilistic Analysis of Dempster–Shafer Theory Part One

Ivan Kramosil[1]

Technical report No. 716
September, 1997

Abstract

Dempster–Shafer theory is an interesting and useful mathematical tool for uncertainty quantification and processing. From one point of view it can be seen as an alternative apparatus to probability theory and mathematical statistics based on this probability calculus, as D.–S. theory can be developed in a way quite independent of probability theory, beginning with a collection of more or less intuitive demands which an uncertainty degree calculus should meet. On the other side, however, D.–S. theory can be developed also as a particular sophisticated application of probability theory, using the notion of non–numerical, in particular, set–valued random variables (random sets) and their numerical characteristics. This later aspect enables to generalize D.–S. theory beyond its classical scopes using appropriately the apparatus of probability theory and measure theory.

This report is the first part of a surveyal work cumulating, and presenting in a systematic way, some former author's ideas and achievements dealing with applications of probability theory and mathematical statistics when defining, developing, and generalizing various parts of D.–S. theory. The more detailed contents of this report can be understood from the list of the titles of the particular chapters presented just below.

Keywords

Dempster–Shafer theory, probability theory, belief function, random variable, random set

[1]This work has been sponsored by the grant no. A1030504 of the GA AS CR.

1 Introduction

From a certain point of view, position of every individual human being, every society, and the mankind as a whole in the surrounding us world can be seen as that of the subject of a continual sequence of decision making acts, terminated by one's death in the case of an individual, and perhaps infinite in the case of a collective agent. Just a very small portion of the decision problem we have to solve are of the deterministic nature when the consequences of the accepted decision can be completely, and with the absolute degree of certainty, foreseen so that we can choose a good, appropriate, acceptable, the best, or almost the best solution, supposing that the adjectives just introduced are sensefully, and with a sufficient degree of preciseness defined. As good examples of uncertainty-free decision procedures let us mention those ones applied in an artificial environment when any influence of uncertainty is avoided a priori. Decision-making in mathematics, or in formal systems in general, or games like chess can be remembered here. One common feature of all these cases consists in the fact that if the decision making fails, e. g., if the accepted decision is a posteriori proved not to possess the expected properties, then the only source of this failure can consist in an error made during the realization of the decision procedure (a computational error or a wrong step in a mathematical proof, for example).

The situation in scientific branches like physics, astronomy, or chemistry is seemingly only slightly different from that in the deductive fields mentioned above, but the shift is very important from the methodological and philosophical point of view. Astronomy, for example, pretends on the same degree of certainty in its decisions and predictions as the deductive sciences like mathematics, and uses the mathematical tools in the most wide scope in order to achieve this goal. However, when some decision or prediction fails, a wrong application of the decision procedure in question, even if it cannot be a priori avoided as a source of this failure, need not be the only possible cause. Or, there may exist some circumstance or phenomenon not known before and not taken, therefore, into consideration when creating the decision procedure, but influencing significantly the correctness or acceptability of the taken decision. E. g., if a planet of the Solar system is not situated at the position forecast by the computation of heaven mechanics, the possibility of an error made during the computations cannot be, of course, a priori avoided. It is also possible, however, that the difference between the predicted and the actual position of the investigated planet is caused by some not discovered yet new planet. From the methodological point of view, this phenomenon can be expressed as follows: the decision about the present position of the planet in question was charged by some (degree or portion of) uncertainty, and this uncertainty was caused by the incompleteness of the information used during the process of decision making. As we shall see below, this interpretation will be of crucial importance in what follows. It is also why we omit a more detailed description of the further way of entering uncertainties of various kinds and proveniences into our prediction and decision-making processes terminated in one's everyday decision making charged by all the uncertainty and vagueness of the real world around us.

Several attempts to build up a consistent and powerful enough mathematical theory of probability, taken as a tool for uncertainty quantification and processing, have resulted in the well-known and in our days already classical Kolmogorov axiomatic probability theory [22]. This theory admits (and, as a matter of fact, has been motivated by) the interpretation of uncertainty as the lack of complete information. In other words said, all the dependences and relations among various phenomena and among values taken by various variables are conditionally deterministic in the sense that occurrences of all phenomena and values taken by all variables are completely defined by the value of a universal hidden parameter which can be understood as "the actual state of the Universe", or "all the history of the Universe since the Big Bang", or in a similar way. The problem is, however, that the actual value of this parameter cannot be identified and the only what we have at our disposal is the information that this value belongs to a subset of possible values of this parameter and that the size of this subset can be numerically quantified by a number from the unit interval of reals, or at least the value of this size can be more or less exactly estimated. As all the sets to which the unknown values of the hidden parameter can be proved to belong are classical crisp sets, and their sizes are quantified by (standard) real numbers, all the tools developed by the classical mathematics, namely set theory, measure theory and the theory of real functions were at the disposal when building the axiomatic probability theory, and they have been, in fact, widely and sophistically used for these purposes.

Let us recall that also the great success of the notion of fuzzy sets has been caused by the fact that this notion can be almost trivially defined and processed using elementary classical mathematical tools – crisp sets and functions taking these sets into the unit interval of real numbers, or into a more general but also already classical structure like Boolean algebra or lattice. On the other side, the theory of semisets, challenged by the same practical and theoretical problems as the theory of fuzzy set but based on a deeply going revision of the set-theoretic, logical and methodological foundations of classical mathematics, has been always situated at the very margins of interest of specialists dealing with applications of mathematical methods in extra-mathematical fields.

Since the moment, sometimes in the middle of the sixtieths, when fuzzy sets emerged, a number of other formal tools for uncertainty quantification and processing have appeared, alternative to the probability theory, Dempster–Shafer theory being one of them. The reasons for which probability theory has lost its position of the unique and universal tool for the purposes of uncertainty quantification and processing are at least twofold; philosophical and methodological at the one side, and practical on the other side.

(i) A more detailed investigation puts into doubts the idea that all the uncertainty in the world around us is of the same nature and can be successfully processed by the tools developed in order to treat a rather special case of uncertainty perhaps identifiable with the notion of statistical or stochastical randomness. The phenomena of vagueness, ambiguity or nonspecificity deserve, perhaps, alternative and special tools to be processed successfully. E. g., there is a qualitative difference between the uncertainty charging the result of a coin tossing (we are not sure which side of the coin occurs) and the uncertainty which side of a coin we are observing (head or tail) supposing that this coin has been digged out in a very damaged state during an archeological investigation, so that the original tails on both the sides of the coin can be hardly identified and distinguished from each other.

Dempster–Shafer theory can be seen as a mathematical model for uncertainty quantification and processing which quantifies the degrees of uncertainty by real numbers from the unit interval. Generalizations to values outside this interval are also possible, interesting, and worth being studied, and will be also introduced in this study. However, the additivity of probabilistic measures is intentionally abandoned and an intuitive interpretation of degrees of uncertainty justifying such a modification is submitted. From the methodological point of view, one feature of Dempster–Shafer theory is interesting and important in our context.

Dempster–Shafer theory can be developed quite independently of probability theory in such a way that a number of more or less natural, intuitive and acceptable demands, which an uncertainty quantification should satisfy, are stated, mathematically formalized, and justified from the point of view of their possible interpretation. Then, the uncertainty degrees investigated and processed in Dempster–Shafer theory are proved to satisfy necessarily certain properties, including, in general, also non-additivity supposing that the imposed demands hold, playing the role of axioms, say. On the other side, however, Dempster–Shafer theory can be obtained as a non-traditional and sophisticated applications of probability theory, when there are not the random events themselves, but other events induced and defined by the original ones, the sizes of which are probabilistically quantified. This approach enables to apply all the powers of the mathematical apparatus developed by probability theory and, more generally, by measure theory, without giving up the possibility of an appropriate non-traditional interpretation of the processed probabilities. E. g., the generalization of Dempster–Shafer theory to the case when uncertainty degrees are beyond the unit interval of real numbers, mentioned above, can be obtained as immediate generalizations of the results obtained for probability measures to more general measures or set functions. Also the generalization of Dempster–Shafer theory to infinite sets and various approximations of uncertainty degrees processed by this theory can be easily developed and processed in the terms of probability theory due to the generalizations of certain methods and results of probability theory and measure theory.

Since several years, the author of this study has published a number of papers and conference contributions dealing with applications of probability theory in order to define, analyze and perhaps generalize and enrich various particular parts or aspects of Dempster–Shafer theory. This work does not bring too many new results when compared with these former papers, but it has been motivated by the aim to present the ideas and results systematically and without any preliminaries which are more or less necessary and tacitly assumed in any special, thematically and/or by its extent limited paper. Hence, this text should be readable for each reader on a more or less elementary level of mathematical education and culture. It is why the work begins with introductory chapters dealing with the elements of probability theory and probabilistic and statistical approaches to decision making under uncertainty. On the other side, the study has been conceived as a mathematical and theoretical one, so that the reader seeking for, say, numerical and practical examples, should consult a more practically oriented book or paper dealing with Dempster–Shafer theory.

Cf. also [9, 44] and [29] for the philosophical bakcgrounds of mathematics in general and uncertainty quantification and processing in particular.

2 The Most Elementary Preliminaries on Axiomatic Probability Theory

This work has been conceived as a purely theoretical and mathematical study dealing with the subject of its interest at a highly abstract and formalized level. Probability theory will serve as one and, as a matter of fact, the most important and the most powerful formal tool used below in order to achieve this goal. Therefore, beginning with a brief survey of the most elementary notions of probability theory, just the most elementary abstract ideas and construction of the axiomatic probability theory, as settled by Kolmogorov in [22] are presented in this chapter, intentionally leaving aside all the informal discussions, motivations, and practical examples preceding the formalized explanations of probability theory in the greatest part of textbooks and monographs dealing with this theory. The reader interested in these informal parts of probability theory is kindly invited to consult an appropriate textbook or monograph, let us mention explicitly the already classical textbooks [8] and [11], where just these informal parts are explained very carefully, in detail, and with a lot of various examples. On the other side, [31] treates probability theory at an exclusively abstract and formalized level.

Let us begin our explanation with the basic notions of σ-field, probability measure and probability space.

Definition 2.1. Let Ω be a nonempty abstract set, let $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ denote the power-set of all subsets of the set Ω, also denoted by 2^Ω (this system of sets is taken as set due to the axiomatics of set theory in its Zermelo–Fraenkel as well as in the Goedel–Bernays setting, cf. [10] or [1]). A nonempty system $\mathcal{A} \subset \mathcal{P}(\Omega)$ of subsets of Ω is called *sigma-field* (*σ-field* or σ-algebra), if for each sets A, A_1, $A_2, \ldots$ from $\mathcal{A}$ also the sets $\Omega - A$ (the complement of A) and $\bigcup_{i=1}^{\infty} A_i$ (the union of the sets A_1, $A_2, \ldots$) belong to $\mathcal{A}$. In other words, σ-field is a nonempty system of subsets closed with respect to the set-theoretic operations of complement and countable union. $\square$

As can be easily seen, each σ-field $\mathcal{A}$ of a nonempty set Ω contains the whole space Ω, the empty subset $\emptyset$ of Ω, and it is closed with respect to finite as well as infinite countable unions and intersections as well as with respect to relative complements $A - B$ for all $A,\ B \in \mathcal{A}$. Or, by definition, $\mathcal{A}$ is nonempty, hence, for $A \subset \Omega$, $A \in \mathcal{A}$, also $\Omega - A \in \mathcal{A}$, and $\Omega = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, where $A_1 = A$ and $A_i = \Omega - A$ for each $i \geq 2$. Consequently, $\emptyset = \Omega - \Omega \in \mathcal{A}$ holds, and $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} B$, where $B_i = A_i$, if $i \leq n$, and $B_i = \emptyset$ otherwise, belongs to $\mathcal{A}$ as well. De Morgan rules, enabling to write $\bigcap_{i=1}^{\infty} A_i$ as $\Omega - (\bigcup_{i=1}^{\infty}(\Omega - A_i))$, prove that $\mathcal{A}$ is closed with respect to countable (and trivially also finite) intersections. If $A,\ B \in \mathcal{A}$, then $A - B = A \cap (\Omega - B)$ is also in $\mathcal{A}$.

Definition 2.2. Let Ω be a nonempty set, let $\mathcal{A} \subset \mathcal{P}(\Omega)$ be a σ-field of subsets of Ω. The ordered pair $\langle \Omega, \mathcal{A} \rangle$ is then called the *measurable space* generated in the set Ω by the σ-field $\mathcal{A}$ of its subsets. $\square$

Definition 2.3. Let $\langle \Omega, \mathcal{A} \rangle$ be a measurable space. A mapping P ascribing to each set $A \in \mathcal{A}$ a real number $P(A)$ from the unit interval of real numbers, in symbols, $P : \mathcal{A} \to \langle 0, 1 \rangle$, is called (*$\sigma$-additive*) *probability measure* defined on the measurable space $\langle \Omega, \mathcal{A} \rangle$, if

(i) $P(\Omega) = 1$, and

(ii) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)P$ for each sequence $A_1,\ A_2, \ldots$ of mutually disjoint (i. e., $A_i \cap A_j = \emptyset$ for all $i,\ j \geq 1,\ \ i \neq j$) sets from $\mathcal{A}$.

In more detail: for each such sequence of sets from $\mathcal{A}$ the series $\sum_{i=1}^{\infty} P(A_i) =$ $= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i)$ is defined and its value equals to $P\left(U_{i=1}^{\infty} A_i\right)$. The ordered triple $\langle \Omega, \mathcal{A}, P \rangle$ is called *probability space*. $\square$

Let us postpone some interpretation remarks a few lines below, just after having introduced the other most important notion of the axiomatic probability theory, namely that of random variable, and its immediate generalization to the notion of generalized random variable.

Definition 2.4. *Borel line* is the measurable space $\langle R, \mathcal{B} \rangle$, where $R = (-\infty, \infty)$ is the space of all real numbers and $\mathcal{B}$ is the σ-field of Borel subsets of R, i. e., the minimal σ-field of subsets of R containing all semi-open intervals $\langle a, b) = \{x \in R : a \leq x < b\},\ \ a \leq b,\ \ a,\ b \in R$ (or, what turns to be the same, the minimal σ-field containing all half-lines $(-\infty, a),\ \ a \in R$, a number of equivalent definitions is also possible). $\square$

Definition 2.5. Let $\langle \Omega, \mathcal{A}, P \rangle$ be a probability space. A total mapping $X : \Omega \to R - (-\infty, \infty)$ is called (*real-valued*) *random variable*, if it is measurable with respect to the σ-field $\mathcal{B}$ of Borel subsets of R, i. e., if for each Borel set $B \in \mathcal{B}$ its inverse image w. r. to X belongs to $\mathcal{A}$, in symbols, if

$$\{\{\omega \in \Omega : X(\omega) \in B\} : B \in \mathcal{B}\} \subset \mathcal{A}. \tag{2.1}$$

As can be easily proved, (2.1) holds iff

$$\{\{\omega \in \Omega : X(\omega) < a\} : a \in R\} \subset \mathcal{A} \tag{2.2}$$

holds. □

An easy checking of the definition just introduced proves that it is just the property that Borel sets form a σ-field which is used here, taking abstraction, in the same time, of all other specific properties of real numbers and their sets. So, we can replace the Borel line in Definition 2.5 by a general case of measurable space.

Definition 2.6. Let $\langle \Omega, \mathcal{A}, P\rangle$ be a probability space, let $\langle Y, \mathcal{Y}\rangle$ be a measurable space. A total mapping $X : \Omega \rightarrow Y$ is called *generalized (Y-valued) random variable*, if it is measurable with respect to the σ-field $\mathcal{Y}$ of subsets of Y, i. e., using the notation as in (2.1), if

$$\{\{\omega \in \Omega : X(\omega) \in Z\} : Z \in \mathcal{Y}\} \subset \mathcal{A}. \tag{2.3}$$

□

Definition 2.7. Let X be a generalized random variable defined on a probability space $\langle \Omega, \mathcal{A}, P\rangle$ and taking its values in a measurable space $\langle Y, \mathcal{Y}\rangle$. The mapping $P_X : \mathcal{Y} \rightarrow \langle 0, 1\rangle$, defined for each $Z \in \mathcal{Y}$ by

$$P_X(Z) = P(\{\omega \in \Omega : X(\omega) \in Z\}) \tag{2.4}$$

is called the *induced probability measure (induced by X) on the σ-field Y*. If $\langle Y, \mathcal{Y}\rangle = \langle R, \mathcal{B}\rangle$, then the *distribution function F_X of the (real-valued) random variable* is the mapping $F_X : R \rightarrow \langle 0, 1\rangle$ defined by $F_X(a) = P(\{\omega \in \Omega : X(\omega) < a\}) = P_X((-\infty, a))$. □

As can be easily proved, P_X is indeed a probability measure on $\mathcal{Y}$, moreover, in the particular case of the Borel line, the induced probability distribution P_X on $\mathcal{B}$ is uniquely defined by the distribution function. Let us also remark, that the most often investigated generalized random variables, in what follows, will be the set-valued ones, when $Y = \mathcal{P}(S)$ will be the power-set of all subsets of a nonempty set S, and $\mathcal{Y} \subset \mathcal{P}(\mathcal{P}(S))$ will be a σ-field of families of subsets of S, in the case when S is finite most often $\mathcal{Y} = \mathcal{P}(\mathcal{P}(S))$ (which is obviously also finite).

Even if we promised to present the elementary ideas of probability theory on a highly abstract level, a small intuitive reconsideration seems to be worth introducing just now. Many specialists, which are not professionally trained in probability theory and want to get familiar with its foundations in order to apply them in their own fields of research, take the constructions based on the notion of probability space and (generalized) random variable as too complicate and propose to begin our formal processing of uncertainty with the induced probability measures defined on the corresponding particular spaces of outcomes of their observations or experiments charged by uncertainty. E. g., they begin their probabilistic description of the experiment consisting in a regular dice tossing by the probability distribution on the six-element set $\{1, 2, \ldots, 6\}$

of possible outcomes ascribing to each result the same probability 1/6, without taking into consideration an abstract probability space $\langle \Omega, \mathcal{A}, P \rangle$ and a random variable $X : \Omega \to R$ such that $P(\{\omega \in \Omega : X(\omega) \in \{i\}\}) = 1/6$ for each $i = 1, 2, \ldots, 6$. However, the idea of probability space enables to pick out or "to put before brackets" our limited abilities to quantify and process probabilities as something a priori given so that it is beyond our powers to enrich it and we are able to quantify and process probabilities of some empirical events just when we are able to express them, through appropriate random variables, as probabilities of some random events, i. e., sets belonging to $\mathcal{A}$, of the fixed probability space being at our disposal. For example, if $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{100}\}$ contains 100 elements and $\mathcal{A}$ is the field (and obviously also σ-field) generated by atoms of the kind $\{\omega_1, \omega_2, \ldots, \omega_{10}\}$, $\{\omega_{11}, \omega_{12}, \ldots, \omega_{20}\}, \ldots, \{\omega_{91}, \omega_{92}, \ldots, \omega_{100}\}$ with a probability measure P defined on $\mathcal{A}$, we are not able, given the probability space $\langle \Omega, \mathcal{A}, P \rangle$, to describe in detail an experiment the results of which can be positive integers from 1 to 1 000 with a positive probability p_i for each $i \leq 1\,000$. Such a restriction can be given by our limited abilities to process probabilistic distributions over more than ten-element spaces, to obtain particular values p_i for each $i \leq 1\,000$, or our limited abilities to distinguish from each other the elements in each of the class $\{\omega_{i_n+1}, \ldots, \omega_{i(n+1)}\}$ for $i = 1, \ldots, 10$, $n = 0, \ldots, 9$. On the other side, the idea of probability space enables to define, even if not always to compute effectively, the probabilities of combined random events defined by particular results of arbitrarily different extra-mathematical nature ("to add apples and pears"). E. g., probabilities of combined results obtained when tossing simultaneously more than one dice can be defined on the same probability space as in the case of a singular tossing.

There are also some intuitive reasons for which just σ-field was chosen as an appropriate structure over the system of random events. It is natural to assume that the phenomenon consisting in the fact that a random event did not occur should be also classified as random event, the same being the case of the simultaneous occurrence of a finite collection of random events or of the occurrence of at least one from a finite collection of random events. The condition according to which σ-field is closed with respect to infinite (countable) unions and, consequently, intersections, enables to define as random events also phenomena defined by infinite sequences of random events like, e. g., "in an infinite sequence of coin tosses sooner or later three immediately following occurrences of head occur". Finite additivity of probability measure is naturally motivated by the finite additivity of relative frequences of occurrences of disjoint random events, its supposed strengthening to the case of σ-additivity is a matter of rather technical routine enabling to define and compute probabilities of random events defined by infinite sequences of more elementary random events using the same rules and ideas as in the finite cases.

The following definition introduces a very important notion of conditional probability which is sometimes considered as the axiom of probability theory (from the point of view of metamathematical principles according to which formalized theories are built the difference between axioms and definitions is not principal in general). We shall see below in this chapter, that conditional probabilities can be defined as very particular cases of conditional expected values.

Definition 2.8. Let $\langle \Omega, \mathcal{A}, P \rangle$ be a probability space, let $A,\ B \in \mathcal{A}$ be random events such that $P(B) > 0$ holds. Then the *conditional probability of (the random event) A given* (or: *under the condition of the random event*) B is defined by $P(A/B) = P(A \cap B)/P(B)$, $P(A/B)$ being undefined otherwise, i.e. if $P(B) = 0$. A slightly modified and shifted definition reads as follows: $P(A/B)$ is a real number satisfying the equality $P(A \cap B) = P(A/B)\, P(B)$; if $P(B) = 0$, then obviously $P(A \cap B) = 0$ as well and the equality holds for each value $P(A/B) \in \langle 0, 1 \rangle$. □

As can be easily seen, for each $B \in \mathcal{A}$ such that $P(B) > 0$, the mapping $P(\cdot/B) : \mathcal{A} \to \langle 0, 1 \rangle$ is a probability measure defined on the σ-field $\mathcal{A}$.

Very important and often used numerical characteristics of (real-valued) random variables are their absolute and central moments, namely the first absolute and the second central ones.

Definition 2.9. Let X be a (real-valued) random variable defined on a probability space, let $k \geq 0$ be a non-negative integer. *The k-th absolute moment* (or: *the absolute moment of the order k*) *of (the random variable) X* is the number $M_k(X)$ from the extended real line $R^* = \langle -\infty, \infty \rangle = R \cup \{-\infty\} \cup \{\infty\}$ defined by

$$M_k(X) = \int_{-\infty}^{\infty} X^k \, \mathrm{d}P_X \tag{2.5}$$

supposing that this number exists. The k-th central moment (or: *the central moment of the order k*) *of (the random variable) X* is the number $M_k^0(X)$ from the extended real line R^* defined by

$$M_k^0(X) = \int_{-\infty}^{\infty} (X - M_1(X))^k \, \mathrm{d}P_X \tag{2.6}$$

supposing that this number exists. The first absolute moment $M_1(X)$ is called the *expected value of (the random variable X)* and it is denoted by EX or $E(X)$, the second central moment $M_2^0(X)$ is called the *dispersion of (the random variable) X* and it is denoted by $D^2\, X$ or $D^2(X)$. Hence,

$$EX = \int_{-\infty}^{\infty} X \, \mathrm{d}P_X, \tag{2.7}$$

$$D^2(X) = \int_{-\infty}^{\infty} (X - EX)^2 \, \mathrm{d}P_X, \tag{2.8}$$

supposing that these two numbers are defined (in the extended real line). □

The most important and most often used characteristics of groups of (generalized) random variables, or of relations among such variables, are the relations of identical distribution and of statistical (stochastical) independence.

Definition 2.10. Let $\mathcal{X}$ be a collection of (generalized) random variables defined on a probability space $\langle \Omega, \mathcal{A}, P \rangle$, each of them taking its values in a measurable space $\langle Y, \mathcal{Y} \rangle$. (Generalized) random variables from $\mathcal{X}$ are called *identically distributed* (or: *$\mathcal{X}$ is a collection of identically distributed (generalized) random variables*), if for each $X_1, X_2 \in \mathcal{X}$ and each $Z \in \mathcal{Y}$ the equality

$$P\left(\{\omega \in \Omega : X_1(\omega) \in Z\}\right) = P\left(\{\omega \in \Omega : X_2(\omega) \in Z\}\right) \tag{2.9}$$

holds. □

Definition 2.11. Let $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$, $n \geq 1$, be a finite collection of (generalized) random variables, each X_i taking its values in a measurable space $\langle Y_i, \mathcal{Y}_i \rangle$ and being defined on the same probability space $\langle \Omega, \mathcal{A}, P \rangle$. The (generalized) random variables $X_1, X_2, \dots, X_n$ are called *statistically* (or: *stochastically*) *independent* (or: $\mathcal{X}$ is called a *collection of statistically or stochastically independent (generalized) random variables*), if for each sequence $\langle Z_1, Z_2, \dots, Z_n \rangle$ of sets such that $Z_i \in Y_i$ holds for each $i \leq n$, the equality

$$P\left(\bigcap\nolimits_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \in Z_i\}\right) = \prod\nolimits_{i=1}^{n} P\left(\{\omega \in \Omega : X_i(\omega) \in Z_i\}\right) \tag{2.10}$$

holds. $\langle X_1, X_2, \dots \rangle$ is an *infinite sequence of statistically* (or: *stochastically*) *independent (generalized) random variables*, if each finite subsequence of $\langle X_1, X_2, \dots \rangle$ defines a collection of statistically (or: stochastically) independent (generalized) random variables in the sense of (2.10). Applying these notions to the particular case of characteristic functions (or: identifiers) $\chi_{A_1}, \chi_{A_2}, \dots, \chi_{A_n}$ of random events $A_1, A_2, \dots, A_n \in \mathcal{A}$, where $\chi_A(\omega) = 1$, if $\omega \in A$ and $\chi_A(\omega) = 0$ otherwise, we obtain that *random events $A_1, A_2, \dots, A_n$ are statistically* (or: *stochastically*) *independent*, if the equality

$$P\left(\bigcap\nolimits_{i=1}^{n} A_i\right) = \prod\nolimits_{i=1}^{n} P(A_i) \tag{2.11}$$

holds; the generalization to the case of an infinite sequence $A_1, A_2, \dots$ of random events is obvious. □

Finite sequences of identically distributed and statistically independent random variables (i.i.d. sequences) play an important role of a bridge between the empirical data obtained on the ground of a finite empirical experience (observations, results of experiments, ...), and the idealized values of probabilities and expected values (cf. [26], e. g. for a more detailed philosophical and methodological discussion). The role of mathematically formalized counterpart of this bridge, enabling to define, in which sense and degree a finite experience (the average value of a finite number of numerically quantified observations, say) can be taken as a good approximation of an ideal characteristics (expected value, say), is played by the *laws of large numbers*. Let us introduce here just one such law, the *strong law of large number*, in the setting simple enough, but sufficient for our purposes. Its more general formulations consists in various kinds of weakenings of its conditions and can be found in all more detailed textbooks or monographs on probability theory; let us recall [31, 8] or [11] just as examples.

Fact 2.1. (A simple formulation of the strong law of large numbers) Let $\langle X_1, X_2, \ldots \rangle$ be an infinite i.i.d. sequence of (real-valued) random variables defined on a probability space $\langle \Omega, \mathcal{A}, P \rangle$, let $|EX_1| < \infty$, let $D^2 X_1 < \infty$. Then

$$P\left(\{\omega \in \Omega : \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} X_i(\omega) = EX_1\}\right) = 1. \tag{2.12}$$

□

Informally: the average value taken from a finite sequence of realizations of statistically independent copies (repetitions) of a random variable X_1 tends almost surely (with the probability one) to the expected value of the random variable X_1. In other words: there are rational reasons to take average value of a large enough sequence of statistically independent realizations as a more or less good approximation or estimate of the expected value of the random variable in question.

The laws of large number including that one just introduced are assertions of limit nature. The degree in which a particular average value (average mean) approximates the expected value in question is quantified by the well-known Chebyshev inequality; let us present here, again, its simple form sufficient enough for our purposes, referring the reader to the books already mentioned or to other sources for some more strong versions of the Chebyshev inequality.

Fact 2.2. (Chebyshev inequality) Let $\langle X_1, X_2, \ldots \rangle$ be the same sequence as in Fact 2.1. Then

$$P\left(\left\{\omega \in \Omega : \left|n^{-1} \sum_{i=1}^{n} X_i(\omega) - EX_1\right| > \varepsilon\right\}\right) < (D^2 X_1)/n\,\varepsilon^2 \tag{2.13}$$

holds for each $n = 1, 2, \ldots$ and each $\varepsilon > 0$. □

In the particular case, when each random variable X_i is identically distributed with the characteristic function (identifier) χ_A of a random event $A \in \mathcal{A}$, i.e., when $P(\{\omega \in \Omega : X_i(\omega) = 1\}) = P(\{\omega \in \Omega : \chi_A(\omega) = 1\}) = P(A)$ and $P(\{\omega \in \Omega : X_i(\omega) = 0\}) = P(\{\omega \in \Omega : \chi_A(\omega) = 0\}) = 1 - P(A)$ hold for each $i = 1, 2, \ldots$, (2.12) and (2.13) reduce to

$$P\left(\left\{\omega \in \Omega : \lim_{n \to \infty} n^{-1} m_n = P(A)\right\}\right) = 1, \tag{2.14}$$

$$P\left(\{\omega \in \Omega : |m_n/n - P(A)| > \varepsilon\}\right) < (D^2 X_n)/n\,\varepsilon^2 < 1/4\,n\,\varepsilon^2, \tag{2.15}$$

where $m_n = m_n(\omega)$ is the number of occurrences of the random event A in a sequence of n statistically independent and identically distributed random experiments each of them taking as its result either A or non A $(\Omega - A)$. The general upper bound in (2.15), independent of $D^2 X_1$, follows from the almost obvious fact that for each random variable X taking as its values only 0 or 1 the inequality $D^2 X \leq 1/4$ holds.

The reader is warned not to overestimate the meaning of strong laws of large numbers, being perhaps suggestively fascinated by the wording "almost surely" used in their formulations. An interesting and more deeply going discussion about the nature of strong laws of large numbers can be found in [9].

Let us close this chapter by the notion of conditional expected value, as both the notions of expected value and that of conditional probability, introduced above, can be obtained as very special cases of this general notion.

Definition 2.12. Let $\langle \Omega, \mathcal{A}, P \rangle$ be a probability space, let $\mathcal{B} \subset \mathcal{A}$ be a nonempty sub-σ-field of $\mathcal{A}$, let X be a (real-valued) random variable defined on $\langle \Omega, \mathcal{A}, P \rangle$, let $P_{\mathcal{B}}$ be the restriction of P to the sub-σ-field $\mathcal{B}$, i.e., $P_{\mathcal{B}} = P(A)$ for each $A \in \mathcal{B}$, $P_{\mathcal{B}}(A)$ being undefined for $A \in \mathcal{A} - \mathcal{B}$. A real-valued random variable $E^{\mathcal{B}} X$ defined on the probability space $\langle \Omega, \mathcal{B}, P_{\mathcal{B}} \rangle$ and such that the equality

$$\int_B (E^{\mathcal{B}} X) \, \mathrm{d}P_{\mathcal{B}} = \int_B X \, \mathrm{d}P \tag{2.16}$$

holds for each $B \in \mathcal{B}$ is called the *conditional expected value of (the random variable) X given* (or: *with respect to*) *(the sub-σ-field) $\mathcal{B} \subset \mathcal{A}$.* □

The well-known Radon–Nikodym theorem implies (cf. [15] or [31]), that the definition just introduced is not vacuous and that the conditional expected value $E^{\mathcal{B}} X$ exists. Moreover, it is defined uniquely up to a null set from $\mathcal{B}$, in more detail, if $E_1^{\mathcal{B}} X$ and $E_2^{\mathcal{B}} X$ both satisfy (2.16) for all $B \in \mathcal{B}$, then there exists $B_0 \subset \Omega$, $B_0 \in \mathcal{B}$, such that $P_{\mathcal{B}}(B_0)$ $(= P(B_0)) = 0$ and the inclusion

$$\left\{\omega \in \Omega : (E_1^{\mathcal{B}} X)(\omega) \neq (E_2^{\mathcal{B}} X)(\omega)\right\} \subset B_0 \tag{2.17}$$

holds.

Let $\mathcal{B} = \{\emptyset, \Omega\}$ be the minimal (the most rough) nonemty sub-σ-field of $\mathcal{A}$. Consequently, $E^{\mathcal{B}} X$ must be constant on each atomic set of $\mathcal{B}$ and there is just one such set, namely, whole the space Ω, so that $(E^{\mathcal{B}_0} X)(\omega) = EX$ (if EX exists) for all $\omega \in \Omega$. $E^{\mathcal{B}_0} X$ is defined unambiguously, as there is no nonempty set of $P_{\mathcal{B}}$-measure zero in $\mathcal{B}_0$. Let $B \in \mathcal{A}$ be such that $1 > P(B) > 0$, let X be the characteristic function (identifier) of A, so that $X(\omega) = 1$, if $\omega \in A$, $X(\omega) = 0$ otherwise. Let $\mathcal{B}_1 = \{\emptyset, B, \Omega - B, \Omega\}$; obviously $\mathcal{B}_1$ is a (sub)-σ-field of $\mathcal{A}$. Then (2.16) is satisfied, if $(E^{\mathcal{B}_1} X)(\omega) = P(A/B) = P(A \cap B)/P(B)$ for each $\omega \in B$, and $(E^{\mathcal{B}_1} X)(\omega) = P(A/\Omega - B) = P(A - B)/(1 - P(B))$ otherwise, i.e., for all $\omega \in \Omega - B$. As can be easily proved, in this case for each $\omega \in \Omega$, $\omega_1 \in B$ and $\omega_2 \in \Omega - B$,

$$\begin{aligned} & E^{\mathcal{B}_0}(E^{\mathcal{B}_1} X)(\omega) = (E^{\mathcal{B}_1} X)(\omega_1) \, P(B) + (E^{\mathcal{B}_1} X)(\omega_2) \, (1 - P(B)) = \\ = \; & P(A/B) \, P(B) + P(A/\Omega - B) \, P(\Omega - B) = P(A) = (E^{\mathcal{B}_0} X)(\omega), \end{aligned} \tag{2.18}$$

as expected.

Some other notions, methods and results of probability theory will be introduced below in the case of their necessity and in the actual context when their necessity arises.

3 General Probabilistic Model of Decision Making under Uncertainty

Like the last chapter, also this one could be conceived at a purely formalized level, speaking about sets, mappings, functions, relations and ordered n-tuples of such objects satisfying some mathematically formalized demands. The difference between the two chapters consists in the fact that the intuition, interpretation and motivation behind the axiomatic probability theory can be found in most of the textbooks and monographs dealing with this theory, on the other side, in the case of general probabilistic and statistical models of decision making under uncertainty the situation is not so simple, let us mention here [30] or [2] as good introductory texts. Therefore we begin our explanation using informal terms charged by some extra-mathematical semantics, but our intention will be to get back to a formalized mathematical language as soon as possible.

Consider a system SYST of no matter which technical, medical, ecological or other nature. At the most general level such a system can be described by a black box such that a subject (agent, user, observer, ...) can put some data or values on the input device of this black box and some output values eventually occur as the reaction (answer) to, or as the impact of, the input values. E. g., in the case of a human being-patient, considered as a medical system, the input values are the applied medical treatments or medicaments, and the output values are various reactions of the patient's organism in particular or her/his global state of health in general. The exhaustive set of input values being at the subject's disposal will be denoted by D and its elements, denoted by d and perhaps indexed, will be called *decisions* as there is just this choice, more or less sophistically taken from the space D, in other words, the decision made by the subject, which enables to intervene into the system SYST with the aim to influence and, in a sense and from the subject's point of view, to optimize the output value by which the system answers to the chosen decision put on its input.

Just in the most trivial cases being, in our context, completely beyond the scope of our interests, the output of the system in question is determined by the input values themselves. To cover a more general case we shall suppose that the answers of the system may be different even if the input value is the same and we shall suppose, in other wording, we shall ontologically interprete this case in such a way, that it is caused by the fact that the system may be situated in various *internal states*. Denoting by S the set of all possible internal states of the system SYST we shall suppose that (i) SYST is situated in just one *actual (internal) state* $s \in S$ and (ii) the output value of the system is completely determined by the pair $\langle d, s \rangle$, where d is the decision taken by the subject and s is the actual state of the system at the time when d was put on the input device of SYST.

When looking for a formal description of the answers output by the system we shall limit ourselves to the case when each such particular output value turns into a numerically quantified gain or profit of the subject in question, e. g., a financial profit. In the case of rather qualitatively than quantitatively classified outputs we can ascribe the value 1 to the good, acceptable or adequate replies, and the value 0 or -1 to the other ones, the solution will always ultimately depend on the particular properties of SYST and the circumstances in question and it is almost useless to give some hints on the most general level adopted here. Formally said, we shall suppose that there is a real-valued function λ ascribing to each decision $d \in D$ and to each internal state $s \in S$ a real number $\lambda(d, s)$ (in symbols, $\lambda : D \times S \rightarrow R = (-\infty, \infty)$) taken as the numerical profit achieved by the subject if she/he applies the decision $d \in D$ and the system is situated in the internal state $s \in S$. Consequently, we can compare two or more decisions potentially applied in the same state s, saying that the decision d_2 is better than d_1 (not worse than d_1, resp.) in the state s, if the inequality $\lambda(d_1, s) < \lambda(d_2, s)$ ($\lambda(d_1, s) \leq \lambda(d_2, s)$, resp.) holds. Similarly, the decision d is the best (one among the best, resp.) in the internal state s of the system SYST, if the inequality $\lambda(d^*, s) < \lambda(d, s)$ ($\lambda(d^*, s) \leq \lambda(d, s)$, resp.) holds for each $d^* \in D,\ d^* \neq d$.

The problem how to choose the decision yielding the maximal profit would be very simple, leaving aside the computational problems, if the subject knew the actual state $s_0 \in S$ of the system SYST, however, as a rule, this is not the case. The only what the subject knows are the results of some observations, measurements, tests or experiments concerning the system and its environment, and only on the grounds of these empirical values the subject can choose a decision. For the sake of formal simplicity we can suppose that all such data being at the subject's disposal are described by a value x from a space E of all possible empirical values. Keeping in mind that E can be also a Cartesian product of some more elementary spaces or a functional space, we can easily see that the simplification just introduced does not bring a substantial loss of generality for our further reasonings. And now it is just the time when the phenomenon of uncertainty enters our model: it is possible, in general, that there are two different internal states s_1, s_2 of SYST such that $d_1 \in D$ is the optimal decision, if s_1 is the actual state of the system, $d_2 \in D,\ d_2 \neq d_1$, is optimal when s_2 is the case, but in both the cases $x \in E$ is observed, so that the subject is not able to choose the optimal decision having at her/his disposal nothing else than the empirical value $x \in E$. In symbols, if $\delta : E \rightarrow D$ is a mapping called *decision function* and ascribing the decision

$\delta(x) \in D$ to the observed empirical value $x \in E$, it is impossible to define δ in such a way that $\delta(x) = d_1$, if s_1 is the actual state of SYST and $\delta(x) = d_2 \neq d_1$, if s_2 is the actual state of SYST, as such a demand contradicts the elementary properties of δ taken as a function defined on E.

In order to describe this situation formally and, perhaps, to solve it somehow, let us apply the basic paradigma, already presented in the first chapter, according to which *uncertainty is nothing else than lack of complete knowledge in a deterministic model of the investigated system and its environment*. Hence, we shall suppose that there is a universal parameter ω, taking its values in a nonempty space Ω such that all the values concerning SYST and its environment, including the actual internal state $s_0 \in S$ of this system and the empirical value $x \in E$, are determined by the actual value $\omega \in \Omega$. This value can be understood as whole the history of Universe since the Big Bang, or as a complete description of positions and movements of all the bodies (all the mass particles, more precisely) in the Universe. Such an interpretation agrees with the idea that having at her/his disposal such an exhaustive description of Universe, the subject would be able to predict, without any risk of failure, all the future phenomena in the Universe in all their details, and it is just the lack of such an exhaustive information which brings uncertainty into our prediction and decision-making processes.

So, let $\sigma : \Omega \to S$ and $X : \Omega \to E$ be two total mappings such that, if $\omega \in \Omega$ is the actual value of the universal parameter or the actual elementary state of the Universe, then $\sigma(\omega) = s_0 \in S$ the actual internal state of the system SYST and $X(\omega) = x_0 \in E$ is the observed empirical value. The phenomenon of uncertainty introduced in the last paragraph can be formally described as follows. Let $opt : \Omega \to D$ be the mapping ascribing to each $\omega \in \Omega$ the best decision with respect to the actual internal state $\sigma(\omega)$ and to the profit function λ. Hence, for each $\omega \in \Omega$ and each $d \in D$ the inequality $\lambda(d, \sigma(\omega)) \leq \lambda(opt(\omega), \sigma(\omega))$ holds, for the sake of simplicity we put aside the problem with respect to which secondary criterion the value $opt(\omega)$ is defined, if there are two or more values $d \in D$ maximizing $\lambda(d, \sigma(\omega))$. However, it is possible that the subject observes a value $x_0 \in E$ such that there exist $\omega_1, \omega_2 \in \Omega$, $\omega_1 \neq \omega_2$, with the property that $X(\omega_1) = X(\omega_2) = x_0$, but $opt(\omega_1) \neq opt(\omega_2)$ (it follows immediately that in this case $\sigma(\omega_1) \neq \sigma(\omega_2)$). How to decide, now, whether to take the decision d_1 or d_2, even when supposing, for the sake of simplicity, that for each $\omega \in \Omega$ such that $X(\omega) = x_0$ either $opt(\omega) = d_1$ or $opt(\omega) = d_2$ hold? The occurred phenomenon of uncertainty is reduced to the subject's lack of information concerning the precise actual value of the parameter ω, the only what the subject knows is that $X(\omega) = x_0$, in other symbols, that $\omega \in \{\omega_1 \in \Omega : X(\omega_1) = x_0\}$, but this information does not enable to identify fully value ω.

A possible solution to this decision problem could be as follows. Consider the subsets $A_1 = \{\omega \in \Omega : X(\omega) = x_0, opt(\omega) = d_1\}$ and $A_2 = \{\omega \in \Omega : X(\omega) = x_0, opt(\omega) = d_2\}$ and compare them from the point of view of their sizes, importance or weights of their elements, or according to another numerically quantified criterion. Consequently, take the decision d_1, if the set A_1 is preferred to A_2 with respect to this criterion, take d_2, if A_2 is preferred to A_1 with respect to the same criterion, and apply some auxiliary criterion, if the primary criterion ascribes the same value to both the sets A_1, A_2.

The reader probably already guesses, and we must admit that it is a good intuition, that it will be a probability measure, defined on an appropriate system of subsets of Ω, which will play the role of a general enough numerical characteristic of the size or weight of the sets A_1 and A_2 supposing that they both belong to the system of subsets of Ω in question. As a matter of fact we shall suppose, since now, that there is a probability space $\langle \Omega, \mathcal{A}, P \rangle$ defined over the space Ω of all possible values of the universal parameter ω and being at the subject's (user's, decision maker's, ...) disposal. The case when P is replaced by some more general measure, e. g., by a signed measure, will be investigated in some of the next chapters. So, the user accepts the decision d_1 if $P(A_1) \geq P(A_2)$, and she/he accepts d_2, if $P(A_1) < P(A_2)$, of course, the case when $P(A_1) = P(A_2)$ can be treated also in the opposite way. If the set $\{\omega \in \Omega : X(\omega) = x_0\} = A_3$ belongs to $\mathcal{A}$ and $P(A_3)$ is positive, then $P(A_1) \geq P(A_2)$ holds iff the inequality

$$\begin{aligned} & P(A_1)/P(A_3) = P(A_1 \cap A_3)/P(A_3) = P(A_1/A_3) = \\ = \; & P\left(\{\omega \in \Omega : opt(\omega) = d_1\}/\{\omega \in \Omega : X(\omega) = x_0\}\right) \geq \\ \geq \; & P(A_2)/P(A_3) = P(A_2 \cap A_3)/P(A_3) = P(A_2/A_3) = \\ = \; & P\left(\{\omega \in \Omega : opt(\omega) = d_2\}/\{\omega \in \Omega : X(\omega) = x_0\}\right) \end{aligned} \tag{3.1}$$

holds, so that the decision function can be defined in the intuitively more easy to understand terms of conditional probabilities.

However, the way of reasoning leading to the solution just introduced is not the most general one, as it does not reflect the different situations which can occur, if a wrong (i. e., not the optimal) decision is taken. It matters, whether the difference $|\lambda(d_1, \sigma(\omega)) - \lambda(d_2, \sigma(\omega))|$ is more or less negligible or whether it is, in fact, a qualitative difference (life or death for a patient or a prisoner) only very poorly described in quantitative terms. A more general way of reasoning leads to the following model.

The optimal decision function $\delta_{opt} : E \to D$ would be such that $\delta_{opt}(X(\omega)) = opt(\omega)$ for each $\omega \in \Omega$, as in this case the inequality

$$\lambda\left(\delta_{opt}(X(\omega)), \sigma(\omega)\right) \geq \lambda(\delta(X(\omega)), \sigma(\omega)) \tag{3.2}$$

would hold for each decision function $\delta : E \to D$. In general, however, δ_{opt} need not exist, as the partition of Ω generated by the system of subsets $\{\{\omega \in \Omega : X(\omega) = x\} : x \in E\}$ may be too rough to enable the definition of δ_{opt}. In the extremal case, when $X(\omega) = x_0$ for each $\omega \in \Omega$, only constant decision functions can be applied to $X(\omega)$. However, the value

$$\alpha(\delta, \sigma)(\omega) = |\lambda(opt(\omega), \sigma(\omega)) - \lambda(\delta(X(\omega)), \sigma(\omega))| \geq 0 \tag{3.3}$$

can be taken as (and will be called) the loss suffered by the subject applying the decision function δ, if ω is the actual value of the universal parameter, and an intuitively reasonable subject's effort will be to minimize this loss to the degree as small as possible.

It is very easy to see that up to trivial cases it does not exist a decision function $\delta_0 : E \to D$ such that $\alpha(\delta_0, \sigma)(\omega) \leq \alpha(\delta, \sigma)(\omega)$ would hold uniformly for all $\omega \in \Omega$. Or, for a fixed $\omega_0 \in \Omega$ and for the constant decision function $\delta_{\omega_0}(x) \equiv opt(\omega_0)$ for all $x \in E$, the value

$$\alpha(\delta_{\omega_0}, \sigma)(\omega_0) = |\lambda(opt(\omega_0), \sigma(\omega_0)) - \lambda(\delta_{\omega_0}(X(\omega_0)), \sigma(\omega_0))| = 0 \tag{3.4}$$

is minimal, however, $\alpha(\delta_{\omega_0}, \sigma)(\omega)$ can be very high for $\omega \neq \omega_0$. (It does not matter, in this context, that the subject is perhaps unable to compute effectively the value $opt(\omega_0)$ and to define effectively, *which* is the decision function δ_{ω_0} in question). The two most frequently used solutions here are the *minimax* and the *Bayes* ones.

The minimax solution is based on the "pessimistic" or "safety first" principle so that we take the value

$$\rho(\delta, \sigma) = \sup\{\alpha(\delta, \sigma)(\omega) : \omega \in \Omega\} \tag{3.5}$$

as the numerical characteristic of the quality of the decision function δ and the subject's aim is to choose δ_0 in such a way that $\rho(\delta_0, \sigma) \leq \rho(\delta, \sigma)$ holds for each decision function δ. It is possible, in general, that such δ_0 does not exists, in other words, there exists a sequence $\delta_1, \delta_2, \ldots$ of decision function such that $\rho(\delta_1, \sigma) \geq \rho(\delta_2, \sigma) \geq \ldots$ holds, but there is no δ_0 such that $\rho(\delta_0, \sigma) = \lim_{i\to\infty} \rho(\delta_i, \sigma)$. But, in this case there exists, for each $\varepsilon > 0$, a decision function $\delta_{0,\varepsilon}$ such that

$$\rho(\delta_{0,\varepsilon}, \sigma) \leq \inf\{\rho(\delta, \sigma) : \delta \in \mathcal{D}\} + \varepsilon \tag{3.6}$$

holds, where $\mathcal{D} = D^E$ is the set of all mappings from E to D.

The often used argument against the minimax principle takes this principle as too pessimistic in the sense that the choice of the decision function δ_0 may be ultimately influenced by the behaviour of the loss function $\alpha(\delta, \sigma)(\omega)$ for a singular value of ω, or for ω's belonging to a subset of non-typical, degenerated, very rarely occurring, useless from the practical point of view, etc., values of the universal parameter. Accepting this argument we have to classify particular decision functions rather with respect to the expected value of the loss function $\alpha(\delta, \sigma)(\omega)$ defined with respect to some apriori probability distribution P defined on a σ-field $\mathcal{A}$ of subsets of Ω. Hence, leaving aside, for the moment, the problem connected with the existence of the integrals in question, we choose such a decision function $\delta_0 : E \to D$ that the inequality

$$\begin{aligned} E_P(\alpha(\delta_0, \sigma)) &= \int_\Omega \alpha(\delta_0, \sigma)(\omega)\,\mathrm{d}P \leq \int_\Omega \alpha(\delta, \sigma)(\omega)\,\mathrm{d}P = \\ &= E_P(\alpha(\delta, \sigma)) \end{aligned} \tag{3.7}$$

holds for each $\delta \in \mathcal{D} = D^E$. Again, due to the properties of the infimum operation in the space of non-negative real numbers there always exists, for each $\varepsilon > 0$, a decision function $\delta_{0,\varepsilon} \in \mathcal{D}$ such that

$$E_P(\alpha(\delta_{0,\varepsilon}, \sigma)) \leq \inf\{E_P(\alpha(\delta, \sigma)) : \delta \in \mathcal{D}\} + \varepsilon \tag{3.8}$$

holds. This is the so called *Bayes solution* corresponding to (or: defined by) the a priori probability measure P on the σ-field $\mathcal{A}$.

As it is not the aim of this work in general and this chapter in particular to investigate decision problems under uncertainty and statistical decision functions in more detail at such a general and abstract level as presented above, let us describe, in more detail, the particular case when

(i) $D = S$, hence, the space of decisions is identical with that of possible internal states of the investigated system SYST; below we shall investigate also the case when $D = \{A, S - A\}$ for a subset $A \subset S$ ($\emptyset \neq A \neq S$, to avoid trivialities), and

(ii) $\lambda(d, s) = 1$, if $d = s$, $\lambda(d, s) = 0$ otherwise.

Hence, the subject's goal is to identify the actual internal state $s_0 = \sigma(\omega)$ of SYST and if she/he identifies the state correctly, she/he obtains a unit profit (suffers no or zero loss, in the dual wording). If the identification is wrong, she/he does not obtain any profit (suffers a unit loss). Let $\langle \Omega, \mathcal{A}, P \rangle$ be a probability space over Ω, let us suppose, for the sake of simplicity, that both the spaces D $(= S)$ and E are finite or countable and that X (σ, resp.) is a random variable taking $\langle \Omega, \mathcal{A}, P \rangle$ into the measurable space $\langle E, \mathcal{P}(E) \rangle$ ($\langle S, \mathcal{P}(S) \rangle$, resp.). Let us try to find the Bayes solution to the decision problem just defined, i. e., let us try to find a decision function $\delta_0 : E \to D$ $(= S)$ satisfying (3.7) for the given random variables X and σ.

We can suppose, without any loss of generality in the case when E is finite or countable, that $P(\{\omega \in \Omega : X(\omega) = x\}) > 0$ holds for each $x \in E$ (we can reduce E to such elements, if it is not a priori the case). Consequently, for mappings $\delta : E \to D$ $(= S)$ and $\sigma : \Omega \to S$ we obtain that

$$\begin{aligned} & E_P(\alpha(\delta, \sigma)) = \qquad (3.9) \\ = \ & \textstyle\sum_{x \in E} [1 \cdot P(\{\omega \in \Omega : \alpha(\delta, \sigma)(\omega) = 1,\ X(\omega) = x\}) + \\ & \quad + 0 \cdot P(\{\omega \in \Omega : \alpha(\delta, \sigma)(\omega) = 0,\ X(\omega) = x\})] = \\ = \ & \textstyle\sum_{x \in E} [P(\{\omega \in \Omega : \alpha(\delta, \sigma)(\omega) = 1\} / \{\omega \in \Omega : X(\omega) = x\}) \cdot \\ & \quad \cdot P(\{\omega \in \Omega : X(\omega) = x\})], \end{aligned}$$

as $\alpha(\delta, \sigma)(\omega)$ takes only the values 0 or 1.

Let $\delta_0 : E \to S$ be defined in this way. For each $x \in E$, $\delta_0(x) = s_x \in S$ iff

$$\begin{aligned} & P(\{\omega \in \Omega : \sigma(\omega) = s_x\} / \{\omega \in \Omega : X(\omega) = x\}) \geq \qquad (3.10) \\ \geq \ & P(\{\omega \in \Omega : \sigma(\omega) = s\} / \{\omega \in \Omega : X(\omega) = x\}) \end{aligned}$$

holds for each $s \in S$, in other terms, s_x is (the or a) value maximizing the conditional probability $P(\{\omega \in \Omega : \sigma(\omega) = s\} / \{\omega \in \Omega : X(\omega) = x\})$ taken as a function of s. Now, for each decision function $\delta : E \to D$ $(= S)$, and for each $x \in E$, due to (3.10) the following equalities and inequalities hold.

$$\begin{aligned} & P\left(\{\omega \in \Omega : \alpha(\delta, \sigma)(\omega) = 1\} / \{\omega \in \Omega : X(\omega) = x\}\right) = \qquad (3.11) \\ = \ & P\left(\{\omega \in \Omega : \delta(X(\omega)) \neq \sigma(\omega)\} / \{\omega \in \Omega : X(\omega) = x\}\right) = \\ = \ & P\left(\{\omega \in \Omega : \delta(x) \neq \sigma(\omega)\} / \{\omega \in \Omega : X(\omega) = x\}\right) \geq \\ \geq \ & P\left(\{\omega \in \Omega : s_x \neq \sigma(\omega)\} / \{\omega \in \Omega : X(\omega) = x\}\right) = \\ = \ & P\left(\{\omega \in \Omega : \delta_0(X(\omega)) \neq \sigma(\omega)\} / \{\omega \in \Omega : X(\omega) = x\}\right) = \\ = \ & P\left(\{\omega \in \Omega : \alpha(\delta_0, \sigma) = 1\} / \{\omega \in \Omega : X(\omega) = x\}\right). \end{aligned}$$

Hence, combining (3.9) and (3.11) we obtain that δ_0 minimizes $E_P(\alpha(\delta, \sigma))$, so that δ_0 is a Bayes solution to the decision problem in question. This result is quite intuitive:

obtaining the empirical value $x \in E$, the subject takes as her/his estimation of the actual internal state of SYST this value $s \in S$, which is the most probable under the condition that x was observed. If there are two or more values from S with the same conditional probabilities, it is a matter of a secondary criterion not so important in our context, which of the possible candidates will be chosen.

When trying to apply the Bayes decision function δ_0 just defined in a practical case we arrive at the following problem: how to obtain the values, or at least good and reliable estimations, of the conditional probabilities $P(\{\omega \in \Omega : \sigma(\omega) = s\}/\{\omega \in \Omega : X(\omega) = x\})$ for $s \in S$ and $x \in E$? The problem is that the dependence between internal states of SYST and the empirical results, even if symmetric from the mathematical point of view, is more intuitively seen as going rather from the actual internal state of the system to the empirical value x than in the opposite sense. E. g., it is quite natural to ask, and not so difficult to compute, which is the probability that just six heads occurs in a sequence of ten statistically independent and equally distributed coin tosses (empirical observation "$X(\omega) = 6$") under the condition that the probability with which head occurs in each toss is $1/2$ (i. e., supposing that $\sigma(\omega) = s_0 = 1/2$), than to ask for the probability that $\sigma(\omega) = 1/2$ under the condition that $X(\omega) = 6$. The well-known Bayes formula (cf. [8, 11] or any elementary textbook on probability theory) reads that

$$P\left(\{\omega \in \Omega : \sigma(\omega) = s\}/\{\omega \in \Omega : X(\omega) = x\}\right) = \tag{3.12}$$

$$\frac{P\left(\{\omega \in \Omega : \sigma(\omega) = s,\, X(\omega) = x\}\right)}{P\left(\{\omega \in \Omega : X(\omega) = x\}\right)}$$

$$\frac{P\left(\{\omega \in \Omega : X(\omega) = x\}/\{\omega \in \Omega : \sigma(\omega) = s\}\right) P\left(\{\omega \in \Omega : \sigma(\omega) = s\}\right)}{\sum_{s \in S} P\left(\{\omega \in \Omega : X(\omega) = x\}/\{\omega \in \Omega : \sigma(\omega) = s\}\right) P\left(\{\omega \in \Omega : \sigma(\omega) = s\}\right)}.$$

Let us recall that the sets S and E are supposed to be finite or countable for the sake of simplicity, more general versions of Bayes formula can be found in the textbooks mentioned above or elsewhere.

Hence, we have escaped from the problem to have at our disposal directly the values or good estimates of conditional probabilities $P(\{\omega \in \Omega : \sigma(\omega) = s\}/\{\omega \in \Omega : X(\omega) = x\})$, but only supposing that we know, besides the probabilities $P(\{\omega \in \Omega : X(\omega) = x\}/\{\omega \in \Omega : \sigma(\omega) = s\})$, also the values of the apriori probabilistic distribution $P(\{\omega \in \Omega : \sigma(\omega) = s\})$ for each $s \in S$. And it is one of the basic problems of the Bayes approach to statistical decision making to obtain this apriori distribution or even to justify the point of view that it is reasonable to assume that the actual state of the system is a random value. E. g., when estimating the probability of life on the surface of Mart given some indirect empirical indices obtained by cosmic sonds and using the Bayes rule, we have to know the apriori probability that the life on Mart exists. However, the life on Mart either exists or not, it is an individual and isolated phenomenon in the Universe, but in order to interprete the probability of life on Mart somehow, we have to suppose that there exist a great number of planets with conditions identical as those on the surface of Mart, and we have to define the apriori probability of life on the surface of Mart by the relative frequence of those "identical copies of Mart" where the life exists. Such a construction seems to be rather artificial and counter-intuitive. And still another problem remains: even if the existence of an apriori distribution can be justified somehow, how to obtain the particular values ascribed by this distribution to various internal states of the system SYST? Happy enough, it is not our aim in this work to discuss this problem in more detail, let us just remark, that there exist a lot of papers, monographs, conference proceedings volumes, etc. dealing with this problem, and many related ones, on various levels and from various points of view. What is important in our context, one of the sources of inspiration for the Dempster–Shafer theory, or rather the Dempster–Shafer approach to (or: model of) uncertainty quantification and processing, consists just in the fact that this approach enables to avoid from our considerations the problem of existence and identification of apriori probability distribution, as will be shown in the next chapter.

4 Basic Elements of Dempster–Shafer Theory

The greatest part of works dealing with the fundamentals of Dempster–Shafer theory is conceived either on the combinatoric, or on the axiomatic, but in both the cases on a very abstract level. The first approach begins by the assumption that S is a nonempty finite set, that m is a mapping which ascribes to each $A \subset S$ a real number $m(A)$ from the unit interval $\langle 0, 1\rangle$ in such a way that $\sum_{A \subset S} m(A) = 1$ (m is called a *basic probability assignment* on S), and that the *(normalized) belief function* induced by m is the mapping $bel_m : \mathcal{P}(S) \rightarrow \langle 0, 1\rangle$ defined, for each $A \subset S$, by $bel_m(A) = (1 - m(\emptyset))^{-1} \sum_{\emptyset \neq B \subset A} m(B)$, if $m(\emptyset) < 1$, bel_m being undefined otherwise ([35]). The other (axiomatic) approach begins with the idea that belief function on a finite nonempty set S is a mapping $bel : \mathcal{P}(S) \rightarrow \langle 0, 1\rangle$, satisfying certain conditions (obeying certain axioms, in other terms). If these conditions (axioms) are strong and reasonable enough, it can be proved that it is possible to define uniquely a basic probability assignment m on S such that the belief function induced by m is identical with the original belief function defined by axioms, so that both the approaches meet each other and yield the same notion of belief function ([39]). The problems how to understand and obtain the probability distribution m over $\mathcal{P}(S)$ in the first case, or how to justify the particular choice of the demands imposed to belief functions in the other case, are put aside or are "picked before brackets" and they are not taken as a part of Dempster–Shafer theory in its formalized setting.

The basic stone of the probabilistic approach to Dempster–Shafer theory consists in a definition and interpretation of belief functions, as the basic quantitative characteristic of uncertainty in this theory, using appropriate terms and tools of probability theory. Like as in the more general case introduced above, we shall begin with some intuitive interpretation of the presented notions, putting this interpretation aside and returning to a purely mathematical formalized style of explanation as quickly as possible.

So, let S be a nonempty, but not necessary finite, set of all possible internal states of an investigated system SYST. As in the particular case of decision making under uncertainty explained in the closing part of Chapter 3, our aim is not to optimize a general statistical decision function with respect to a given loss function and to a global decision strategy (the minimax, a Bayes, or another one), but rather to identify the actual internal state s_0 of the system SYST or at least to decide, whether $s_0 \in A$ holds or does not hold for a (proper, to avoid the trivial case) subset A of S. The hidden assumption behind such a simplification is that if the decision about the internal state of SYST is correct, also the consecutive activity of the subject concerning in her/his intervention into the system will be the best possible. More generally, the better is the decision about the actual internal state of SYST, the better will be the consecutive operation executed by the subject.

Again, as in the general case above, the subject is not supposed to be able to observe immediately the actual state of SYST or to draw this information simply and beyond any risk of error from her/his knowledge concerning the system and its environment. The only what the subject knows are the results of some observations, measurements, experiments, etc., cumulated into a value x from a nonempty (and possibly vector) space E. At the general level the finiteness of E also need not be assumed. However, in order to be sure that there is at least some degree of sensefulness and rationality when taking some decisions concerning the actual state of SYST on the ground of empirical values from E, some relation between the states from S and values from E must be assumed to exist and to be known to the subject. In the case of statistical decision functions such relations are given by the conditional probabilities $P(\{\omega \in \Omega : X(\omega) = x\}/\{\omega \in \Omega : \sigma(\omega) = s\})$ for $x \in X$ and $s \in S$ in the case when the spaces E and S are at most countable, or by the conditional probabilities $P(\{\omega \in \Omega : X(\omega) \in F\}/\{\omega \in \Omega : \sigma(\omega) \in T\})$ for at least some subsets $F \subset E$ and $T \subset S$ in the general case. If the Bayes approach is to be applied also the apriori probability distribution $P(\{\omega \in \Omega : \sigma(\omega) = s\})$ or $P(\{\omega \in \Omega : \sigma(\omega) \in T\})$ must be known. When developing the Dempster–Shafer theory, such a basic relation between states and observations is defined by the so called compatibility relation.

Definition 4.1. *Compatibility relation* over a state space S and an observational space E is a subset of the Cartesian product $S \times E$ or, what obviously turns to be the same, a binary function $\rho : S \times E \rightarrow \{0, 1\}$, i.e., for each $s \in S$ and $x \in E$, either $\rho(s, x) = 1$ or $\rho(s, x) = 0$. □

The intuition behind this definition is as follows. The case $\rho(s, x) = 0$ denotes, for a particular state $s \in S$ and a particular empirical value $x \in E$, that the subject knows, or is able to deduce, using her/his knowledge and within the frameworks of her/his deductive abilities, that s *cannot be* the actual internal state of the system under consideration supposing that the empirical value x was obtained. The state s and the value x are then called *(mutually) incompatible*. E.g., a doctor can eliminate certain diagnoses on the ground of the medical data obtained during an examination of a patient, even if this doctor is still not able to say exactly, which is the true diagnosis. Consequently, the case when $\rho(s, x) = 1$ describes the situation when the subject is not able to avoid the possibility that s is the actual internal state of SYST when x was observed; s and x are then called *(mutually) compatible*. Two assumptions are imposed to this interpretation of compatibility relation, namely:

(i) If $\rho(s,x)=0$, then it is taken as granted and objectively valid, that s and x are incompatible, hence, this case describes the objectively true state of affairs, no matter of the obvious fact that subject's knowledge and deductive abilities must be limited. This condition will be abandoned, in what follows, only locally, when introducing and discussing the so called dual (or: pessimistic) Dempster combination rule.

(ii) On the other side, if $\rho(s,x)=1$, it is possible that, according to some laws of nature or other rules governing the system and its environment, s and x are incompatible, but the subject does not know about this fact because of her/his limited knowledge base and deductive abilities. E. g., a young doctor lacking a sufficient experience is not able to avoid a diagnosis which her/his older colleague eliminates almost immediately having seen the results of the patient's examinations. This condition plays an important role in the Dempster combination rule (cf. the next chapter), as this rule enables to improve one subject's knowledge (i. e., to enlarge the set of pairs $\langle s,x\rangle$ known by her/him to be incompatible) by sharing knowledge with another subject-specialist in the field of discourse.

Given an empirical value $x\in E$, Definition 4.1 enables to define the set $U_\rho(x)=\{s\in S:\rho(s,x)=1\}$ of states compatible with this empirical value. The phenomenon of uncertainty will be embedded into our model when supposing that empirical values are charged by uncertainties which can be defined, quantified and processes using the tools of classical Kolmogorov axiomatic probability theory. In other terms, we shall suppose that $x\in E$ is the value taken by a (generalized) random variable X defined on an abstract probability space $\langle\Omega,\mathcal{A},P\rangle$ and taking its values in a measurable space $\langle E,\mathcal{E}\rangle$ generated over the observational space E when choosing a σ-field $\mathcal{E}$ of subsets of E. Now, the composed mapping $U_\rho(X(\cdot))$ takes the space Ω into the power-set $\mathcal{P}(S)$ of all subsets of the space S and we may ask, given $\omega\in\Omega$ and $A\subset S$, whether the inclusion $U_\rho(X(\omega))\subset A$ holds or does not hold. If $\{\omega\in\Omega:U_\rho(X(\omega))\subset A\}$ belongs to the σ-field $\mathcal{A}$ of subsets of Ω, we may quantify the size of this set using the probability measure P and we can define the value $P(\{\omega\in\Omega:U_\rho(X(\omega))\subset A\})$. If this is the case also for the empty set $\emptyset\subset S$, i. e., if $\{\omega\in\Omega:U_\rho(X(\omega))=\emptyset\}\in\mathcal{A}$ holds, we may define the value

$$bel^*_\rho(A)=P\left(\{\omega\in\Omega:\emptyset\neq U_\rho(\omega)\subset A\}\right) \tag{4.1}$$

and we can call it the *non-normalized degree of belief* (or: *the value of non-normalized belief function*) defined by the compatibility relation ρ and random variable X, ascribed to the subset A of S ([38]). If, moreover, $bel^*_\rho(\emptyset)<1$ holds, then the *normalized degree of belief* (or: *the value of normalized belief function*) defined by the compatibility relation ρ and random variable X, ascribed to the subset A of S, is defined by the conditional probability

$$bel_\rho(A)=P\left(\{\omega\in\Omega:U_\rho(X(\omega))\subset A\}/\{\omega\in\Omega:U_\rho(X(\omega))\neq\emptyset\}\right). \tag{4.2}$$

A slightly more formalized definition reads as follows.

Definition 4.2. Let S be a nonempty state space, let E be a nonempty observational space, let $\mathcal{E} \subset \mathcal{P}(E)$ be a σ-field of subsets of E, let $\langle \Omega, \mathcal{A}, P \rangle$ be a probability space, let $X : \langle \Omega, \mathcal{A}, P \rangle \to \langle E, \mathcal{E} \rangle$ be a (generalized) random variable, let $\rho \subset S \times E$ be a compatibility relation. Let $U_{\rho,X} : \Omega \to \mathcal{P}(S)$ be a mapping defined, for each $\omega \in \Omega$, by

$$U_{\rho,X}(\omega) = \{s \in S : \rho(s,\, X(\omega)) = 1\}, \tag{4.3}$$

let $\mathcal{S} \subset \mathcal{P}(\mathcal{P}(S))$ be a σ-field of systems of subsets of S such that $U_{\rho,X}$ is a (generalized set-valued) random variable taking $\langle \Omega, \mathcal{A}, P \rangle$ into $\langle \mathcal{P}(S), \mathcal{S} \rangle$ (such $\mathcal{S}$ always exists, at least $\mathcal{S} = \{\emptyset, \mathcal{P}(S)\}$ will do). Then *non-normalized degree of belief (belief function)* $bel^*_{\rho,X}$ is the (partial, in general) mapping which takes $\mathcal{P}(S)$ into $\langle 0, 1 \rangle$ in such a way that

$$bel^*_{\rho,X}(A) = P\left(\{\omega \in \Omega : \emptyset \neq U_{\rho,X}(\omega) \subset A\}\right), \tag{4.4}$$

if $A \subset S$, $\{\emptyset\} \in \mathcal{S}$ and $\mathcal{P}(A) \in \mathcal{S}$, $bel^*_{\rho,X}(A)$ being undefined otherwise. If $bel^*_{\rho,X}(\emptyset) < 1$ holds for the empty subset $\emptyset$ of S, then *normalized degree of belief (belief function)* $bel_{\rho,X}$ is the (partial, in general) mapping which takes $\mathcal{P}(S)$ into $\langle 0, 1 \rangle$ in such a way that

$$bel_\rho(A) = P\left(\{\omega \in \Omega : U_{\rho,X}(\omega) \subset A\} / \{\omega \in \Omega : U_{\rho,X}(\omega) \neq \emptyset\}\right), \tag{4.5}$$

if $A \subset S$ and $\mathcal{P}(A) \in \mathcal{S}$, $bel_{\rho,X}(A)$ being undefined otherwise. If $bel^*_{\rho,X}(\emptyset) = 1$, the normalized degree of belief (belief function) $bel_{\rho,X}$ is undefined. □

Belief functions, just defined, are the basic numerical quantifications or characteristics of uncertainty in the Dempster–Shafer theory and it is why their definition needs several more detailed comments and remarks to which the greatest part of the rest of this chapter will be devoted.

According to Definition 4.2, belief functions bel^* and bel depend on the compatibility relation ρ, hence, also on the spaces S and E, on the probability space $\langle \Omega, \mathcal{A}, P \rangle$, and on the random variable X. In symbols, $bel^* = bel^*_{S,E,\rho,\langle \Omega,\mathcal{A},P \rangle,X}$, and similarly for the normalized version. However, in what follows, only ρ and X will be, occasionally, introduced explicitly, namely in the cases when belief functions induced by different compatibility relations ρ_1, ρ_2, say, and/or by different random variables, X_1, X_2, say, will be considered and perhaps combined with each other, as it will be the case of the Dempster combination rule, introduced and investigated in the next chapter. All the other parameters will be either taken as fixed (e. g., the probability space $\langle \Omega, \mathcal{A}, P \rangle$), or assumed to be clear from the context. The author believes that the resulting simplification in the used notation is worth accepting the perhaps possible risk of a misunderstanding.

A very important property of Definition 4.2 consists in its compatibility with the usual combinatoric definition of belief function for finite spaces S, briefly mentioned at the very beginning of this chapter. In other words, Definition 4.2 generalizes this elementary definition in a natural conservative way. Let the space S and, consequently, also the power-set $\mathcal{P}(S)$, be finite, let $\mathcal{S} = \mathcal{P}(\mathcal{P}(S))$ be the maximal (the finest) σ-field of subsets of $\mathcal{P}(S)$, let the random variable $X : \langle \Omega, \mathcal{A}, P \rangle \to \langle E, \mathcal{E} \rangle$ be such that the composed mapping $U_\rho(X(\cdot)) : \Omega \to \mathcal{P}(S)$ is a random variable defined on the probability space $\langle \Omega, \mathcal{A}, P \rangle$ and taking its values in the measurable space $\langle \mathcal{P}(S),\, \mathcal{P}(\mathcal{P}(S)) \rangle$.

Such a measurability of $U_\rho(X(\cdot))$ can be easily achieved, e. g. when also the observational space E is finite, $\mathcal{E} = \mathcal{P}(E)$ is the σ-field of all subsets of E, and X is a random variable taking $\langle \Omega, \mathcal{A}, P \rangle$ into $\langle E, \mathcal{P}(E) \rangle$. Then we can define and denote, for each $A \subset S$,

$$m(A) = P\left(\{\omega \in \Omega : U_\rho(X(\omega)) = A\}\right). \tag{4.6}$$

It is obvious that m is a probability distribution over $\mathcal{P}(S)$, in other words, a basic probability assignment on S, as $\sum_{A \subset S} m(A) = 1$. Now,

$$bel_m^*(A) = P\left(\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset A\}\right) = \sum_{\emptyset \neq B \subset A} m(B), \tag{4.7}$$

and

$$\begin{aligned} bel_m(A) &= P\left(\{\omega \in \Omega : U_\rho(X(\omega)) \subset A\} / \{\omega \in \Omega : U_\rho(X(\omega)) \neq \emptyset\}\right) = \\ &= \frac{P\left(\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset A\}\right)}{P\left(\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega))\}\right)} = \\ &= (1 - m(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A} m(B), \end{aligned} \tag{4.8}$$

if $m(\emptyset) < 1$ holds, according to the combinatoric definition. Obviously, in this special case the values $bel_m^*(A)$ (and $bel(A)$, if $m(\emptyset) < 1$) are defined for each $A \subset S$.

In our reasonings above we have not avoided the case when $U_\rho(X(\omega)) = \emptyset$ or even when $P(\{\omega \in \Omega : U_\rho(X(\omega)) = \emptyset\})$ is positive, only the case when this probability equals one has been eliminated from our considerations when defining the normalized belief function. On the other side, it is clear that when $U_\rho(X(\omega)) \neq \emptyset$ holds for each $\omega \in \Omega$ or, in a weaker setting, if $P(\{\omega \in \Omega : U_\rho(X(\omega))\}) = 0$, then $bel_\rho^*(A) = bel_\rho(A)$ for each $A \subset S$ for which $bel^*(A)$ is defined. We shall write, in what follows, $m_1 \approx m_2$ ($bel_{m_1} \approx bel_{m_2}$, resp.), if $m_1(A) = m_2(A)$ ($bel_{m_1}(A) = bel_{m_2}(A)$, resp.) holds for each $A \subset S$ for which the value in question are defined. The case $U_\rho(X(\omega)) = \emptyset$ occurs, if there is *no* state $s \in S$ compatible with the observed value $X(\omega)$. Such a situation seems to be contradictory at the first sight, as all empirical results should be compatible at least with the actual internal state of the system SYST and SYST is supposed to be situated in just one internal state $s_0 \in S$. An explication or interpretation of this phenomenon can be twofold:

(i) Our assumption that S contains *all* the possible internal states of SYST is true (the so called *closed world assumption*), but $X(\omega)$ is not immediately the value of the observation(s) and/or experiment(s) in question, but the value obtained through a communication channel in which the original empirical value could have been subjected to a deformation. E. g., if $X(\omega) = \langle x_1, x_2 \rangle$, where x_1 says "there is a snow covering all the countryside" and x_2 says "there is a temperature $+20°$ C in the open air", so that $X(\omega)$ is incompatible with any state of affairs supposing that the compatibility relation describes the usual meteorological laws, it is possible that the original version of x_2 was "there is a temperature $-20°$ C in the open air", the sign "$-$" being lost during the transcription or other communication of this message. The result would be, of course, that there is no state of nature compatible with $\langle x_1, x_2 \rangle$, at least under the usual physical conditions.

(ii) Another possible explication reads as follows. The obtained empirical data are correct, but our assumption that the actual internal state s_0 of the system under consideration must be in S was wrong (the so called *open world assumption*). Under this interpretation the seemingly inconsistent data say to the subject, that the actual internal state of SYST is beyond the space S, so that her/his assumption about the exhaustive character of S has proved to fail. Considering the example just introduced, the simultaneous observation of snow outside the windows and the thermometer showing $+20°$ C imply, that the atmospheric pressure outside is much more higher than the normal one, even if this possibility has not been (wrongly) taken into consideration when defining the space S. It is obvious, but perhaps important to say explicitly, that it is impossible to decide, using only the mathematical apparatus being at the subject's disposal, which of the two possibilities (inconsistent deformation of data vs. an a priori unconsidered state of the system) took place in the real world.

In order to discuss the sense of the random event $\emptyset \neq U_\rho(X(\omega)) \subset A$, the probability of which is quantified by the belief functions bel^* and bel, let us adopt the basic idea of Bayes decision making under uncertainty and suppose that the actual internal state s_0 of SYST is defined as the value of a random variable σ, which takes the probability space $\langle \Omega, \mathcal{A}, P \rangle$ into a measurable space $\langle S, \mathcal{S}^* \rangle$. Here $\mathcal{S}^*$ is a σ-field of subsets of S. Suppose that the following *condition of semantical correctness* is satisfied, namely, that $\rho(\sigma(\omega),\, X(\omega)) = 1$ holds for each $\omega \in \Omega$. Hence, we suppose that each empirical value is compatible with the actual internal state of the system by which this value has been generated. Consequently, $\sigma(\omega) \in U_\rho(X(\omega))$ and $U_\rho(X(\omega)) \neq \emptyset$ hold for each $\omega \in \Omega$. Hence, for each $\omega \in \Omega$ and $A \subset S$, $U_\rho(X(\omega)) \subset A$ implies that $\sigma(\omega) \in A$ (but not vice versa, in general, so that the inequality

$$\begin{aligned} bel^*(A) = bel(A) &= P\left(\{\omega \in \Omega : \emptyset \neq U_\rho(\omega) \subset A\}\right) \leq \\ &\leq P\left(\{\omega \in \Omega : \sigma(\omega) \in A\}\right) \end{aligned} \tag{4.9}$$

hold for each $A \subset S$ for which the values in question are defined. Only under this condition of semantical correctness it is possible to take the value $bel(A)$ as a probabilistically reasonable characteristic of the random event occurring when the actual state of the system under investigation belongs to the subset A of S, namely, $bel(A)$ is a lower bound of the a priori probability that $\sigma(\omega) \in A$ holds. On the other side, if the condition of semantical correctness does not hold, then it is possible that the actual state s_0 of SYST is not in $U_\rho(X(\omega))$ (in a particular case it may trivially follow from the fact that $U_\rho(X(\omega))$ is empty), so that from the inclusion $U_\rho(X(\omega)) \subset A$ the conclusion $s_0 \in A$ cannot be drawn. It is just for this reason, and not only because of the resulting technical and computational simplifications, why the assumption $m(\emptyset) = 0$ (i. e., a weakened version of the condition of semantical correctness) is often supposed to be valid in works dealing with the Dempster–Shafer theory. We should keep in mind, however, that in this weakened version this condition does not imply that the actual state of SYST is in $U_\rho(X(\omega))$ so that the conclusion $s_0 \in A$ drawn from the inclusion $U_\rho(X(\omega)) \subset A$ is not justified. It is evident that the case $U_\rho(X(\omega)) = \emptyset$ must be, somehow, avoided from our considerations, as $\emptyset \subset A$ holds trivially for each $A \subset S$ no matter whether $s_0 \in A$ or not. It is a matter of continued discussions, whether such an elimination should be realized by a simple erasing of $\emptyset$ from the field of subsets of A, as it is the case when defining the non-normalized version of belief functions, or whether to re-normalize the obtained values to the case when $U_\rho(X(\omega)) \neq \emptyset$. In our work we shall follow both the patterns, but the non-normalized version will be preferred because of the (at least) three following reasons: (i) it is more adequate for the generalizations investigated below, (ii) it is more compatible with the minimax idea on which belief functions are (perhaps implicitly) based and, what follows partially from (ii), (iii) the properties of non-normalized belief functions are more close to some intuitive demands behind.

5 Elementary Properties of Belief Functions

In this chapter we shall survey the most elementary properties of belief functions and some other characteristics derived from them (cf. [38], e. g., for more detail). We shall suppose, throughout this chapter, that the probability space $\langle \Omega, \mathcal{A}, P \rangle$ and the measurable spaces $\langle \mathcal{P}(S), \mathcal{S} \rangle$ and $\langle E, \mathcal{E} \rangle$ are fixed, the dependence of belief functions on possible variations or modifications of these basic stones of our constructions will be investigated in some of the following chapters. We shall also suppose that if the state space S is finite, then the σ-field $\mathcal{S}$ is the maximal one, i. e., $\mathcal{S} = \mathcal{P}(\mathcal{P}(S))$, so that the values $m(A)$, $bel^*(A)$ (and $bel(A)$, if $m(\emptyset) < 1$) are defined for each $A \subset S$ and obeys the usual combinatoric definitions. The properties of belief functions concerning their possible combinations and actualizations will be investigated in the next chapter dealing with the Dempster combination rule.

Definition 5.1. Let the notations and conditions of Definition 4.2 hold. *Nonnormalized plausibility function* induced by the compatibility relation ρ is the mapping $pl^*_\rho : \mathcal{P}(S) \to \langle 0, 1 \rangle$ defined, for each $A \subset S$ for which $bel^*(S - A)$ and $m(\emptyset) = P(\{\omega \in \Omega : U_\rho(X(\omega)) = \emptyset\})$ are defined, by the relation

$$pl^*_\rho(A) = \Big(1 - bel^*_\rho(S - A)\Big) - m(\emptyset). \tag{5.1}$$

If, moreover, $m(\emptyset) < 1$ holds, then the *normalized plausibility function* pl_ρ is defined by the relation

$$pl_\rho(A) = 1 - bel_\rho(S - A). \tag{5.2}$$

□

An easy calculation yields that, if $pl^*_\rho(A)$ is defined, then

$$pl^*_\rho(A) = \Big(1 - bel^*_\rho(S - A)\Big) - m(\emptyset) = \tag{5.3}$$

$$
\begin{aligned}
&= (1 - P(\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset S - A\})) - m(\emptyset) = \\
&= P\left(\{\omega \in \Omega : U_\rho(X(\omega)) \cap A \neq \emptyset\}\right).
\end{aligned}
$$

Hence, if S is finite and $m(A)$ is defined, for each $A \subset S$, by (4.6), then evidently

$$pl^*(A) = \sum\nolimits_{B \subset S,\, B \cap A \neq \emptyset} m(B). \tag{5.4}$$

For the normalized version we obtain that

$$
\begin{aligned}
pl(A) &= 1 - P\left(\{\omega \in \Omega : U_\rho(X(\omega)) \subset S - A\} / \{\omega \in \Omega : U_\rho(X(\omega)) \neq \emptyset\}\right) = (5.5) \\
&= P\left(\{\omega \in \Omega : U_\rho(X(\omega)) \cap A \neq \emptyset\} / \{\omega \in \Omega : U_\rho(X(\omega)) \neq \emptyset\}\right),
\end{aligned}
$$

so that, if S is finite,

$$pl_\rho(A) = (1 - m(\emptyset))^{-1} \sum\nolimits_{B \subset S,\, B \cap A \neq \emptyset} m(B). \tag{5.6}$$

The relation between bel^* (bel, resp.) and pl^* (pl, resp.) is dual in both the direction, as for each $A \subset S$ such that $pl^*(S - A)$ and $m(\emptyset)$ are defined, $bel^*(A)$ is also defined and (5.1) yields that

$$pl^*_\rho(S - A) = (1 - bel^*_\rho(A)) - m(\emptyset), \tag{5.7}$$

hence,

$$bel^*_\rho(A) = (1 - pl^*(A)) - m(\emptyset). \tag{5.8}$$

For the normalized case we obtain easily that

$$bel_\rho(A) = 1 - pl_\rho(S - A). \tag{5.9}$$

The simplified forms of (5.1) and (5.8) in the case when $m(\emptyset) = 0$ are obvious.

The interpretation of the values of plausibility functions is also dual to that in the case of belief functions. Under the condition of semantical correctness the value $bel^*(A)$ can be taken as the probability with which such empirical values (data) have been obtained, that the subject is able to decide or deduce, within the framework of her/his deductive abilities given by the compatibility relation ρ, that the actual internal state of the system under investigation is (must be, in a sense) in the subset A of the state space S. The value $bel(A)$ can be interpreted in the same way just considering the probability normalized to the case when the obtained data are consistent (under the closed world assumption). The value $pl^*(A)$ can be taken as the probability with which such empirical data are obtained that the subject is not able (by proving the contrary) to disprove the hypothesis that the actual internal state of SYST is in A.

Fact 5.1. For each $A \subset B \subset S$ for which the values under consideration are defined the following relations hold.

$$
\begin{aligned}
0 &= bel^*_\rho(\emptyset) = bel_\rho(\emptyset) \leq bel^*_\rho(A) \leq bel^*_\rho(B) \leq bel^*_\rho(S) \leq \qquad (5.10) \\
&\leq bel_\rho(S) = 1,
\end{aligned}
$$

$$
\begin{aligned}
bel_\rho(A) &\leq bel_\rho(A), \qquad bel_\rho(A) \leq bel_\rho(B), && (5.11)\\
0 &= pl^*_\rho(\emptyset) = pl_\rho(\emptyset) \leq pl^*_\rho(A) \leq pl^*_\rho(B) \leq pl^*_\rho(S) \leq && (5.12)\\
&\leq pl_\rho(S) = 1, &&\\
pl^*_\rho(A) &\leq pl_\rho(A), \qquad pl_\rho(A) \leq pl_\rho(B), && (5.13)\\
bel^*_\rho(A) &\leq pl^*_\rho(A), \qquad bel_\rho(A) \leq pl_\rho(A). && (5.14)
\end{aligned}
$$

□

Proof. (5.10) – (5.13) follow immediately from the definitions of the values in question, (5.14) follows from the fact that if $U(X(\omega)) = B \neq \emptyset$ and $B \subset A$, then $U(X(\omega)) \cap A = B \neq \emptyset$ holds as well. □

If the basic space S is finite (and $\mathcal{S} = \mathcal{P}(\mathcal{P}(S))$), then there is a one-to-one relation between non-normalized belief functions and basic probability assignments (b.p.a.'s). Or, if m_1, m_2 are b.p.a.'s on S such that $m_1 \equiv m_2$, i.e. $m_1(A) = m_2(A)$ for all $A \subset S$, then $bel^*_{m_1} \equiv bel^*_{m_2}$ obviously holds due to (4.7). On the other side, let $m_1 \not\equiv m_2$, i.e., let there exist $A \subset S$ such that $m_1(A) \neq m_2(A)$. Obviously, there must exist $A \neq \emptyset$ with this property, as if $m_1(\emptyset) \neq m_2(\emptyset)$ and $m_1(A) = m_2(A)$ for all $\emptyset \neq A \subset S$, then the equality $\sum_{A \subset S} m_1(A) = \sum_{A \subset S} m_2(A) = 1$ cannot hold. Let $A \subset S$ be such that $\emptyset \neq A$, $m_1(A) \neq m_2(A)$, but $m_1(B) = m_2(B)$ for all $\emptyset \neq B \subset S$, $\mathrm{card}(B) < \mathrm{card}(A)$; such $A \subset S$ obviously exists. Then, however,

$$
\begin{aligned}
bel^*_{m_1}(A) = \textstyle\sum_{\emptyset \neq B \subset A} m_1(B) = \sum_{\emptyset \neq B \subset A,\, B \neq A} m_1(B) + m_1(A) \neq \qquad (5.15)\\
\neq \textstyle\sum_{\emptyset \neq B \subset A,\, B \neq A} m_2(B) + m_2(A) = bel^*_{m_2}(A),
\end{aligned}
$$

as $m_1(B) = m_2(B)$ for all $\emptyset \neq B \subset A$, $B \neq A$, so that $bel^*_{m_1} \not\equiv bel^*_{m_2}$.

In the case of normalized belief functions the situation is as follows.

Fact 5.2. Let m_1, m_2 be b.p.a.'s defined on a finite set S. Then $bel_{m_1} \equiv bel_{m_2}$ holds iff $m_1(\emptyset) < 1$, $m_2(\emptyset) < 1$, and

$$
m_1(A) = (1 - m_1(\emptyset))\,(1 - m_2(\emptyset))^{-1}\, m_2(A) \qquad (5.16)
$$

holds for each $\emptyset \neq A \subset S$. □

Proof.

$$
\begin{aligned}
\textstyle\sum_{A \subset S} m_1(A) = \sum_{\emptyset \neq A \subset S} (1 - m_1(\emptyset))\,(1 - m_2(\emptyset))^{-1}\, m_2(A) + m_1(\emptyset) = \qquad (5.17)\\
= (1 - m_1(\emptyset))\,(1 - m_2(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq A \subset S} m_2(A) + m_1(\emptyset) =\\
= (1 - m_1(\emptyset))\,(1 - m_2(\emptyset))^{-1}\,(1 - m_2(\emptyset)) + m_1(\emptyset) = 1,
\end{aligned}
$$

hence, if m_2 is a b.p.a. on S, then m_1 is also b.p.a. on S and vice versa. The equality $0 = bel_{m_1}(\emptyset) = bel_{m_2}(\emptyset)$ holds due to (4.8), as each sum over the empty set of items equals zero. If $\emptyset \neq A \subset S$, then

$$\begin{aligned} bel_{m_1}(A) &= (1-m_1(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset S} m_1(B) = \\ &= (1-m_1(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A} (1-m_1(\emptyset))\,(1-m_2(\emptyset))^{-1}\, m_1(B) = \\ &= (1-m_2(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A} m_2(B) = bel_{m_2}(A), \end{aligned} \tag{5.18}$$

hence, (5.16) implies that $bel_{m_1} \equiv bel_{m_2}$ holds. To prove the inverse implication, let there exist $\emptyset \neq A \subset S$ such that (5.16) does not hold. We can suppose, without any loss of generality, that (5.16) holds for each $B \subset S$ such that $\text{card}(B) < \text{card}(A)$. Then

$$\begin{aligned} bel_{m_1}(A) &= (1-m_1(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A} m_1(B) = \\ &= (1-m_1(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A,\, B \neq A} m_1(B) + (1-m_1(\emptyset))^{-1}\, m_1(A) = \\ &= (1-m_1(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A,\, B \neq A} (1-m_1(\emptyset))\,(1-m_2(\emptyset))^{-1}\, m_2(B) + \\ &\qquad (1-m_1(\emptyset))^{-1}\, m_1(B) \neq \\ &\neq (1-m_1(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A,\, B \neq A} (1-m_1(\emptyset))\,(1-m_2(\emptyset))^{-1}\, m_2(B) + \\ &\qquad +(1-m_1(\emptyset))^{-1}\,(1-m_1(\emptyset))\,(1-m_2(\emptyset))^{-1}\, m_2(A) = \\ &= (1-m_2(\emptyset))^{-1} \textstyle\sum_{\emptyset \neq B \subset A} m_2(B) \;=\; bel_{m_2}(A), \end{aligned} \tag{5.19}$$

so that $bel_{m_1} \equiv bel_{m_2}$ does not hold. The assertion is proved. □

Another important property of belief functions consists in their super-additivity.

Fact 5.3. Let the notations and conditions of Definition 4.2 hold, let $\mathcal{S}_0 \subset \mathcal{P}(S)$ be a nonempty, finite or countable, system of mutually disjoint sets, i.e., $T_1, T_2 \in \mathcal{S}_0$ and $T_1 \neq T_2$ implies that $T_1 \cap T_2 = \emptyset$, such that $bel_\rho^*(T)$, $bel_\rho(T)$ for each $T \in \mathcal{S}_0$ as well as $bel_\rho^*(\bigcup_{T \in \mathcal{S}_0} T)$ and $bel_\rho(\bigcup_{T \in \mathcal{S}_0} T)$ are defined. Then the inequalities

$$bel_\rho^*\left(\textstyle\bigcup_{T \in \mathcal{S}_0} T\right) \geq \textstyle\sum_{T \in \mathcal{S}_0} bel_\rho^*(T) \tag{5.20}$$

and

$$bel_\rho\left(\textstyle\bigcup_{T \in \mathcal{S}_0} T\right) \geq \textstyle\sum_{T \in \mathcal{S}_0} bel_\rho(T) \tag{5.21}$$

hold. □

Proof. If $\emptyset \neq U_\rho(X(\omega)) \subset T_1$ holds for some $\omega \in \Omega$ and $T_1 \in \mathcal{S}_0$, it is obviously impossible that the same nonempty set $U_\rho(X(\omega))$ is a subset of some $T_2 \in \mathcal{S}_0$, $T_2 \neq T_1$, as $T_1 \cap T_2 = \emptyset$. In other terms, for each $T_1, T_2 \in \mathcal{S}_0$, $T_1 \neq T_2$, the relation

$$\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset T_1\} \cap \{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset T_2\} = \emptyset \tag{5.22}$$

is valid. At the same time the inclusion

$$\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset T\} \subset \left\{\omega \in \Omega : \emptyset \neq U_\rho(X(\omega)) \subset \textstyle\bigcup_{T \in \mathcal{S}_0} T\right\} \tag{5.23}$$

also holds for each $T \in \mathcal{S}_0$, so that

$$\begin{aligned} & \textstyle\sum_{T \in \mathcal{S}_0} P\left(\{\omega \in \Omega : \emptyset \neq U\rho(X(\omega)) \subset T\}\right) = \sum_{T \in \mathcal{S}_0} bel^*_\rho(T) \leq \\ \leq \; & P\left(\{\omega \in \Omega : \emptyset \neq U\rho(X(\omega)) \subset \textstyle\bigcup_{T \in \mathcal{S}_0} T\}\right) = bel^*_\rho\left(\textstyle\bigcup_{T \in \mathcal{S}_0} T\right) \end{aligned} \tag{5.24}$$

also holds and (5.20) is proved. Dividing all the values in (5.24) by $P(\{\omega \in \Omega : U_\rho(X(\omega)) \neq \emptyset\})$, we obtain (5.21). $\square$

In particular, for $T_1,\, T_2 \subset S,\; T_1 \cap T_2 = \emptyset,\; bel^*_\rho(T_1 \cup T_2) \geq bel^*_\rho(T_1) + bel^*_\rho(T_2)$ and analogously for bel_ρ. This property generalizes the usual σ-additivity of probabilistic measures, when $Pr\left(\bigcup_{T \in \mathcal{S}_0} T\right) = \sum_{T \in \mathcal{S}_0} Pr(T)$ holds for each nonempty finite or countable system $\mathcal{S}_0$ such that the probabilities in question are defined. The generalization goes still farther, as for each $T_1,\, T_2 \subset S$ for which the values of belief functions are defined, the inequalities

$$bel^*_\rho(T_1 \cup T_2) \geq bel^*_\rho(T_1) + bel^*_\rho(T_2) - bel^*_\rho(T_1 \cap T_2) \tag{5.25}$$

and

$$bel_\rho(T_1 \cup T_2) \geq bel_\rho(T_1) + bel_\rho(T_2) - bel_\rho(T_1 \cap T_2) \tag{5.26}$$

hold. These inequalities can be generalized to any finite system $\mathcal{S}_0 = \{T_1, T_2, \ldots, T_n\}$ of subsets of S for which, and for their union, bel^*_ρ and bel_ρ are defined. Namely, we obtain that

$$\begin{aligned} & bel^*_\rho\left(\textstyle\bigcup_{k=1}^n T_k\right) \geq \\ \geq \; & \textstyle\sum_{k=1}^n (-1)^{k+1} \sum_{\substack{\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\} \\ \text{card}\{i_1, i_2, \ldots, i_k\} = k}} bel^*_\rho\left(\bigcap_{m=1}^k T_{i_m}\right) \end{aligned} \tag{5.27}$$

and similarly for bel_ρ. In general, belief functions, share many properties with the so called *inner measures* induced by probabilistic measures (cf. [15], e. g., for more detail), and we shall take profit of these relations in one of the following chapters, when we shall look for an appropriate approximation of belief functions in the cases when they are not definable by the set-valued random variable $U_\rho(X(\cdot))$, i. e., in the case of such subsets $A \subset S$ for which the power-set $\mathcal{P}(A)$ does not belong to the σ-field $\mathcal{S} \subset \mathcal{P}(\mathcal{P}(S))$.

Some particular cases of belief functions and (if S is finite) the corresponding basic probability assignments (b.p.a.'s) are perhaps worth being introduced explicitly. For the sake of simplicity we shall consider only the case when the state space S is finite and $\mathcal{S} = \mathcal{P}(\mathcal{P}(S))$, even if, as will be evident, some of the notions can be extended also to the case when S is infinite, supposing that $m(A) = P(\{\omega \in \Omega : U_\rho(X(\omega)) = A\})$ is defined for each $A \subset S$ occurring in the definition of the particular notion, i. e., supposing that $\{A\} \in \mathcal{S}$ holds in such cases.

Let m be a b.p.a. on a finite set S. A subset $A \subset S$ is called a *focal element* of m, if $m(A) > 0$ holds. A b.p.a. m_S on S is called *vacuous*, if $m_S(S) = 1$, consequently, $m_S(A) = 0$ for all $A \subset S,\; A \neq S$. The corresponding belief functions $bel^*_{m_S}$ and bel_{m_S} are obviously identical with each other and are called *vacuous belief functions* on S, evidently, $bel_{m_S}(S) = 1$ and $bel_{m_S}(A) = 0$ for all $A \subset S,\ A \neq S$. A b.p.a. m_A is called *singular* in $A \subset S$, if $m_A(A) = 1$, hence, $m_A(B) = 0$ for all $B \subset S,\ B \neq A$, so that vacuous belief function m_S is singular in S. For the belief functions $bel^*_{m_A}$ and bel_{m_A}, which are also identical, if $A \neq \emptyset$, we obtain that $bel_{m_A}(B) = 1$, if $A \subset B$, $bel_{m_A}(B) = 0$ otherwise. *(Totally) inconsistent* b.p.a. is the b.p.a. $m_\emptyset$ singular in the empty set $\emptyset$. In this case $bel^*_{m_\emptyset}(A) = 0$ for all $A \subset S$ and $bel_{m_\emptyset}$ is not defined. A b.p.a. m is called *partially (in)consistent*, if $0 < m(\emptyset) < 1$ holds and it is called *(totally) consistent*, if $m(\emptyset) = 0$. The intuition behind the adjectives "(in)consistent" just introduced is based on the interpretation preferring the closed world assumption and explained in the foregoing chapter. A b.p.a. $m_{A,\varepsilon}$ on S is called ε-quasi-singular in $A \subset S$, for a given real number $0 \leq \varepsilon \leq 1$, if $m_{A,\varepsilon}(A) = 1 - \varepsilon,\ m_{A,\varepsilon}(S) = \varepsilon$, and $m_{A,\varepsilon}(B) = 0$ for all $B \subset S,\ B \neq A,\ B \neq S$. Hence, $m_A \equiv m_{A,0}$ and $m_{A,1} \equiv m_S$ hold for each $A \subset S$. In [40] the author introduces the exponential form of notation for ε-quasi-singular b.p.a.'s, when he writes A^ε for $m_{A,\varepsilon}$. The fact that it is not the probability $1 - \varepsilon$ ascribed to A but the complementary probability ε left to the whole state space S which plays the role of the exponent, is motivated by the resulting simplified form of the Dempster combination rule.

The following special case of b.p.a.'s and belief functions should be also mentioned explicitly. If m is such a b.p.a. on a finite set S that all focal elements of m are singletons of S, i.e., if $m(A) = 0$ for all $A \subset S$ such that $\mathrm{card}(A) \neq 1$, then the induced belief function bel_m, obviously identical with $bel^*_{m_1}$, as it is the case for all totally consistent b.p.a.'s, is a probability measure on S. Or, for each $A, B \subset S$ such that $A \cap B = \emptyset$ we obtain that $A \cap B = \emptyset$ we obtain that

$$\begin{aligned} bel_m(A \cup B) = \textstyle\sum_{\emptyset \neq C \subset A \cup B} m(C) = \sum_{s \in A \cup B} m(\{s\}) = \\ = \textstyle\sum_{s \in A} m(\{s\}) + \sum_{s \in B} m(\{s\}) = bel_m(A) + bel_m(B). \end{aligned} \tag{5.28}$$

In the same case, i.e., when $m(A) > 0$ implies $\mathrm{card}(A) = 1$, also the identities $bel_m(A) = pl_m(A)$ and $bel^*_m(A) = pl_m(A)$ hold for each $A \subset S$, so that also the plausibility function is a probability measure on $\mathcal{P}(S)$. Or, for each $A \subset S$ and each $s \in S$, $\{s\} \cap A \neq \emptyset$ iff $\{s\} \subset A$ iff $s \in A$, so that

$$\begin{aligned} bel_m(A) = bel^*_m(A) = \textstyle\sum_{\emptyset \neq B \subset A} m(B) = \sum_{s \in A} m(\{s\}) = \\ = \textstyle\sum_{\{s\} \cap A \neq \emptyset} m(\{s\}) = \sum_{B \cap A \neq \emptyset} m(B) = pl^*_m(A) = pl_m(A). \end{aligned} \tag{5.29}$$

When the focal elements of a b.p.a. m defined on a finite state space S are *nested*, belief function bel^*_m converts into the so called *possibilistic measure* ([6]). In its most simple setting, and if the state space S is finite, possibilistic (or: possibility) measure on S is a mapping $\pi : \mathcal{P}(S) \to \langle 0, 1 \rangle$ such that $\pi(A \cup B) = \max\{\pi(A), \pi(B)\}$ for each $A, B \subset S$. Possibilistic measure in *normalized*, if $\pi(\Omega) = 1$. Focal elements of a b.p.a. m on S are nested, if for each $A, B \subset S$ such that $m(A) > 0$ and $m(B) > 0$ holds, either $A \subset B$ or $B \subset A$. It follows that there exists, in such a case, a finite (and uniquely determined) sequence $A_1 \subset A_3 \subset A_3 \subset \cdots \subset A_n$ of subsets of S such that $m(A) > 0$ iff $A \subset \{A_1, A_2, \ldots, A_n\}$ holds for each $A \subset S$. So,

$$bel^*_m(A) \ = \ \textstyle\sum_{\emptyset \neq C \subset A} m(C) = \sum_{i=1}^{i(A)} m(A_i), \tag{5.30}$$

$$bel^*_m(B) \;=\; \textstyle\sum_{\emptyset \neq C \subset B} m(C) = \sum_{i=1}^{i(B)} m(A_i), \tag{5.31}$$

where integers $i(A)$ and $i(B) \leq n$ are uniquely defined. Moreover,

$$\begin{aligned} bel^*_m(A \cup B) &= \textstyle\sum_{i=1}^{\max\{i(A), i(B)\}} m(A_i) = \\ = \;\; &\max\left\{\textstyle\sum_{i=1}^{i(A)} m(A_i), \sum_{i=1}^{i(B)} m(A_i)\right\} = \max\{bel^*_m(A),\ bel^*_m(B)\}. \end{aligned} \tag{5.32}$$

Obviously, in the same case bel_m is a normalized possibilistic measure.

There is perhaps the best time and place, now, to mention explicitly one of the important differences between the Dempster–Shafer theory and probability theory or, to be more correct, between belief functions and probability measures, favourizing basic probability assignments and belief functions when describing the case of total ignorance. As already mentioned, the Bayes approach to statistical decision making under uncertainty requests the a priori probability distribution on the state space S to be defined and known to the subject. If this distribution is completely unknown and if the set S is finite, the Bayes approach often applies the so called Laplace principle: the lack of any reason for which we have to prefer one possibility to another can be taken as a sufficient reason to take both the possibilities as equivalent. Taking the words "to prefer one alternative to another" in the sense "to ascribe greater a priori probability to the first alternative than to the other one", we arrive at the solution to ascribe the same a priori probability $(\mathrm{card}(S))^{-1}$ to each $s \in S$. If S is infinite, a generalization of this principle results in the so called *principle of maximal entropy*, the uniform (equiprobable) probability distribution on a finite S being its special case.

Within the framework of Dempster–Shafer theory the uniform probability distribution on finite S can be easily defined by the b.p.a. m_{eq} on S such that $m_{\mathrm{eq}}(\{s\}) = (\mathrm{card}(S))^{-1}$ for all singletons $\{s\}$, $s \in S$, consequently, $m_{\mathrm{eq}}(A) = 0$ for each $A \subset S$, $\mathrm{card}(A) \neq 1$. However, the case of total ignorance, as far as the actual internal state of the system SYST under consideration is concerned, is better defined by the vacuous b.p.a. m_S, when $m_S(S) = 1$, hence, $m_S(A) = 0$ for all $A \subset S$, $A \neq S$. This vacuous b.p.a. corresponds to the situation when there are absolutely no arguments, neither of uncertain or stochastic nature, in favour of the hypothesis that the actual internal state of SYST belongs to some *proper* subset of S. The only fact which is taken as granted is the closed world assumption according to which the space S is the exhaustive list of *all* possible internal states of SYST. On the other side, the "equiprobable" b.p.a. m_{eq} describes the situation when every data item brings an argument in favour of one particular $s \in S$, and the numbers or weights of arguments for particular values $s \in S$ are the same or at least are not distinguishable from each other within some reasonable tolerance bounds. The fact that probability theory works only with equiprobable distributions as mathematical models of total ignorance is caused by historical reasons. Probabilistic models were based on games and bets ideas and the estimation that there was the probability 1/2 that head occurs and the same probability 1/2 that tail occurs when tossing a fair coin was based on the results of the past tosses when approximately 1/2 of results were heads, i. e., arguments in favour of the hypothesis that head occurs also in the next toss, and approximately 1/2 of the results were tails, i. e., arguments in favour of the hypothesis that tail occurs again in the next toss. The possibility of splitting the coin in such a way that *both* the sides occur simultaneously, or the possibility of disappearing of the coin during the toss so that *no* side occurs, was taken as a priori avoided due to the accepted tossing rules. In the case of b.p.a.'s we can admit also the possibility that the coin splits, setting $m(\{T, H\}) > 0$, and/or the possibility that the coin disappears, setting $m(\emptyset) > 0$.

Let us also discuss, very briefly, the often posed question whether Dempster–Shafer theory is a *generalization* or an *application* of the probability theory. The answer can be affirmative in both the cases, having very carefully re-formulated the question. So, Dempster–Shafer theory is a generalization of the probability theory in the sense that belief function, as a measure of uncertainty, is a generalization of probability measure. In other words, probability *measure* results as a special case of belief functions (degrees or *measures* of belief) under some additional conditions imposed, namely that $m(A) = 0$ for each $A \subset S$ with $\mathrm{card}(A) \neq 1$. On the other side, Dempster–Shafer *theory* is an application of probability *theory* in the sense that all notions and characteristics used in Dempster–Shafer theory can be defined by appropriate notions of probability theory, including the degrees of beliefs which are defined by the probabilities with which certain random sets (values of set-valued random variables) satisfy some relation of set-theoretic inclusion. The situation is similar to that with the inner and outer (probabilistic) measures (cf. [15]) which are generalizations of the original (probabilistic) measures, as they are defined also for non-measurable sets and their values agree with those of the original (probabilistic) measures for measurable sets (random events), but at the same time inner and outer (probabilistic) measures are applications of probability or measure theory, as they are defined by (probabilistic) measures through appropriate supremum or infimum operations.

6 Probabilistic Analysis of Dempster Combination Rule

In the real world around us, a subject's knowledge concerning this world in general, and investigated system(s) and their (its) environments in particular, are not of static, but rather of dynamic nature. In other words, this knowledge is subjected to changes involved by the time passing. These changes can be caused either by the changes taking places either in the world itself, or by changes of the body of evidence and laws of the nature known to the subject. The changes should be applied to the knowledge of sure deterministic nature (more correctly, the knowledge taken as sure in the given context and under the given circumstances), as well as to the knowledge charged by uncertainty. In this work we focus our attention to the knowledge expressed in the terms of compatibility relations, basic probability assignments and belief functions, so that our aim will be, in this chapter, to investigate the ways in which one compatibility relation, b.p.a., or belief function can and should be modified when obtaining some more information described by another compatibility relation, b.p.a., or belief function. As a rule, in Dempster–Shafer theory such a modification (actualization) is realized applying the so called *Dempster combination rule*. In this chapter we shall introduce this rule using the probabilistic model and terms presented above and we shall discover and formalize explicitly the usually only tacitly assumed hidden assumptions behind this combination rule. As in the foregoing chapters, we shall begin with an informal intuition behind our explanation, leaving this intuition aside and and returning to a formalized mathematical level of presentation as soon as possible.

Let us consider, as above, the system SYST with the space S of possible internal states. The task is, again, to identify the actual internal state of SYST or at least to decide, whether this state belongs to a (proper, to avoid trivialities) subset A of S. However, the problem is solved, now, by *two* subjects, SUB_1 and SUB_2. For both of them, the actual state of the system is not directly observable, so that SUB_1, as well as SUB_2, must solve the problem, whether the internal state of SYST is in A, indirectly, using some observations or other empirical data concerning the system and its environment. These data can be, in general, not only different, but even of different nature for both the subjects, hence, we shall assume that the empirical data being at the disposal of SUB_1, take values in a nonempty space E_1, the empirical data of SUB_2 belong to an $E_2 \neq \emptyset$. As before, the empirical data of both the subjects are supposed to be of random character, and because of the general abstract nature of the notion of probability space we can assume, that the empirical data being at the disposal of SUB_1 are formally described by a random variable X_1, defined on a fixed probability space $\langle \Omega, \mathcal{A}, P \rangle$ and taking its values in a measurable space $\langle E_1, \mathcal{E}_1 \rangle$ defined over the set E_1. Analogously, X_2 is a random variable defined on the same probability space $\langle \Omega, \mathcal{A}, P \rangle$ and taking its values in $\langle E_2, \mathcal{E}_2 \rangle$, where $\mathcal{E}_2$ is an appropriate σ-field of subsets of E_2; X_2 describes the empirical data being at the disposal of SUB_2. The bodies of a priori knowledge of both the subjects are defined by corresponding compatibility relations: $\rho_1 \subset S \times E_1$ for SUB_1 and $\rho_2 \subset S \times E_2$ for SUB_2. The degrees of beliefs of both the subjects can be quantified by basic probability assignments m_1, m_2, and by belief functions bel_1^*, bel_2^* or bel_1, bel_2, namely, as before, for each $A \subset S$ and for both $i = 1, 2$,

$$\begin{aligned} U(X_i(\omega)) &= \{s \in S : \rho_i(s, X_i(\omega)) = 1\}, & (6.1)\\ m_i(A) &= P\left(\{\omega \in \Omega : U(X_i(\omega)) = A\}\right), & (6.2)\\ bel_i^*(A) &= P\left(\{\omega \in \Omega : \emptyset \neq U(X_i(\omega)) \subset A\}\right), & (6.3)\\ bel_i(A) &= P\left(\{\omega \in \Omega : U(X_i(\omega)) \subset A\} / \{\omega \in \Omega : U(X_i(\omega)) \neq \emptyset\}\right), & (6.4) \end{aligned}$$

supposing that the probabilities in question are defined.

Both the subjects, however, can arrive at the decision to co-operate with each other and to combine sophistically their a priori knowledge and empirical data in order to obtain better (in a sense to be explicitly defined later) results concerning the actual internal state of SYST than the results achievable by each of them separately. Another interpretation can read that there is a third (meta) subject SUB_{12} who has at her/his disposal the apriori knowledge (i. e., the compatibility relations) and the empirical data of both SUB_1 and SUB_2 and combines them sophistically together.

First of all, the empirical spaces E_1, E_2 are combined into their Cartesian product $E_{12} = E_1 \times E_2$ and the σ-field $\mathcal{E}_{12} \in \mathcal{P}(E_1 \times E_2)$ is defined by the minimal σ-field containing all the rectangles $F_1 \times F_2$ such that $F_1 \in \mathcal{E}_1$, $F_2 \in \mathcal{E}_2$. Random variables X_1 and X_2 are combined into the vector random variable $X_{12} = \langle X_1, X_2 \rangle$ defined on the probability space $\langle \Omega, \mathcal{A}, P \rangle$ and taking its values in $\langle E_{12}, \mathcal{E}_{12} \rangle$. The well-known theorem about the extension of measure (cf., e. g. [15]) yields that for each $F_{12} \in \mathcal{E}_{12}$ the probability

$$P\left(\{\omega \in \Omega : \langle X_1(\omega), X_2(\omega) \rangle \in F_{12}\}\right) \tag{6.5}$$

is correctly and unambiguously defined, so that X_{12} is, in fact, a random variable.

The next step consists in a combination of compatibility relations $\rho_1 \subset S \times E_1$ and $\rho_2 \subset S \times E_2$ into a new compatibility relation $\rho_{12} \subset S \times E_{12}$. Dempster combination rule is based on the assumption that

$$\rho_{12}(s, \langle x_1, x_2 \rangle) = \min\{\rho_1(s, x_1), \rho_2(s, x_2)\} \tag{6.6}$$

for all $s \in S,\ x_1 \in E_1,\ x_2 \in E_2$. Hence, a state $s \in S$ is taken as incompatible with the empirical values x_1, x_2, if it is taken as incompatible by at least one of the two (or more, as can be immediately generalized) subjects in question. The interpretation behind is that the pieces of knowledge, according to which one of the subject is able to refuse a state $s \in S$ on the ground of a data item $x_1 \in E_1$, are objectively valid beyond any doubts and they are, therefore, accepted by the other subject(s). In other words, the situation when $\rho_1(s,x) = 0$ and $\rho_2(s,x) = 1$ for some $s \in S$ and $x \in E_1 \cap E_2$ must be caused by the fact that the knowledge of SUB_2 is only fragmental so that she/he does not know that s and x are incompatible, even if it is objectively true. Having been informed about this fact by SUB_1, SUB_2 immediately accepts this fact and modifies her/his compatibility relation ρ_1 into $\rho_{12}(s,\ \langle x, x\rangle) = \rho_1(s,x) = 0$. So, the phenomenon of subjectivity of compatibility relations concerning the same system reduces, for two or more subjects, to their partial knowledge (or ignorance) of one common objectively valid compatibility relation $\rho_0(s,\ \langle x_1, x_2\rangle)$. The combination of compatibility relations defined by (6.6) can be called the *optimistic* one as there are the beliefs in validity of data which are shared by other subject(s), not the doubts concerning this validity. The *condition of optimistic combination* of compatibility relations is one of the two basic assumptions of Dempster combination rule (the other will be introduced a few-lines below), even if these conditions are often only tacitly assumed and not explicitly stated. An alternative approach, based on the dual idea that there are just the doubts which are shared and that a state s is taken as incompatible iff it is taken as incompatible by all the subjects separately, in symbols, the approach when the combined compatibility relation $\hat\rho$ is defined by

$$\hat\rho_{12}(s,\ \langle x_1, x_2\rangle) = \max\{\rho_1(s,x_1),\ \rho_2(s,x_2)\}, \tag{6.7}$$

will be briefly investigated at the end of this chapter.

Relation (6.6) immediately implies that for each $\langle x_1, x_2\rangle \in E_{12}$,

$$\begin{aligned} & U_{\rho_{12}}(\langle x_1, x_2\rangle) = \{s \in S : \rho_{12}(s,\ \langle x_1, x_2\rangle) = 1\} = \\ = \ & \{s \in S : \rho_1(s,x_1) = 1\} \cap \{s \in S : \rho_2(s,x_2) = 1\} = \\ = \ & U_{\rho_1}(x_1) \cap U_{\rho_2}(x_2). \end{aligned} \tag{6.8}$$

Applying (6.8) to the case when $x_1 = X_1(\omega)$ and $x_2 = X_2(\omega)$ are values of random variables, and supposing that all the probabilities in question are defined, we can define the b.p.a. m_{12} and belief functions bel^*_{12} and bel_{12}, for $A \subset S$, by

$$m_{12}(A) = P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) = A\}\right), \tag{6.9}$$

$$bel^*_{12}(A) = P\left(\{\omega \in \Omega : \emptyset \neq U_{\rho_1}(\omega) \cap U_{\rho_2}(X_2(\omega)) \subset A\}\right), \tag{6.10}$$

$$\begin{aligned} bel_{12}(A) = & \ P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_1(\omega)) \subset A\}/\right. \\ & \left.\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) \neq \emptyset\}\right). \end{aligned} \tag{6.11}$$

The other basic assumption on which Dempster combination rule relies is that of *statistical (stochastical) independence* of the random variables X_1 and X_2. Hence, we suppose that for each $F_1 \subset E_1,\ F_1 \in \mathcal{E}_2$ and $F_2 \subset E_2,\ F_2 \in \mathcal{E}_2$, the equality

$$\begin{aligned} & P\left(\{\omega \in \Omega : \langle X_1(\omega),\ X_2(\omega)\rangle \in F_1 \times F_2\}\right) \\ & P\left(\{\omega \in \Omega : X_1(\omega) \in F_1\}\right)\ P\left(\{\omega \in \Omega : X_2(\omega) \in F_2\}\right) \end{aligned} \tag{6.12}$$

holds. Suppose, moreover, that S is finite and that both the mappings $U_{\rho_1}(X_1(\cdot))$ and $U_{\rho_2}(X_2(\cdot))$ are set-valued random variables defined on the probability space $\langle \Omega, \mathcal{A}, P \rangle$ and taking their values in the measurable space $\langle \mathcal{P}(S), \mathcal{P}(\mathcal{P}(S)) \rangle$. Then, for each $B, C \subset S$ such that

$$\begin{aligned} U_{\rho_1}^{-1}(B) &= \{x_1 \in E_1 : U_{\rho_1}(x_1) = B\} \in \mathcal{E}_1, && (6.13) \\ U_{\rho_2}^{-1}(B) &= \{x_2 \in E_2 : U_{\rho_2}(x_2) = C\} \in \mathcal{E}_2 && (6.14) \end{aligned}$$

hold, we obtain that

$$\begin{aligned} & P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) = B,\, U_{\rho_2}(X_2(\omega)) = C\}\right) = && (6.15) \\ = \; & P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) = B\} \cap \{\omega \in \Omega : U_{\rho_2}(X_2(\omega)) = C\}\right) = \\ = \; & P\left(\{\omega \in \Omega : X_1(\omega) \in U_{\rho_1}^{-1}(B)\} \cap \{\omega \in \Omega : X_2(\omega) = U_{\rho_2}^{-1}(C)\}\right) = \\ = \; & P\left(\{\omega \in \Omega : X_1(\omega) \in U_{\rho_1}^{-1}(B)\}\right) \cdot P\left(\{\omega \in \Omega : X_2(\omega) \in U_{\rho_2}^{-1}(C)\}\right) = \\ = \; & P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) = B\}\right) \cdot P\left(\{\omega \in \Omega : U_{\rho_2}(X_2(\omega)) = C\}\right). \end{aligned}$$

Consequently, also the set-valued random variables $U_{\rho_1}(X_1(\cdot))$ and $U_{\rho_2}(X_2(\cdot))$ are statistically independent. Due to the condition that the state space S is finite the following factorization holds true. For each $A \subset S$, (6.9) yields that

$$\begin{aligned} & m_{12}(A) = P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) = A\}\right) = && (6.16) \\ = \; & \sum\nolimits_{B,C \subset S,\, B \cap C = A} P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) = B,\, U_{\rho_2}(X_2(\omega)) = C\}\right) = \\ = \; & \sum\nolimits_{B,C \subset S,\, B \cap C = A} P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) = B\}\right) \cdot P\left(\{\omega \in \Omega : U_{\rho_2}(X_2(\omega)) = C\}\right) = \\ = \; & \sum\nolimits_{B,C \subset S,\, B \cap C = A} m_1(B)\, m_2(C). \end{aligned}$$

Analogously, the relations (6.10) and (6.11) yield that

$$\begin{aligned} & bel_{12}^{*}(A) = P\left(\{\omega \in \Omega : \emptyset \neq U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) \subset A\}\right) = && (6.17) \\ = \; & \sum\nolimits_{D \subset S,\, \emptyset \neq D \subset A} P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) = D\}\right) = \\ = \; & \sum\nolimits_{B,C \subset S,\, \emptyset \neq B \cap C \subset A} P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) = B,\, U_{\rho_2}(X_2(\omega)) = C\}\right) = \\ = \; & \sum\nolimits_{B,C \subset S,\, \emptyset \neq B \cap C \subset A} m_1(B)\, m_2(C), \end{aligned}$$

and

$$\begin{aligned} bel_{12}(A) &= P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) \subset A\} / \right. && (6.18) \\ & \qquad \left. / \{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) \neq \emptyset\}\right) = \\ &= \frac{P\left(\{\omega \in \Omega : \emptyset \neq U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) \subset A\}\right)}{P\left(\{\omega \in \Omega : \emptyset \neq U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega))\}\right)} = \\ &= \frac{\sum_{\emptyset \neq D \subset A} m_{12}(D)}{\sum_{\emptyset \neq D}} = \frac{\sum_{B,C \subset S,\, \emptyset \neq B \cap C \subset A} m_1(B)\, m_2(C)}{\sum_{B,C \subset S,\, \emptyset \neq B \cap C} m_1(B)\, m_2(C)}, \end{aligned}$$

supposing that the conditional probability in (6.18) is defined, i. e., supposing that there exist $B, C \subset S$ such that $B \cap C \neq \emptyset$ and $m_1(B) > 0$, $m_2(C) > 0$ hold. The relations (6.16), (6.17) and (6.18) are the well-known combinatoric formulas for the b.p.a. m_{12} and the belief functions bel^*_{12} and bel_{12}. The b.p.a. m_{12} is called the *Dempster product* of the b.p.a.'s m_1 and m_2 and is usually denoted by $m_1 \oplus m_2$, bel^*_{12} is called the *(nonnormalized) Dempster product* of the nonnormalized belief functions bel^*_1 and bel^*_2 and is denoted by $bel^*_1 \oplus bel^*_2$. Finally, bel_{12} is called the *(normalized) Dempster product* of (normalized) belief functions bel_1 and bel_2 and is denoted by $bel_1 \oplus bel_2$.

As can be easily proved, the Dempster operation $\oplus$ is commutative and associative in the space of b.p.a.'s as well as in the space of belief functions. In symbols, the equalities

$$
\begin{aligned}
(m_1 \oplus m_2)(A) &= (m_2 \oplus m_1)(A), \qquad (6.19)\\
\left(bel^*_{m_1} \oplus bel^*_{m_2}\right)(A) &= \left(bel^*_{m_2} \oplus bel^*_{m_1}\right)(A),\\
\left(bel_{m_1} \oplus bel_{m_2}\right)(A) &= \left(bel_{m_2} \oplus bel_{m_1}\right)(A),\\
((m_1 \oplus m_2) \oplus m_3)(A) &= (m_1 \oplus (m_2 \oplus m_3))(A),\\
\left((bel^*_{m_1} \oplus bel^*_{m_2}) \oplus bel^*_{m_3}\right)(A) &= \left(bel^*_{m_1} \oplus (bel^*_{m_2} \oplus bel^*_{m_3})\right)(A),\\
((bel_{m_1} \oplus bel_{m_2}) \oplus bel_{m_3})(A) &= (bel_{m_1} \oplus (bel_{m_2} \oplus bel_{m_3}))(A)
\end{aligned}
$$

holds for each $A \subset S$, supposing that the values on one side of the corresponding equality are defined. The validity of these equalities follows from the commutativity and associativity of the set-theoretic operation of intersection. E. g.,

$$
\begin{aligned}
& (m_1 \oplus m_2)(A) = P(\{\omega \in \Omega : U_{\rho_1}(\omega) \cap U_{\rho_2}(\omega) = A\}) = \qquad (6.20)\\
= \ & P(\{\omega \in \Omega : U_{\rho_2}(\omega) \cap U_{\rho_1}(\omega) = A\}) = (m_2 \oplus m_1)(A),\\
& ((m_1 \oplus m_2) \oplus m_3)(A) = P(\{\omega \in \Omega : (U_{\rho_1}(\omega) \cap U_{\rho_2}(\omega)) \cap U_{\rho_3}(\omega) = A\}) \qquad (6.21)\\
= \ & P(\{\omega \in \Omega : U_{\rho_1}(\omega) \cap (U_{\rho_2}(\omega) \cap U_{\rho_3}(\omega)) = A\}) = (m_1 \oplus (m_2 \oplus m_3))(A).
\end{aligned}
$$

Replacing the equalities "$\ldots = A$" by "$\emptyset \neq \ldots \subset A$" we obtain the proofs for bel^*, the modification for bel is obvious. Relations for belief functions follow also from the fact that $bel^*_{m_1} \oplus bel^*_{m_2}$ is defined by $bel^*_{m_1 \oplus m_2}$, so that commutativity and associativity follow from (6.20) and (6.21); the case of bel is the same. If the state space S is finite, all the equalities in (6.19) can be proved also in a purely combinatoric way descending from the combinatoric definition of the Dempster operation, cf. (6.16), (6.17) and (6.18).

The vacuous b.p.a. m_S (i. e., $m_S = 1$) and the induced vacuous belief function bel_{m_S} ($\equiv bel^*_{m_S}$) are the *unit* elements with respect to the Dempster operation $\oplus$ understood as product (of course, we can call m_S and bel_{m_S} the *zero* elements, if $\oplus$ is understood as addition). In symbols,

$$
\begin{aligned}
m \oplus m_S &\equiv m_S \oplus m \equiv m, \qquad (6.22)\\
bel^*_m \oplus bel^*_{m_S} &\equiv bel^*_{m_S} \oplus bel^*_m \equiv bel^*_m, \qquad (6.23)
\end{aligned}
$$

hold for each b.p.a. m, if bel_m is defined, (6.23) holds also for bel_m. Preferring the terminology under which $\oplus$ is taken as product, the role of the zero element is played by the absolutely inconsistent b.p.a. $m_\emptyset$ (i.e., $m_\emptyset(\emptyset) = 1$) in the sense that

$$m \oplus m_\emptyset \equiv m_\emptyset \oplus m \equiv m_\emptyset \tag{6.24}$$

holds for each b.p.a. m. The relation

$$bel_m^* \oplus bel_{m_\emptyset}^* \equiv bel_{m_\emptyset}^* \oplus bel_m^* \equiv 0 \tag{6.25}$$

holds trivially, but $bel_{m_\emptyset}$ obviously is not defined. The only what we have to take into consideration when proving (6.22) to (6.25) is that m_S is defined by the compatibility relation $\rho_S = S \times E$, so that $U_{\rho_S}(\omega) = S$ for each $\omega \in \Omega$, and $m_\emptyset$ is defined by the compatibility relation $\rho_\emptyset = \emptyset \subset S \times E$, so that $U_{\rho_\emptyset}(\omega) = \emptyset$ for each $\omega \in \Omega$.

Perhaps the most trivial problem arising almost immediately, but unsolvable within the framework developed till now, is that of invertibility of b.p.a.'s (and, consequently, also of belief functions). Given a b.p.a. m on a state space S, does there exist a b.p.a. m^{-1} on S such that $m \oplus m^{-1} \equiv m_S$ (the vacuous b.p.a.) would hold? In other terms, does there exist an inverse b.p.a., with respect to the Dempster combination rule, for a given b.p.a. on S? A simple reasoning proves that such an inverse b.p.a. exists only when $m \equiv m_S$, and in this case $m_S^{-1} \equiv m_S$ holds. Or, if m_1, m_2 are b.p.a.'s on S generated by compatibility relations ρ_1, ρ_2, then

$$(m_1 \oplus m_2)(S) = P\left(\{\omega \in \Omega : U_{\rho_1}(\omega) \cap U_{\rho_2}(\omega) = S\}\right), \tag{6.26}$$

so that $(m_1 \oplus m_2)(S) = 1$ implies that

$$P\left(\{\omega \in \Omega : U_{\rho_1}(\omega) = S\}\right) = P\left(\{\omega \in \Omega : U_{\rho_2}(\omega) = S\}\right) = 1, \tag{6.27}$$

as $U_{\rho_i}(\omega) \subset S$ holds for both $i = 1, 2$. Hence, $m_1(S) = m_2(S) = 1$ follows, consequently, $(m_1 \oplus m_2)(S) = 1$ implies that $m_1 \equiv m_2 \equiv m_S$. A nontrivial solution to the invertibility problem, enabling to define, at least partially, an inverse operation to the Dempster combination rule $\oplus$, can be achieved when generalizing appropriately the notion of basic probability assignment. This problem will be investigated, in more detail, in one of the following chapters.

The dual combination $\hat{\rho}_{12}$ of compatibility relations ρ_1 and ρ_2, defined by (6.7), can be processed like ρ_{12} above in order to obtain dual combinations $m_1 \hat{\oplus} m_2$, $bel_1^* \hat{\oplus} bel_2^*$, and $bel_1 \hat{\oplus} bel_2$ of b.p.a.'s m_1, m_2 and belief functions bel_1^*, bel_2^* (bel_1, bel_2, resp.). As can be easily seen, for each $\langle x_1, x_2 \rangle \in E_{12} = E_1 \times E_2$,

$$\begin{aligned} & U_{\hat{\rho}_{12}}(\langle x_1, x_2 \rangle) = \{s \in S : \hat{\rho}_{12}(s, \langle x_1, x_2 \rangle) = 1\} = \\ = \; & \{s \in S : \rho_1(s, x_1) = 1\} \cup \{s \in S : \rho_2(s, x_2) = 1\} = U_{\rho_1}(x_1) \cup U_{\rho_2}(x_2). \end{aligned} \tag{6.28}$$

If x_1, x_2 are values taken by random variables X_1, X_2 as above, we can define

$$(m_1 \hat{\oplus} m_2)(A) = P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cup U_{\rho_2}(X_2(\omega)) = A\}\right), \tag{6.29}$$

$$(bel_1^* \hat{\oplus} bel_2^*)(A) = P\left(\{\omega \in \Omega : \emptyset \neq U_{\rho_1}(X_1(\omega)) \cup U_{\rho_2}(X_2(\omega)) \subset A\}\right), \tag{6.30}$$

$$\begin{aligned} (bel_1 \hat{\oplus} bel_2)(A) = \; & P\left(\{\omega \in \Omega : U_{\rho_1}(X_1(\omega)) \cup U_{\rho_2}(X_2(\omega)) \subset A\} \,/\right. \\ & \left.\{\omega \in \Omega : U_{\rho}(X_1(\omega)) \cup U_{\rho_2}(X_2(\omega)) \neq \emptyset\}\right). \end{aligned} \tag{6.31}$$

As the inclusion

$$U_{\rho_1}(X_1(\omega)) \cap U_{\rho_2}(X_2(\omega)) \subset U_{\rho_1}(X_1(\omega)) \cup U_{\rho_2}(X_2(\omega)) \tag{6.32}$$

trivially holds for each $\omega \in \Omega$, the inequality

$$(bel_1^* \hat{\oplus} bel_2^*)(A) \leq (bel_1^* \oplus bel_2^*)(A) \tag{6.33}$$

is also obvious for each $A \subset S$. For the case of normalized belief functions the situation is not so simple.

Again, as in the case of Dempster combination rule defined above, let us assume that the random variables X_1, X_2 and, consequently, also the set-valued random variables $U_{\rho_1}(X_1(\cdot))$, $U_{\rho_2}(X_2(\cdot))$, are statistically independent, and that the basic space S is finite. Due to the same factorization as in (6.16), also the dual combination rule $\hat{\oplus}$ can be defined in a combinatoric way dual to the formulas for Dempster products. Namely,

$$(m_1 \hat{\oplus} m_2)(A) = \sum_{B,C \subset S,\, B \cup C = A} m_1(B)\, m_2(C), \tag{6.34}$$

$$(bel_1^* \hat{\oplus} bel_2^*)(A) = \sum_{B,C \subset S,\, \emptyset \neq B \cup C \subset A} m_1(B)\, m_2(C), \tag{6.35}$$

$$(bel_1 \hat{\oplus} bel_2)(A) = \frac{\sum_{B,C \subset S,\, \emptyset \neq B \cup C \subset A} m_1(B)\, m_2(C)}{\sum_{B,C \subset S,\, \emptyset \neq B \cup C} m_1(B)\, m_2(C)} \tag{6.36}$$

supposing, in the case of (6.36), that there exist $B \subset S$, $C \subset S$ such that $B \neq \emptyset$ or $C \neq \emptyset$, $m_1(B) > 0$ and $m_2(C) > 0$ hold. Again, the dual combination rule $\hat{\oplus}$ is commutative and associative for b.p.a.'s as well as for both the kinds of belief functions, due to the obvious commutativity and associativity of the set-theoretic operation of union.

As could be expected, the roles of the vacuous b.p.a. m_S and the totally inconsistent b.p.a. $m_\emptyset$ are also dually interchanged so that, for each b.p.a. m, the identities

$$m \hat{\oplus} m_S \equiv m_S \hat{\oplus} m \equiv m_S, \tag{6.37}$$

$$m \hat{\oplus} m_\emptyset \equiv m_\emptyset \hat{\oplus} m \equiv m \tag{6.38}$$

hold. Corresponding relations for belief functions can be easily obtained.

An interpretation behind the dual combination rule can be as follows. Let the original compatibility relation ρ_1 represent the basic undergraduate textbook knowledge of a young and just graduated physician, so that this compatibility relation reflects the typical medical cases and does not take into consideration non-typical and rare examples. On the other side, compatibility relation ρ_2 may describe the knowledge base of an older experienced physician who has met and examined a number of non-typical cases, so that she/he takes into account also the compatibility of some observations and diagnoses not explicitly stated as compatible by undergraduate medical textbooks. E. g., a young physician, having examined a patient, concludes that the patient does not suffer from an inflection disease on the ground of the observation that she/he is not feverish. An older experienced physician knows, however, that in certain, even if very rare cases, an infection disease need not be combined with a fewer, so that the possibility of such a diagnosis cannot be ultimately avoided.

In the next chapter we shall see that the dual combination rule possesses some properties dual to those possessed by the Dempster rule also from the point of view of an explicit numerical criterion measuring the qualities of b.p.a.'s or of corresponding belief functions.

7 Nonspecificity Degrees of Basic Probability Assignments

As we have already defined, in the case of a finite basic space S, basic probability assignment (BPA) on S is a probability distribution on the power-set $\mathcal{P}(S)$ (set of all subsets of S). In this chapter we shall define and investigate the nonspecificity degree of a BPA given by the normalized expected value of the size (cardinality) of subsets of S with respect to the probability distribution defining the BPA in question. This notion enables to express formally and to prove the intuitive feelings of improving one's basic probability assignment and belief function when combining it with another one by the Dempster combination rule. It enables also to define a basic probability assignment which can play the role, at least in certain relations, of the BPA inverse to the original one with respect to the Dempster combination rule, even if we know that such an inverse BPA cannot be defined up to the most trivial case of the vacuous BPA $m_S(m_S(S) = 1)$. Analogous properties of the combination rule dual to the Dempster one will be also briefly investigated.

Definition 7.1. Let m be a BPA on a finite set S. The *(degree of) nonspecificity* $W(m)$ of m is the real number from the unit interval defined by

$$W(m) = \sum\nolimits_{A \subset S} m(A) \, (\|A\| / \|S\|), \tag{7.1}$$

where $\|A\|$ denotes the cardinality, i. e., in our case of finite sets, simply the number of elements of a subset A of S. $\square$

The intuition behind the substantive "nonspecificity" consists in the intuition according to which we take a BPA m_1 as less specific than m_2, if the same probabilities are ascribed to larger sets in the case of m_1 than in the case of m_2. Hence, the (degree of) nonspecificity of m should increase with the average size of focal elements increasing (a subset $A \subset S$ is called a *focal element* of m, or with respect to m, if $m(A) > 0$). In agreement with this intuition we obtain that $W(m_S) = \|S\|/\|S\| = 1$ is the maximal value taken by the function W in the space of BPA's, on the other side, $W(m_\emptyset) = 0$ for the (totally) inconsistent BPA $m_\emptyset$ ascribing all the probability 1 to the empty subset $\emptyset$ of S. In this case, the intuition behind the term "minimal degree of nonspecificity", i.e., the "maximal degree of specificity" does not intuitively correspond to $m_\emptyset$, as $m_\emptyset$ does not contain any specification of the actual value $s_0 \in S$. Moreover, $m_\emptyset$ is the *only* BPA on S for which the nonspecificity takes the zero value, in other words, $m \not\equiv m_\emptyset$ implies that $W(m) > 0$ holds. If m defines a probability distribution on singletons of $\mathcal{P}(S)$, i.e., if $m(A) = 0$ for $A = \emptyset$ and for each $A \subset S$ such that $\|A\| \geq 2$ (consequently, the corresponding belief function bel_m is a probability distribution on $\mathcal{P}(S)$), then obviously $W(m) = 1/\|S\|$, if m is (totally) consistent, i.e., if $m(\emptyset) = 0$, then $W(m) \geq 1/\|S\|$ holds. If m is the uniform probability distribution on $\mathcal{P}(S)$, hence, if $m(A) = 1/2^{\|S\|}$ for each $A \subset S$, then $W(m) \to 1/2$ with $\|S\| \to \infty$. Or, due to the strong law of large numbers (cf. [8] or any textbook on elementary probability theory), for each $\varepsilon > 0$, $\delta > 0$, there exists $n \in \mathcal{N}^+ = \{1, 2, \dots\}$ such that, for all S with $\|S\| \geq n$, the relation

$$(\|\mathcal{P}(S)\|)^{-1} \; \|\{A \subset S : (1/2) - \varepsilon < \|A\|/\|S\| < (1/2) + \varepsilon\}\| > 1 - \delta \tag{7.2}$$

holds, so that, for the uniform probability distribution m on S, the inequalities

$$(1/2) - \varepsilon - \delta < W(m) < (1/2) + \varepsilon + \delta \tag{7.3}$$

are satisfied.

Instead of deducing some more or less interesting properties of nonspecificity degrees for particular BPA's let us focus our attention to the manner in which nonspecificity degrees reflect the model of combination of two or more BPA's or belief functions called Dempster combination rule, introduced and analyzed in the last chapter. Let us recall that for two BPA's m_1, m_2 on a finite set S, their non-normalized Dempster product $m_1 \oplus m_2$ is defined by

$$(m_1 \oplus m_2)(A) = \sum\nolimits_{B, C \subset S,\, B \cap C = A} m_1(B)\, m_2(C) \tag{7.4}$$

for each $A \subset S$, and their normalized product $m_1 \oplus_n m_2$ by

$$\begin{aligned} (m_1 \oplus_n m_2)(A) &= (1 - (m_1 \oplus m_2)(\emptyset))^{-1} (m_1 \oplus m_2)(A) = \\ &= \frac{\sum_{B, C \subset S,\, B \cap C = A} m_1(B)\, m_2(C)}{\sum_{B, C \subset S,\, B \cap C \neq \emptyset} m_1(B)\, m_2(C)} \end{aligned} \tag{7.5}$$

for each $\emptyset \neq A \subset S$, and by $(m_1 \oplus m_2)(\emptyset) = 0$, supposing that $(m_1 \oplus m_2)(\emptyset) < 1$ holds. The expressions for the Dempster products of the corresponding belief functions can be found in the last chapter and we do not recall them here as only BPA's will be investigated throughout this chapter.

Let us introduce and prove the following statement postponing a discussion concerning its sense and importance to an appropriate place below.

Theorem 7.1. Let m_1, m_2 be BPA's on a finite set S. Then

$$W(m_1 \oplus m_2) \leq \min\{W(m_1),\, W(m_2)\}. \tag{7.6}$$

$\square$

Proof. Set, for each $A \subset S$, $\mathcal{C}_A = \{\langle B, C\rangle : B \subset S,\, C \subset S,\, B \cap C = A\}$, so that (7.4) can be written as

$$(m_1 \oplus m_2)\,(A) = \sum\nolimits_{\langle B,C\rangle \in \mathcal{C}_A} m_1(B)\, m_2(C). \tag{7.7}$$

For each $\langle B, C\rangle$, $B \subset S$, $C \subset S$, the inclusion $\langle B, C\rangle \in \mathcal{C}_{B\cap C}$ trivially holds. If $A_1 \neq A_2$ are subsets of S, then $\langle B, C\rangle \in \mathcal{C}_{A_1}$ implies that $B \cap C = A_1$, hence, $B \cap C \neq A_2$ and $\langle B, C\rangle \notin \mathcal{C}_{A_2}$, so that $\mathcal{C}_{A_1} \cap \mathcal{C}_{A_2} = \emptyset$. Hence, $\{\mathcal{C}_A : A \subset S\}$ is a disjoint covering (i. e., partition) of the Cartesian product $\mathcal{P}(S) \times \mathcal{P}(S)$. Definition 7.1 now yields that

$$\begin{aligned} W(m_1 \oplus m_2) &= \sum\nolimits_{A \subset S} (m_1 \oplus m_2)\,(A)\,(\|A\|/\|S\|) \\ &= \sum\nolimits_{A \subset S} (\|A\|/\|S\|) \left(\sum_{\langle B,C\rangle \in \mathcal{C}_A} m_1(B)\, m_2(C) \right) = \\ &= \sum\nolimits_{A \subset S} \left(\sum\nolimits_{\langle B,C\rangle \in \mathcal{C}_A} m_1(B)\, m_2(C)\,(\|A\|/\|S\|) \right). \end{aligned} \tag{7.8}$$

By the same definition,

$$\begin{aligned} W(m_1) &= \sum\nolimits_{B \subset S} m_1(B)\,(\|B\|/\|S\|) = \\ &= \sum\nolimits_{B \subset S} \left[\sum\nolimits_{C \subset S} m_2(C) \right] m_1(B)\,(\|B\|/\|S\|), \end{aligned} \tag{7.9}$$

as $\sum_{C \subset S} m_2(C) = 1$. Hence,

$$\begin{aligned} W(m_1) &= \sum\nolimits_{B \subset S} \left(\sum\nolimits_{C \subset S} m_1(B)\, m_2(C) \right) (\|B\|/\|S\|) = \\ &= \sum\nolimits_{\langle B,C\rangle,\, B \subset S,\, C \subset S} [m_1(B)\, m_2(C)\,(\|B\|/\|S\|)] = \\ &= \sum\nolimits_{A \subset S} \left[\sum\nolimits_{\langle B,C\rangle \in \mathcal{C}_A} m_1(B)\, m_2(C)\,(\|B\|/\|S\|) \right] \geq \\ &\geq \sum\nolimits_{A \subset S} \left[\sum\nolimits_{\langle B,C\rangle \in \mathcal{C}_A} m_1(B)\, m_2(C)\,(\|A\|/\|S\|) \right] = W(m_1 \oplus m_2) \end{aligned} \tag{7.10}$$

by (7.8), as for all $\langle B, C\rangle \in \mathcal{C}_A$, $B \cap C = A$, hence, $B \supset A$, so that $\|B\| \geq \|A\|$ holds. As both the Dempster combination rules (the non-normalized as well as the normalized one) are commutative, we obtain that

$$W(m_1 \oplus m_2) = W(m_2 \oplus m_1) \leq W(m_2) \tag{7.11}$$

holds as well, so that (7.6) immediately follows from (7.10) and (7.11). $\square$

This result agrees with the intuition behind the way in which the Dempster rule combines the partial knowledge and degrees of belief of two subjects. The focal elements of a BPA m are those subsets of the basic space S which can play, with a positive probability, the role of sets of states compatible, up to the subject's knowledge, with the random empirical data being at the disposal of this subject. Dempster rule combines the knowledge of two subjects in such a way that only the states considered as compatible by both the subjects are taken as compatible. Consequently, the probability values ascribed to the focal elements of the original particular BPA's are now ascribed to their subsets, hence, in general, to smaller sets, so that the degree of nonspecificity decreases, informally said, the degree of specificity increases.

Let us note that, in general, the inequality in (7.6) cannot be replaced by equality. Or, take, e.g. $\emptyset \neq A, B \subset S$ such that $A \cap B = \emptyset$, then $W(m_A) = \|A\|/\|S\| > 0$, $W(m_B) = \|B\|/\|S\| > 0$, but $W(m_A \oplus m_B) = W(m_\emptyset) = 0$ (let us recall that m_A is the BPA defined by $m_A(A) = 1$).

Let S be a nonempty finite set taken as the set of all possible states of a system SYST (other interpretations are also possible), just one state s_0 from S being the actual one. For a BPA m on S, the value $m(B)$ can be interpreted as the probability with which such random empirical data were obtained that the subject can be sure that $s_0 \in B \subset S$ holds, no more specification of s_0 to a proper subset of B being possible. Now, the subject obtains a piece of information saying that $s_0 \in A$ holds for a subset $A \subset S$ ($A \neq S$, if we want to avoid the trivial case). The corresponding modification of the subject's original BPA m can be formally defined by the Dempster product $m \oplus m_A$ (in probability theory, this modification is expressed by replacing an original probability measure on S by the conditional probability measure $P(\cdot|A)$). Now, let us consider the situation when the reliability of the additional information defined by the BPA m_A has been put into serious doubts by some next meta-information and the subject wants to turn back from the modified BPA $m \oplus m_A$ to the original BPA m. Let us emphasize here, that it is the *reliability* of the information that $s_0 \in A$ holds, what is taken as doubtful, not the *validity* of this information; this difference is very important in the context of the Dempster–Shafer theory. Hence, a deconditioning operation leading from $m \oplus m_A$ to m would serve as a useful and desirable tool within the framework of Dempster–Shafer theory.

Unfortunately, this problem is unsolvable not only within the framework developed till now, but also within a substantially broader framework defined and investigated in the following chapters. This problem can be transformed into a more general problem how to define, if even possible, an "inversion" to the Dempster combination rule, i.e., how to define an operation $\ominus$ such that $(m_1 \oplus m_2) \ominus m_2 \equiv m_1$ would hold for each BPA's m_1, m_2 on a finite set S.

This generalized problem can be solved, within our framework, if there exists, for each BPA m_2, an "inverse" BPA m_2^{-1} such that, for each BPA m_1, the identity $(m_1 \oplus m_2) \oplus m_2^{-1} \equiv m_1$ holds. Due to the associativity of the Dempster rule this identity implies that $m_2 \oplus m_2^{-1} \equiv m_S$, as $m_1 \oplus m_S \equiv m_S$ holds for each BPA m_1 on S. However, $m_2 \oplus m_2^{-1} \equiv m_S$ implies that $(m_2 \oplus m_2^{-1})(S) = m_2(S)\,m_2^{-1}(S) = m_S = 1$, hence, $m_2(S) = m_2^{-1}(S) = 1$, so that $m_2 \equiv m_2^{-1} \equiv m_S$ follows. So, if $m_2 \not\equiv m_S$, the "inverse" BPA m_2^{-1} does not exist. In particular, m_A^{-1} does not exist, if $A \neq S$. Let us recall that an analogous deconditioning problem is unsolvable also within the framework of probability theory; given a probability distribution P on S and $A \subset S$, there is, in general, no $B \subset S$ such that the conditional probability $(P(\cdot|A)|B)$, i.e.,

$P(\cdot|A\cap B)$ would be identical with P on S. In [40] or [27] we can found a partial solution to the invertibility problem in Dempster–Shafer theory consisting in an appropriate generalization of the notion of BPA (cf. also the next chapters, Part II of this work), however, for the particular case of deconditioning the problem remains unsolvable even within this enriched space of BPA's.

Keeping in mind what we have just quoted, we shall focus our attention, in the rest of this chapter, to a less pretentious task which can be called *quasi-deconditioning*. Instead of an effective operation giving, for each input set $A \subset S$ and input BPA $m \oplus m_A$ on S, as its result the BPA m, we shall seek for a weaker procedure yielding, for each $A \subset S$ and each $m \oplus m_A$, a BPA m^* satisfying the three following demands:

(i) $m^* \oplus m_A \equiv m \oplus m_A$,

(ii) m^* is effectively defined given $m \oplus m_A$ and A,

(iii) m^* takes an extraordinary position (to be specified later) in the set of all BPA's satisfying (i).

Let us note that (ii) avoids, in general, the BPA m itself from the set of possible candidates to m^*.

Definition 7.2. Let S be a finite set, let $A \subset S$, let m be a BPA on S such that $m(B) = 0$ for each $B \subset S$, $B \not\subset A$. The *extension* of m from A to S, denoted by $\mathrm{Ext}(A,S)(m)$, is the BPA m^* on S defined by $m^*(B) = m(B \cap A)$ for each $B \subset S$ such that $B \cap (S-A) = S-A$, otherwise written, $m^*(C \cup (S-A)) = m(C)$ for each $C \subset A$. □

As $\sum_{C\subset A} m^*(C\cup(S-A)) = \sum_{C\subset A} m(C) = 1$, it follows immediately, that $m^*(B) = 0$ for all $B \subset S$ which cannot be written in the form $C \cup (S-A)$, i. e., for all $B \subset S$ such that $B \cap (S-A) \neq S-A$.

For each BPA m on a finite set S and each $A \subset S$, the BPA $m \oplus m_A$ is such that $(m \oplus m_A)(B) = 0$ for each $B \subset S$, $B \not\subset A$, so that, setting $m^* = \mathrm{Ext}(A,S)(m \oplus m_A)$, m^* is correctly and effectively defined and satisfies the demand (ii) above. Moreover, given $B \subset S$,

$$\begin{aligned}(m^* \oplus m_A)(B) &= \textstyle\sum_{\langle C,D\rangle, C\cap D=B} m^*(C)\, m_A(B) = \sum_{C\subset S,\, C\cap A=B} m^*(C) = \\ &= m^*(B \cup (S-A)) = (m \oplus m_A)(B),\end{aligned} \tag{7.12}$$

as $B \cup (S-A)$ is the only subset C of S such that $C \cap A = B$ and, at the same time, $m^*(C)$ can be positive. Hence, m^* meets the demand (i) above. Set, for each BPA m on S and each $A \subset S$,

$$\mathrm{Red}(S,A)(m) = \{m_1 : m_1 \text{ is a BPA on } S,\ m_1 \oplus m_A \equiv m \oplus m_A\}\,. \tag{7.13}$$

The following assertion proves that, and in which sense, $\mathrm{Ext}(A,S)(m \oplus m_1)$ meets also the demand (iii) above.

Theorem 7.2. For each BPA m on a finite set S,

$$W(\mathrm{Ext}(A,S)\,(m\oplus m_A)) = \max\left\{W(m_1) : m_1 \in \mathrm{Red}(S,A)\,(m)\right\}. \tag{7.14}$$

$\square$

Hence, $\mathrm{Ext}(A,S)\,(m\oplus m_A)$ can be called the *quasi-solution to the deconditioning problem* based on the principle of maximum nonspecificity.

Proof. Let $m_1 \in \mathrm{Red}(S,A)\,(m)$, let $m^* = \mathrm{Ext}(A,S)\,(m\oplus m_A)$. By definition,

$$\begin{aligned}(m_1\oplus m_A)\,(B) &= \textstyle\sum_{C\subset S,\,C\cap A=B} m_1(C) = \sum_{X\subset S-A} m_1(B\cup X), \\ (m^*\oplus m_A)\,(B) &= \textstyle\sum_{C\subset S,\,C\cap A=B} m^*(C) = \sum_{X\subset S-A} m^*(B\cup X) = \\ &= m^*(B\cup(S-A)) = m(A),\end{aligned} \tag{7.15}$$

as $B\cup(S-A)$ is the only set $C\subset S$ such that $C\cap(S-A)=S-A$ and $C\cap A=B$, and only for such C the value $m^*(C)$ can be positive. We also assume that the summation over the empty set of items yield zero, as is the case for all $B\subset S$, $B\not\subset A$. For $m_1\in\mathrm{Red}(S,A)\,(m)$ the equalities $(m_1\oplus m_A)\,(B) = (m^*\oplus m_A)\,(B) = \sum_{C\subset S-A} m_1(X) = m^*(B\cup(S-A)) = m(B)$ hold. Let $\mathcal{C}_A(B) = \{C\subset S : C\cap A = B\}$. Then, for $B\subset A$,

$$\begin{aligned}&\textstyle\sum_{C\in\mathcal{C}_A(B)} m_1(C)\,(\|C\|/\|S\|) = \sum_{X\subset S-A} m_1(B\cup X)\,(\|B\cup X\|/\|S\|) = \\ =\ &\textstyle\sum_{X\subset S-A} m_1(B\cup X)\ (\|B\|+\|X\|)/\|S\|)) \leq \\ \leq\ &\textstyle\sum_{X\subset S-A} m_1(B\cup X)\ ((\|B\|+\|S-A\|)/\|S\|),\end{aligned} \tag{7.16}$$

as $B\subset A$ implies that the sets B and $X\subset S-A$ are disjoint, the last inclusion implies also that $\|X\|\leq\|S-A\|$. Hence,

$$\begin{aligned}&\textstyle\sum_{C\in\mathcal{C}_A(B)} m_1(C)\,(\|C\|/\|S\|) \leq \\ \leq\ &\left(\textstyle\sum_{X\subset S-A} m_1(B\cup X)\right)((\|B\|+\|S-A\|)/\|S\|) = \\ =\ &m(B)\ ((\|B\|+\|S-A\|)/\|S\|) = \\ =\ &m^*(B\cup(S-A))\,(\|B\cup(S-A)\|/\|S\|) = \\ =\ &\textstyle\sum_{C\in\mathcal{C}_A(B)} m^*(C)\,(\|C\|/\|S\|).\end{aligned} \tag{7.17}$$

As we have already proved (cf. the proof of Theorem 7.1 above), the system $\{\mathcal{C}_A(B) : B\subset A\}$ is a disjoint covering of the power-set $\mathcal{P}(S)$, i.e., $B_1, B_2\subset A$, $B_1\neq B_2$ implies that $\mathcal{C}_A(B_1)\cap\mathcal{C}_A(B_2)=\emptyset$, and for each $C\subset S$, $C\in\mathcal{C}_A(A\cap C)$ holds. Hence, for $m_1\in\mathrm{Red}(S,A)\,(m)$,

$$\begin{aligned}W(m_1) &= \textstyle\sum_{D\subset S} m_1(D)\,(\|D\|/\|S\|) = \\ &= \textstyle\sum_{B\subset A}\left(\sum_{C\in\mathcal{C}_A(B)} m_1(C)\,(\|C\|/\|S\|)\right) \leq \\ &\leq \textstyle\sum_{B\subset A}\left(\sum_{C\in\mathcal{C}_A(B)} m^*(C)\,(\|C\|/\|S\|)\right) = \\ &= \textstyle\sum_{D\subset S} m^*(D)\,(\|D\|/\|S\|) = W(m^*) = W\left(\mathrm{Ext}(A,S)\,(m\oplus m_A)\right).\end{aligned} \tag{7.18}$$

As the set $\mathrm{Red}(S, A)\,(m)$ of BPA's is finite, the assertion is proved. $\square$

An assertion dual to Theorem 7.1 in the sense that Dempster combination rule is replaced by its dual version $\hat{\oplus}$ introduced in the last chapter, can be also stated and easily proved.

Theorem 7.3. Let m_1, m_2 be BPA's on a finite set S. Then

$$W(m_1 \hat{\oplus} m_2) \geq \max\{W(m_1),\, W(m_2)\}. \tag{7.19}$$

$\square$

Proof. Also the proof is dual to that of Theorem 7.1. Set, for each $A \subset S$, $\mathcal{D}_A = \{\langle B, C\rangle : B \subset S,\, C \subset S,\, B \cup C = A\}$, so that, after the factorization analogous to that used for Dempster rule,

$$(m_1 \hat{\oplus} m_2)\,(A) = \textstyle\sum_{\langle B,C\rangle \in \mathcal{D}_A} m_1(B)\, m_2(C). \tag{7.20}$$

For each $\langle B, C\rangle$, $B \subset S$, $C \subset S$, the relation $\langle B, C\rangle \in \mathcal{D}_{B\cup C}$ trivially holds. If $A_1 \neq A_2$ are different subsets of S, then $\langle B, C\rangle \in \mathcal{D}_{A_1}$ yields that $B \cup C = A_1$, hence, $B \cup C \neq A_2$ and $\langle B, C\rangle \notin \mathcal{D}_{A_2}$, consequently, $\mathcal{D}_{A_1} \cap \mathcal{D}_{A_2} = \emptyset$. Hence, $\{\mathcal{D}_A : A \subset S\}$ is a disjoint covering, i. e., a partition, of the Cartesian product $\mathcal{P}(S) \times \mathcal{P}(S)$. Definition 7.1 yields that

$$\begin{aligned} W(m_1 \hat{\oplus} m_2) &= \textstyle\sum_{A \subset S} (m_1 \hat{\oplus} m_2)\,(A)\,(\|A\|/\|S\|) = \\ &= \textstyle\sum_{A \subset S} (\|A\|/\|S\|) \sum_{\langle B,C\rangle \in \mathcal{D}_A} m_1(B)\, m_2(C) = \\ &= \textstyle\sum_{A \subset S} \left(\sum_{\langle B,C\rangle \in \mathcal{D}_A} m_1(B)\, m_2(C)\right) (\|A\|/\|S\|). \end{aligned} \tag{7.21}$$

By the same definition,

$$\begin{aligned} W(m_1) &= \sum_{B \subset S} m_1(B)\,(\|B\|/\|S\|) = \\ &= \textstyle\sum_{B \subset S} \left[\displaystyle\sum_{C \subset S} m_2(C)\right] m_1(B)\,(\|B\|/\|S\|), \end{aligned} \tag{7.22}$$

as $\sum_{C \subset S} m_2(C) = 1$. Hence,

$$\begin{aligned} W(m_1) &= \textstyle\sum_{B \subset S} \left(\sum_{C \subset S} m_1(B)\, m_2(C)\right) (\|B\|/\|S\|) = \\ &= \textstyle\sum_{\langle B,C\rangle,\, B \subset S,\, C \subset S} [m_1(B)\, m_2(C)\,(\|B\|/\|S\|)] = \\ &= \textstyle\sum_{A \subset S} \left[\sum_{\langle B,C\rangle \in \mathcal{D}_A} m_1(B)\, m_2(C)\right] (\|B\|/\|S\|) \leq \\ &\leq \textstyle\sum_{A \subset S} \left[\sum_{\langle B,C\rangle \in \mathcal{D}_A} m_1(B)\, m_2(C)\right] (\|A\|/\|S\|) = W(m_1 \hat{\oplus} m_2) \end{aligned} \tag{7.23}$$

by (7.21), as for all $\langle B, C\rangle \in \mathcal{D}_A$, $B \cup C = A$, hence, $B \subset A$. so that $\|B\| \leq \|A\|$. As the dual combination rule $\hat{\oplus}$ is commutative, we obtain that

$$W(m_1 \hat{\oplus} m_2) = W(m_2 \hat{\oplus} m_1) \geq W(m_2) \tag{7.24}$$

also holds, so that (7.19) immediately follows from (7.23) and (7.24). $\square$

Going on with our investigation of the dual combination rule $\hat{\oplus}$, we shall consider the dual version of the conditioning operation, i.e., the dual product $m \hat{\oplus} m_A$ for a subset $A \subset S$ ($A \neq S$, if we want to avoid the trivial case already investigated above). Let us recall that $m_A(A) = 1$, i.e., $m_A(B) = 0$ for each $B \subset S$, $B \neq A$. An easy calculation yields that

$$\begin{aligned} (m \hat{\oplus} m_A)(B) &= \textstyle\sum_{C \subset S,\, D \subset S,\, C \cup D = B} m(C)\, m_A(D) = \\ &= \textstyle\sum_{C \subset S,\, C \cup A = B} m(B), \end{aligned} \tag{7.25}$$

so that $(m \hat{\oplus} m_A)(B) = 0$, if $B \not\supset A$. Let $B_1 \supset A$, $B_2 \supset A$, let $B_1 \neq B_2$. Then there exists $s \in B_1 - B_2$ or $s \in B_2 - B_1$. If $s \in B_1 - B_2$, then $s \in B_1$ and $s \notin B_2$, so that $s \notin A$, as $B \subset B_2$. Consequently, $s \in B_1 - A$, $s \notin B_2 - A$, and $B_1 - A \neq B_2 - A$. The case when $s \in B_2 - B_1$ is processed analogously. Hence, $B_1 \supset A$, $B_2 \supset A$, and $B_1 \neq B_2$ implies that $B_1 - A \neq B_2 - A$, so that there exists a $1-1$ mapping between the systems $\{B \subset S : B \supset A\}$ and $\{B : B \subset S - A\}$ of subsets of S. Consequently, given a BPA m on S and $A \subset S$, we can define the BPA m^{**} on S in such a way that all the value $(m \hat{\oplus} m_A)(B)$ for $B \supset A$ is shifted to $B - A$. In symbols,

$$\begin{aligned} m^{**}(B) &= (m \hat{\oplus} m_A)(B \cup A), \text{ if } B \subset S - A, \\ m^{**}(B) &= 0 \text{ otherwise, i.e., if } B \not\subset S - A. \end{aligned} \tag{7.26}$$

An easy calculation yields that, for each $B \subset S$,

$$\begin{aligned} (m^{**} \hat{\oplus} m_A)(B) &= \textstyle\sum_{C \subset S,\, D \subset S,\, C \cup D = B} m^{**}(C)\, m_A(D) = \\ &= \textstyle\sum_{C \subset S,\, C \cup A = B} m^{**}(C). \end{aligned} \tag{7.27}$$

The only set $C \subset S - A$ (and only for those sets m^{**} may take positive values) such that $C \cup A$=B is the set $B - A$, if $B \supset A$; if $B \not\supset A$, no such C exists. Hence,

$$(m^{**} \hat{\oplus} m_A)(B) = 0, \text{ if } B \not\supset A, \tag{7.28}$$

$$\begin{aligned} (m^{**} \hat{\oplus} m_A)(B) &= m^{**}(B - A) = (m \hat{\oplus} m_A)((B - A) \cup A) = \\ &= (m \hat{\oplus} m_A)(B), \end{aligned} \tag{7.29}$$

if $B \supset A$, so that the equivalence $m^{**} \hat{\oplus} m_A \equiv m \hat{\oplus} m_A$ is valid. The only we have to prove in order to show that the duality between m^{**} and * defined above is complete, is the inequality $W(m^{**}) \leq W(m_1)$ for each BPA m_1 such that $m_1 \hat{\oplus} m_A \equiv m \hat{\oplus} m_A$ holds.

Theorem 7.4. Let m be a BPA on a finite set S, let A be a subset of S, let m^{**} be the BPA on S defined by (7.26), let m_1 be any BPA on S such that $m_1 \hat{\oplus} m_A \equiv m \hat{\oplus} m_A$. Then $W(m^{**}) \leq W(m_1)$ holds. $\square$

Proof. Given $B \subset S$, set $\mathcal{D}_A(B) = \{C \subset S : C \cup A = B\}$. There is a $1-1$ mapping between $\mathcal{D}_A(B)$ and the set $\mathcal{P}(A)$ of all subsets of A so that $\mathcal{D}_A(B)$ can be written as $\{C : C = X \cup (B-A),\ X \subset A\}$. Let us consider the case when $B \supset A$. An easy calculation yields that

$$\begin{aligned}
& \sum\nolimits_{C \in \mathcal{D}_A(B)} m_1(C)\,(\|C\|/\|S\|) = \\
= \ & \sum\nolimits_{X \subset A} m_1(X \cup (B-A))\,(\|X \cup (B-A)\|/\|S\|) = \\
= \ & \sum\nolimits_{X \subset A} m_1(X \cup (B-A))\,((\|X\| + \|B-A\|)/\|S\|) \geq \\
\geq \ & \sum\nolimits_{X \subset A} m_1(X \cup (B-A))\,(\|B-A\|/\|S\|) = \\
= \ & \left(\sum\nolimits_{X \subset A} m_1(X \cup (B-A))\right)(\|B-A\|/\|S\|) = \\
= \ & \left(\sum\nolimits_{C \in \mathcal{D}_A(B)} m_1(C)\right)(\|B-A\|/\|S\|) = \\
= \ & \left(\sum\nolimits_{C \subset S,\, C \cup A = B} m_1(C)\right)(\|B-A\|/\|S\|) = \\
= \ & (m \hat{\oplus} m_A)(B)\,(\|B-A\|/\|S\|) = \\
= \ & m^{**}(B-A)\,(\|B-A\|/\|S\|) = \\
= \ & \sum\nolimits_{C \in \mathcal{D}_A(B)} m^{**}(C)\,(\|C\|/\|S\|),
\end{aligned} \tag{7.30}$$

as $B - A$ is the only set C such that $C \subset S - A$ and $C \cup A = B$, and just for those subsets of S the value $m^{**}(C)$ may be positive. So, for each $B \supset A$, the inequality

$$\sum\nolimits_{C \subset \mathcal{D}_A(B)} m_1(C)\,(\|C\|/\|S\|) \geq \sum\nolimits_{C \in \mathcal{D}_A(B)} m^{**}(C)\,(\|C\|/\|S\|) \tag{7.31}$$

holds. For each $C \subset S$, $C \in \mathcal{D}_A(C \cup A)$ and $C \cup A \supset A$ trivially hold, moreover, if $B_1 \supset A$, $B_2 \supset A$ and $B_1 \neq B_2$ hold, then $C \in \mathcal{D}_A(B_1)$ yields that $C \cup A = B_1$, hence, $C \cup A \neq B_2$, so that $C \notin \mathcal{D}_A(B_2)$, consequently, $\mathcal{D}_A(B_1) \cap \mathcal{D}_A(B_2) = \emptyset$. So, $\{\mathcal{D}_A(B) : B \supset A\}$ is a disjoint covering, i. e., partition, of the power-set $\mathcal{P}(S)$. Consequently,

$$\begin{aligned}
W(m_1) \ &= \ \sum\nolimits_{D \subset S} m_1(D)\,(\|D\|/\|S\|) = \\
&= \ \sum\nolimits_{B \subset A} \left(\sum\nolimits_{C \in \mathcal{D}_A(B)} m_1(C)\,(\|C\|/\|S\|)\right) \geq \\
&\geq \ \sum\nolimits_{B \subset A} \left(\sum\nolimits_{C \in \mathcal{D}_A(B)} m^{**}(C)\,(\|C\|/\|S\|)\right) = \\
&= \ \sum\nolimits_{D \subset S} m^{**}(D)\,(\|D\|/\|S\|) = W(m^{**}).
\end{aligned} \tag{7.32}$$

The assertion is proved. $\square$

The list of references introduced below contains not only the items explicitly referred to in the text above, but also some books and papers which have served as perhaps not immediate, but in no case less important sources of inspiration and motivation for the ideas presented and results achieved in this work. Some of these references will be explicitly referred to in the forthcoming Part II of this report.

Bibliography

[1] B. Balcar, P. Štěpánek: Teorie množin (Set Theory – in Czech). Academia, Prague, 1986.

[2] D. Blackwell, M. A. Girshick: Theory of Games and Statistical Decisions. John Wiley, New York, 1954.

[3] W. Brachinger, J. Kohlas: Argumentation systems and evidence theory. In: Advances in Intelligent Computing – IPMU 94 (Bouchon–Meunier and Yager, Eds.). Lecture Notes in Computer Science 945, Springer Verlag, Berlin – Heidelberg – New York, 1994, pp. 41–50.

[4] G. Choquet: Theory of capacities. Annales de l'Institut Fourier 5 (1953–4), pp. 131–295.

[5] A. P. Dempster: Upper and lower probabilities induced by a multi-valued mapping. Annals of Mathematical Statistics 38 (1967), no. 2, pp. 325–339.

[6] D. Dubois, H. Prade: Théorie de Possibilités – Applications à le Représentation de Connaissances en Informatique. Mason, Paris, 1985.

[7] R. Faure, E. Heurgon: Structures Ordonnées et Algèbres de Boole. Gauthier–Villars, Paris, 1971.

[8] W. Feller: An Introduction to Probability Theory and its Applications I, II. J. Wiley and Sons, New York – London, Vol. I 1957, Vol. II 1965.

[9] T. L. Fine: Theories of Probability – an Examination of Foundations. Academic Press, New York, 1973.

[10] A. A. Fraenkel, Y. Bar–Hillel: Foundations of Set Theory. North Holland, Amsterdam, 1959.

[11] B. V. Gnedenko: Kurs tĕoriji vĕrojatnostĕj (A Course in Probability Theory – in Russian). Nauka, Moscow, 1965.

[12] J. A. Goguen: L–fuzzy sets. J. of Math. Analysis and Applications 18 (1967), pp. 145–174.

[13] S. Guiasu: Information Theory with Applications. McGraw–Hill, New York, 1977.

[14] P. Hájek, T. Havránek, R. Jiroušek: Uncertain Information Processing in Expert Systems, CRC Press, 1992.

[15] P. R. Halmos: Measure Theory. D. van Nostrand, New York – Toronto – London, 1950.

[16] D. Harmanec: Uncertainty in Dempster–Shafer Theory. Ph. D.–dissertation, Univ. of New York, Binghamton, 1996.

[17] S. J. Henkind, M. C. Harrison: An analysis of four uncertainty calculi. IEEE Transactions on Systems, Man, and Cybernetics 18 (1988), no. 5, pp. 700–714.

[18] R. Kennes: Evidential reasoning in a categorial perspective: Conjunction and disjunction of belief functions. In: Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence (B. D. d'Ambrosio, P. Smets, P. P. Bonissone, Eds.). Morgan Kaufman, San Matteo, California, 1991, pp. 174–181.

[19] G. J. Klir, T. A. Folger: Fuzzy Sets, Uncertainty, and Information. Prentice–Hall, Englewood Cliffs,, New Jersey, 1988.

[20] J. Kohlas: Mathematical foundations of evidence theory. In: Mathematical Models for Handling Partial Knowledge in Artificial Intelligence (G. Coletti et al, Eds.). Plenum Press, New York, 1994, pp. 31–64.

[21] J. Kohlas, P. A. Monney: A Mathematical Theory of Hints – An Approach to the Dempster–Shafer Theory of Evidence. Lecture Notes in Economics and Mathematical Systems, Springer Verlag, New York, 1995.

[22] A. N. Kolmogorov: Osnovnyje ponjatija teoriji věrojatnostěj (Basic Notions of Probability Theory – in Russian). Nauka, Moscow, 1974.

[23] I. Kramosil: Strong law of large numbers for set-valued random variables. In: WUPES 94 – Proceedings of the 3rd Workshop on Uncertainty Processing in Expert Systems, University of Economics, Prague, 1994, pp. 122–142.

[24] I. Kramosil: Believeability functions induced by partial generalized compatibility relations. In: Fuzzy Sets 94 – Proceedings of the Second International Conference on Fuzzy Sets Theory and its Applications (R. Mesiar and V. Novák, Eds.). Tatra Mountains Mathematical Publications vol. 6 (1995), pp. 103–116.

[25] I. Kramosil: Nonstandard approach to possibility measures. Int. J. of Uncertainty, Fuzziness and Knowledge–Based Systems 4 (1996), no. 3, pp. 275–301.

[26] I. Kramosil: Some remarks on philosophical aspects of the laws of large numbers. In: LOGICA 95 – Proceedings of the Symposium (T. Childers, P. Kolář and V. Svoboda, Eds.), Filosofia, Prague, 1996, pp. 179–198.

[27] I. Kramosil: Belief functions with nonstandard values. In: ECSQARU 97 – Proceedings of the Conference (D. Gabbay, R. Kruse, A. Nonnengart, H. J. Ohlbach, Eds.). Lecture Notes in Artificial Intelligence 1244, Springer Verlag, Berlin – Heidelberg – New York, 1997, pp. 380–391.

[28] I. Kramosil: Probabilistic analysis of Dempster combination rule. Submitted for publication to LOGICA 97 – Proceedings of the Symposium.

[29] H. E. Kyburg: Probability and Inductive Logic. Collier–Macmillan, London, 1970.

[30] E. L. Lehman: Testing Statistical Hypotheses. John Wiley, New York, 1947.

[31] M. Loève: Probability Theory. D. van Nonstrand, New York, 1960.

[32] V. Novák: Fuzzy Sets and their Applications. Adam Hilger, Bristol–Philadelphia, 1989.

[33] J. Pearl: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Matteo, California, 1988.

[34] A. Rényi: Probability Theory. North Holland, Amsterdam, 1970.

[35] G. Shafer: A Mathematical Theory of Evidence. Princeton Univ. Press, Princeton, 1976.

[36] G. Shaffer: Allocations of probability. The Annals of Probability 7 (1979), no. 5, pp. 827–839.

[37] R. Sikorski: Boolean Algebras. Springer Verlag, Berlin – Göttingen – Heidelberg, 1960.

[38] Ph. Smets: The nature of the unnormalized beliefs encountered in the transferable belief model. In: Uncertainty in Artificial Intelligence 92 (D. Dubois, M. P. Wellman, B. d'Ambrosio, Ph. Smets, Eds.). Morgan Kaufmann, San Matteo, California, 1992, pp. 292–297.

[39] Ph. Smets: The representation of quantified belief by the transferable belief model. Technical Report no TR/IRIDIA/94–19.1, Institut de Récherches Interdisciplinaires et de Dévelopments en Intelligence Artificielle, Université Libre de Bruxelles, 1994, 49 pp.

[40] Ph. Smets: The canonical decomposition of a weighted belief. In: IJCAI 95 – Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montréal, August 20–25, 1995, Vol. 2, pp. 1896–1901.

[41] J. Štěpán: Teorie pravděpodobnosti (Probability Theory – in Czech). Academia, Prague, 1987.

[42] J. Vejnarová: A few remarks on measures of uncertainty in Dempster–Shafer theory. In: Proceedings of Workshop on Uncertainty in Expert Systems. Czechoslovak Academy of Sciences, Prague, 1991, pages not numbered.

[43] P. Vopěnka, P. Hájek: The Theory of Semisets. Academia, Prague, 1972.

[44] P. Vopěnka: Rozpravy s geometrií I (Discussions with Geometry – in Czech). Panorama, Prague, 1989.

[45] J. Štěpán: Teorie pravděpodobnosti (Probability Theory – in Czech). Academia, Prague, 1987.

[46] R. A. Yager: Entropy and specificity in a mathematical theory of evidence. International Journal of General Systems 9 (1983), no. 4, pp. 249–260.

Contents