Tatra Mt. Math. Publ. **16** (1999), 339–357

TATRA MOUNTAINS
Mathematical Publications

MONTE-CARLO ESTIMATIONS FOR BELIEF FUNCTIONS

IVAN KRAMOSIL

ABSTRACT. Belief functions serve as the main numerical characteristics of the degrees of uncertainty in the so called Dempster-Shafer theory. Even if belief functions do not possess, in general, the properties of probability measures, particular values of belief functions can be defined by probabilities with which an appropriately defined generalized set-valued random variable takes values satisfying certain quite natural properties. Consequently, the values of belief functions can be estimated, using the most elementary principles of Monte-Carlo methods, by relative frequencies of successes in series of statistically independent and identically distributed random samples. An algorithm based on this idea is proposed and its asymptotical properties as well as its qualitative characteristics in non-asymptotic cases are stated and proved.

1. Degrees of belief and their probabilistic interpretation

There are already numerous references dealing with the so called Dempster-Shafer (D.–S.) theory in general and with belief functions as quantitative characteristics (degrees) of uncertainty in particular. In these references, this theory is investigated, developed and discussed from philosophical, methodological, mathematical, theoretical and practical points of view. It is not the aim of this paper to enrich these discussions, but rather to investigate a special problem concerning the computability of a belief functions in actual situations, namely in those connected with statistical data acquisition and processing. It is why we choose a way as short as possible to arrive at the notion of a belief function.

The usual definition of a belief function, given for the case of finite basic space S in a combinatoric way, reads as follows. Let S be a finite nonempty set, let m be a probability distribution on the (obviously finite) power-set $\mathscr{P}(S)$ of all subsets of the set S, i.e., $m\colon \mathscr{P}(S) \to \langle 0,1\rangle$ is such that $\sum_{A\subset S} m(A) = 1$.

2000 Mathematics Subject Classification: 68T30, 60B99.
Keywords: uncertainty quantification and processing, belief function, compatibility relation, probability measure, strong law of large numbers, Monte-Carlo methods.

This work was sponsored by the grant no. A 1030803 of the GA AS CR.

Then m is called a *basic probability assignment* over the basic space S and the *belief function* bel_m induced by the b.p.a. m is the mapping defined, for each $A \subset S$, by

$$\mathrm{bel}_m(A) = \big(1 - m(\emptyset)\big)^{-1} \sum_{\emptyset \neq B \subset A} m(B) \tag{1.1}$$

supposing that $m(\emptyset) < 1$ holds; if $m(\emptyset) = 1$, bel_m is not defined. In what follows, we shall mostly assume that $m(\emptyset) = 0$, so that (1.1) reduces to

$$\mathrm{bel}_m(A) = \sum_{B \subset A} m(B) \tag{1.2}$$

and the difference between the so called *normalized* and *non-normalized* version of a belief function disappears.

The following interpretation enables to define the value $\mathrm{bel}_m(A)$ directly as the probability of certain random event, what will be of key importance in the sequel.

Let us consider a nonempty finite set S of possible internal states of a system SYST, just one $s_0 \in S$ being the actual state, let E be a nonempty and perhaps many-dimensional set of possible empirical values (observations, results of experiments, etc.) concerning SYST and its environment. Both the actual internal state s_0 and the observed empirical value $x_0 \in E$ are supposed to be of random character, formally described by random variables σ and X, both defined on a fixed abstract probability space $\langle \Omega, \mathscr{A}, P \rangle$ and taking their values in a measurable space $\langle S, \mathscr{S} \rangle$ in the case of the random variable σ, and in a measure space $\langle E, \mathscr{E} \rangle$ in the case of the random variable X. Here $\mathscr{S}$ is a nonempty σ-field of subsets of S and we shall suppose, for S finite, that $\mathscr{S} = \mathscr{P}(S)$, and $\mathscr{E}$ is a nonempty σ-field of subsets of E. The subject's (observer's; user's) a priori knowledge concerning the system in question and its environment is described by the so called *compatibility relation* $\rho\colon E \times S \to \{0, 1\}$ (or $\rho \subset E \times S$, if this notation is preferred). For each $s \in S$ and $x \in E$, $\rho(x, s) = 0$ iff the subject knows, or is able to deduce within the framework of her/his deductive abilities, that if x was observed, then s cannot be the actual internal state of SYST. Consequently, $\rho(x, s) = 1$ iff the subject cannot avoid that SYST is situated in the internal state s having observed the empirical value $x \in E$, in other terms said, the values s and x are *compatible.* We shall suppose that the compatibility relation ρ is consistent in the sense that $\rho\big(X(\omega), \sigma(\omega)\big)$ holds for each $\omega \in \Omega$, i.e., each observation is compatible (at least) with the actual internal state of SYST under which it was observed.

Denoting, for $x \in E$, by $U_\rho(x) = \big\{s \in S\colon \rho(x, s) = 1\big\}$ the set of all states compatible with the empirical value x we can easily see that the composed mapping $U_\rho\big(X(\cdot)\big)$ takes Ω into $\mathscr{P}(S)$. If

$$\Big\{\big\{\omega \in \Omega\colon U_\rho\big(X(\omega)\big) = A\big\}\colon A \subset S\Big\} \subset \mathscr{A} \tag{1.3}$$

holds, then $U_\rho(X(\cdot))\colon \Omega \to \mathscr{P}(S)$ can be taken as a (generalized) random variable defined on the probability space $\langle \Omega, \mathscr{A}, P\rangle$ and taking its values in the measurable space $\langle \mathscr{P}(S), \mathscr{P}(\mathscr{P}(S))\rangle$. Set, for each $A \subset S$,

$$m(A) = m_{\rho,X}(A) = P\Big(\big\{\omega \in \Omega\colon U_\rho(X(\omega)) = A\big\}\Big). \tag{1.4}$$

If S is finite, then $m_{\rho,X}$ is obviously a basic probability assignment over S and (1.1) can be written as

$$\mathrm{bel}_{m_\rho,X}(A) = P\Big(\big\{\omega \in \Omega\colon U_\rho(X(\omega)) \subset A\big\}\Big), \tag{1.5}$$

as the consistence condition implies that $U_\rho(X(\omega)) \neq \emptyset$ for each $\omega \in \Omega$, hence, $m_{\rho,X}(\emptyset) = 0$.

Taken as a mapping from $\mathscr{P}(S)$ into $\langle 0, 1\rangle$, a belief function does not meet, in general, the demands imposed to a probability measure (as a matter of fact, $\mathrm{bel}_{m_\rho,X}$ is additive iff $\mathrm{bel}_{m_\rho,X}(A) = 0$ for all $A \subset S$ such that $\mathrm{card}(A) \neq 1$). However, as the values of belief functions can be defined, due to (1.5), by values of probability measures ascribed to certain random events, they can be also estimated by the well-known Monte–Carlo methods (or probabilistic algorithms, in other terms), based on the simple idea to take arithmetical means, defined by a large enough finite number of stochastically independent and identically distributed (i.i.d.) random samples, as reasonable and good estimations of the corresponding expected values.

2. Elementary preliminaries – strong law of large numbers and Chebyshev inequality

The soundness of the basic idea of Monte–Carlo method, as described in the last paragraph of the last section, is based on the well-known *strong law of large numbers* (SLLN). For the sake of reader's convenience let us present this statement here in the formulation as simple as sufficient for our purposes. The corresponding proof and other details can be found in any textbook of the elementary probability theory.

Fact 2.1. (Strong law of large numbers). *Let $\langle \Omega, \mathscr{A}, P\rangle$ be a probability space, let X be a real-valued random variable defined on this probability space, i.e., let X be a measurable mapping taking $\langle \Omega, \mathscr{A}, P\rangle$ into the Borel line $\langle R, \mathscr{B}\rangle$, let the expected value $EX = \int_{-\infty}^{\infty} X \,\mathrm{d}P$ as well as the dispersion $D^2X = E(X - EX)^2$ be finite. Let $X_1, X_2, \ldots$ be an infinite sequence of stochastically (statistically)* *independent random variables defined on* $\langle \Omega, \mathscr{A}, P\rangle$, *each of them possessing the same probability distribution as* X. *Then*

$$P\left(\left\{\omega \in \Omega\colon \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} X_i(\omega) = EX\right\}\right) = 1\,. \tag{2.1}$$

□

By words, the average value taken from an increasing number of independent and identically distributed samples tends with the probability one (almost surely) to the expected value EX (obviously identical with EX_i for all $i = 1, 2, \ldots$). The assertion (2.1) can be improved (generalized) by appropriate weakenings of the conditions imposed to the random variables $X, X_1, X_2, \ldots$, cf. textbooks and monographs on probability theory, but we shall neither need nor use these generalizations in what follows. A formulation of strong law of large numbers for non-numerical, namely set-valued random variables was proposed and investigated in [5], and it could be perhaps of interest in the context of our further explanation, but we shall not use it as immediate preliminaries.

Because of its limit character, SLLN is rather useless in practical cases, unless combined with a quantitative evaluation of the degree in which the average value $n^{-1} \sum_{i=1}^{n} X_i(\omega)$ approximates the expected value EX. Such an evaluation can be deduced from the following statement.

Fact 2.2. (Chebyshev inequality). *Let* X *be a real-valued random variable defined on a probability space* $\langle \Omega, \mathscr{A}, P\rangle$, *let* EX *and* D^2X *be finite. Then, for any real* $\varepsilon > 0$, *the inequality*

$$P\left(\left\{\omega \in \Omega\colon \left|X(\omega) - EX\right| \geq \varepsilon\right\}\right) \leq (D^2X)/\varepsilon^2 \tag{2.2}$$

holds. □

Let $X_1, X_2, \ldots$ be an i.i.d. sequence as in Fact 2.1. An easy calculation yields that, for each $n = 1, 2, \ldots$,

$$D^2\left(n^{-1} \sum_{i=1}^{n} X_i\right) = n^{-1}\, D^2X\,, \tag{2.3}$$

so that (2.2) yields that

$$P\left(\left\{\omega \in \Omega\colon \left|n^{-1} \sum_{i=1}^{n} X_i(\omega) - EX\right| \geq \varepsilon\right\}\right) \leq (D^2X)/n\varepsilon^2\,. \tag{2.4}$$

Hence, for each $\varepsilon > 0$, $\delta > 0$ there exists $n_0(\varepsilon, \delta)$ such that, for all $n \geq n_0(\varepsilon, \delta)$, the inequality

$$P\left(\left\{\omega \in \Omega\colon \left|n^{-1} \sum_{i=1}^{n} X_i(\omega) - EX\right| \geq \varepsilon\right\}\right) \leq \delta \tag{2.5}$$

holds, and it is a matter of elementary technical routine to obtain an explicit expression for n_0 as a function of ε, δ, and D^2X. For the special case of binary random variables X, i.e., for the case when $X(\omega) = 0$ or $X(\omega) = 1$ for all $\omega \in \Omega$, an expression similar to (2.4) but independent of D^2X can be easily obtained. Or, denoting by p the probability $P(\{\omega \in \Omega\colon X(\omega) = 1\})$, so that $P(\{\omega \in \Omega\colon X(\omega) = 0\}) = 1 - p$, we easily obtain that $D^2X = p(1-p) \leq 1/4$ holds for each $p \in \langle 0, 1\rangle$, so that (2.4) yields that

$$P\left(\left\{\omega \in \Omega\colon \left|n^{-1}\sum_{i=1}^{n} X_i(\omega) - p\right| \geq \varepsilon\right\}\right) \leq 1/4\,n\varepsilon^2 \tag{2.6}$$

holds for each binary random variable X such that $P(\{\omega \in \Omega\colon X(\omega) = 1\}) = p$, in this case evidently $EX = p$.

In the case of necessity, more detailed consequences of the Facts 2.1 and 2.2 will be drawn below in appropriate contexts.

3. Monte–Carlo algorithms for belief functions

The most obvious and immediate application of the SLLN to the case of belief functions reads as follows.

Fact 3.1. *Let $\langle \Omega, \mathscr{A}, P\rangle$ be a probability space, let $\langle \mathscr{P}(S), \mathscr{S}\rangle$ be a measurable space over the power-set $\mathscr{P}(S)$ of a nonempty set S such that $\mathscr{P}(T) \in \mathscr{S}$ holds for each $T \subset S$, let U be a (generalized set-valued) random variable defined on $\langle \Omega, \mathscr{A}, P\rangle$, taking its values in $\langle \mathscr{P}(S), \mathscr{S}\rangle$, and such that $U(\omega) \neq \emptyset$ holds for each $\omega \in \Omega$. Let $U_1, U_2, \dots$ be a sequence of statistically independent random variables defined on $\langle \Omega, \mathscr{A}, P\rangle$, taking their values in $\langle \mathscr{P}(S), \mathscr{S}\rangle$, and possessing the same probability distribution as U. Hence, for each $k \in \mathbb{N}^+ = \{1, 2, \dots\}$, each $n_1 < n_2 < \cdots < n_k$, and each $\mathscr{V}_1, \mathscr{V}_2, \dots, \mathscr{V}_k \in \mathscr{S}$, the equality*

$$P\left(\bigcap_{i=1}^{k} \{\omega \in \Omega\colon U_{n_i}(\omega) \in \mathscr{V}_i\}\right) = \prod_{i=1}^{n} P\left(\{\omega \in \Omega\colon U_{n_i}(\omega) \in \mathscr{V}_i\}\right) \tag{3.1}$$

holds. Moreover, for each $i \in \mathbb{N}^+$ and each $\mathscr{V} \in \mathscr{S}$,

$$P\left(\{\omega \in \Omega\colon U_i(\omega) \in \mathscr{V}\}\right) = P\left(\{\omega \in \Omega\colon U(\omega) \in \mathscr{V}\}\right). \tag{3.2}$$

Let $T \subset S$, let $\chi_{\mathscr{P}(T)}$ be the characteristic function (identifier) of the system $\mathscr{P}(T)$ of all subsets of T, so that $\chi_{\mathscr{P}(T)}(A) = 1$ if $A \in \mathscr{P}(T)$, i.e., if $A \subset T$, and $\chi_{\mathscr{P}(T)}(A) = 0$ otherwise. Let

$$\mathrm{bel}_U(T) = P\left(\{\omega \in \Omega\colon U(\omega) \subset T\}\right) \tag{3.3}$$

denote the belief function induced by the random variable U. Then

$$P\left(\left\{\omega \in \Omega\colon \lim_{n\to\infty}\left[n^{-1}\sum_{i=1}^{n}\chi_{\mathscr{P}(T)}\big(U_i(\omega)\big)\right] = \mathrm{bel}_U(T)\right\}\right) = 1\,. \tag{3.4}$$

□

P r o o f . It is an immediate consequence of SLLN. Setting

$$A_i = \big\{\omega \in \Omega\colon U_i(\omega) \in \mathscr{P}(T)\big\} \quad (= \{\omega \in \Omega\colon U_i(\omega) \subset T\})\,, \tag{3.5}$$

we can immediately observe that $\langle A_1, A_2, \dots, \rangle$ is a sequence of statistically independent random events such that $P(A_i) = \mathrm{bel}_U(T)$ holds for each $i \in \mathbb{N}^+$. Hence, $n^{-1}\sum_{i=1}^{n}\chi_{\mathscr{P}(T)}\ \big(U_i(\omega)\big)$ is the relative frequency of occurrences of a random event with the probability $\mathrm{bel}_U(T)$ in a sequence of n statistically independent realizations, and this relative frequency tends to $\mathrm{bel}_U(T)$ almost surely. □

Before focusing our attention to more sophisticated and complex Monte-Carlo algorithms for belief functions, we should mention, at least very briefly, our motivation for such an effort. In order to define and to compute the value $\mathrm{bel}_U(T)$ we need not only the notion of a "state compatible with an observation", defined by a binary compatibility relation between states and empirical values at the *intensional* level, but also the *extension* of this notion in the form of a subset of the state space S. If S is finite, such an extension can be obtained by testing, sequentially for each $s \in S$, whether s is compatible with the observation in question, if neglecting, for the moment, the problems with high computational complexity of such a procedure for large S. If S is infinite, this approach is impossible and problems arise with the definability of the set of compatible states in terms acceptable by the used formal apparatus. So, our aim will be to suggest a Monte-Carlo algorithm for belief functions using the binary compatibility relation between states and empirical values just at the level of a black box outputting the answers YES (compatible) or NO (incompatible) for each input pair "state-empirical value", but incapable to generate sets of states compatible with a given empirical observation.

Let S, $\langle E, \mathscr{E}\rangle$ and $\rho\colon E \times S \to \{0,1\}$ be as above. A random variable X defined on the probability space $\langle \Omega, \mathscr{A}, P\rangle$ and taking its values in the measurable space $\langle E, \mathscr{E}\rangle$ is called *regular with respect to the compatibility relation* ρ, if the mapping $U_{X,\rho}\colon \Omega \to \mathscr{P}(S)$, defined by

$$U_{X,\rho}(\omega) = U_{m_{\rho,X}}\big(X(\omega)\big) = \big\{s \in S\colon \rho\big(X(\omega),\, s\big) = 1\big\} \tag{3.6}$$

for each $\omega \in \Omega$, is a random variable defined on $\langle \Omega, \mathscr{A}, P\rangle$ and taking its values in the measurable space $\big\langle \mathscr{P}(S), \mathscr{S}\big\rangle$, where $\mathscr{S}$ is the minimal σ-field over the system $\{\mathscr{P}(T)\colon T \subset S\} \subset \mathscr{P}(\mathscr{P}(S))$. Hence, if X is regular with respect to ρ, then $\{\omega \in \Omega\colon U(X(\omega)) \subset T\} \in \mathscr{A}$ holds for each $T \subset S$, so that the value $\mathrm{bel}_{U_{X,\rho}}(T) = P(\{\omega \in \Omega\colon U_{X,\rho}(\omega) \subset T\})$ is defined for all $T \subset S$. Let $X_1, X_2, \ldots$ be a sequence of statistically independent and identically distributed random variables defined on $\langle \Omega, \mathscr{A}, P\rangle$, taking their values in $\langle E, \mathscr{E}\rangle$ and regular with respect to ρ.

FACT 3.2. *Let $U_i = U_{X_i,\rho}$ for each $i = 1, 2, \ldots$ Then $U_1, U_2, \ldots$ is a sequence of statistically independent and identically distributed random variables defined on $\langle \Omega, \mathscr{A}, P\rangle$ and taking their values in $\langle \mathscr{P}(S), \mathscr{S}\rangle$.* □

Remark. Throughout the rest of this paper we shall omit the symbols "$\ldots \omega \in \Omega\colon \ldots$" in the expressions defining particular subsets of Ω, if no misunderstanding menaces.

Proof. Let $\mathscr{V} \in \mathscr{S}$, let $i \in \mathbb{N}^+$. Then

$$
\begin{aligned}
&P\Big(\{U_i(\omega) \in \mathscr{V}\}\Big) = P\Big(\{U(X_i(\omega)) \in \mathscr{V}\}\Big) = P\Big(\{X_i(\omega) \in U^{-1}(\mathscr{V})\}\Big)\\
&= P\Big(\{X_i(\omega) \in \{x \in E\colon U(X) \in \mathscr{V}\}\}\Big) = P\Big(\{X_1(\omega) \in \{x \in E\colon U(X) \in \mathscr{V}\}\}\Big)\\
&= P\Big(\{U(X_1(\omega)) \in \mathscr{V}\}\Big) = P\Big(\{U_1(\omega) \in \mathscr{V}\}\Big)
\end{aligned}
\tag{3.7}
$$

due to the identical distribution of X_i and X_1, so that the random variables U_i and U_1 are also identically distributed. Let $k \in \mathbb{N}^+$, let $n_1 < n_2 < \cdots < n_k$ be positive integers, let $\mathscr{V}_1, \mathscr{V}_2, \ldots, \mathscr{V}_k \in \mathscr{S}$. Then

$$
\begin{aligned}
&P\left(\bigcap_{i=1}^{k} \{U_{n_i}(\omega) \in \mathscr{V}_i\}\right) = P\left(\bigcap_{i=1}^{k} \{U(X_{n_i}(\omega)) \in \mathscr{V}_i\}\right)\\
&= P\left(\bigcap_{i=1}^{k} \{X_{n_i}(\omega) \in U^{-1}(\mathscr{V}_i)\}\right) = \prod_{i=1}^{k} P\Big(\{X_{n_i}(\omega) \in U^{-1}(\mathscr{V}_i)\}\Big)\\
&= \prod_{i=1}^{k} P\Big(\{U(X_{n_i}(\omega)) \in \mathscr{V}_i\}\Big) = \prod_{i=1}^{k} P\Big(\{U_{n_i}(\omega) \in \mathscr{V}_i\}\Big)
\end{aligned}
\tag{3.8}
$$

due to the statistical independence of the random variables $X_1, X_2, \ldots$ So, $U_1, U_2, \ldots$ is a sequence of statistically independent and identically distributed (generalized set-valued) random variables. □

Let us suppose, from now on, that the state space S is finite or countable. Let $\{Y_{ij}\}_{i,j=1}^{\infty}$ be a system of statistically independent and identically distributed random variables defined on $\langle \Omega, \mathscr{A}, P\rangle$, taking their values in the measurable space $\langle S, \mathscr{P}(S)\rangle$ and such that, for each $s \in S$, $P(\{Y_{11}(\omega) = s\}) > 0$ holds.

What we have at our disposal, it is the system $\{\rho(X_i(\omega),\ Y_{ij}(\omega))\}_{i=1,\,j=1}^{n \quad m(i)}$ for some n, $m(i) \in \mathbb{N}^+$ of binary values, and our aim will be to estimate the value $\mathrm{bel}_{U_{X_1,\rho}}(T)$, for a given $T \subset S$, on the basis of these data.

Informally said, if for some $i \in \mathbb{N}^+$, for $m(i)$ large enough, and for all $j \leq m(i)$ such that $Y_{ij}(\omega)$ is compatible with $X_i(\omega)$, also the relation $Y_{ij}(\omega) \in T$ holds, we are tempted to believe that *all* $s \in S$ compatible with $X_i(\omega)$ belong to T, in other words, that the inclusion $U(X_i(\omega)) \subset T$ holds. So, setting $\lambda_i(T,\omega) = 1$, if for all $j \leq m(i)$ such that $\rho(X_i(\omega),\ Y_{ij}(\omega)) = 1$ also $\chi_T(Y_{ij}(\omega)) = 1$ holds, and setting $\lambda_i(T,\omega) = 0$ otherwise, we feel that $\lambda_i(T,\omega)$ is a reasonable and relatively good approximation of the value $\chi_{\mathscr{P}(T)}(U_{X_i,\rho}(\omega))$. Defining $\lambda_i(T,\omega)$ more formally, we can write that

$$\lambda_i(T,\omega) = \begin{cases} 1, & \text{if } (\forall\, j \leq m(i)) \Big(\rho(X_i(\omega),\ Y_{ij}(\omega)) \leq \chi_T(Y_{ij}(\omega))\Big), \\ 0, & \text{otherwise}, \end{cases} \tag{3.9}$$

hence,

$$\lambda_i(T,\omega) = \prod_{j=1}^{m(i)} \varphi_{ij}(T,\omega), \tag{3.10}$$

where

$$\varphi_{ij}(T,\omega) = \begin{cases} 1, & \text{if } \rho(X_i(\omega),\, Y_{ij}(\omega)) \leq \chi_T(Y_{ij}(\omega)), \\ 0, & \text{otherwise}. \end{cases} \tag{3.11}$$

Accepting the convention that $0^0 = 1$ we can easily obtain that

$$\begin{aligned} \varphi_{ij}(T,\omega) &= \min\Big(1,\ 1 - \rho(X_i(\omega),\ Y_{ij}(\omega)) + \chi_T(Y_{ij}(\omega))\Big) \\ &= \big[\chi_T(Y_{ij}(\omega))\big]^{\rho(X_i(\omega),Y_{ij}(\omega))} \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} \lambda_i(T,\omega) &= \prod_{j=1}^{m(i)} \min\Big(1,\ 1 - \rho(X_i(\omega),Y_{ij}(\omega)) + \chi_T(Y_{ij}(\omega))\Big) \\ &= \prod_{j=1}^{m(i)} \big[\chi_T(Y_{ij}(\omega))\big]^{\rho(X_i(\omega),Y_{ij}(\omega))}. \end{aligned} \tag{3.13}$$

Using the values $\lambda_i(T,\omega)$ as approximations of $\chi_{\mathscr{P}(T)}\big(U_{X_i,\rho}(\omega)\big)$ as in Fact 3.1, we arrive at the following random variable.

$$\begin{aligned} B_{X_1,\rho}\big(T,n,\langle m(i)\rangle_{i=1}^{n},\omega\big) &= n^{-1}\sum_{i=1}^{n}\lambda_i(T,\omega) \\ &= n^{-1}\sum_{i=1}^{n}\prod_{j=1}^{m(i)}\min\big(1,\,1-\rho\big(X_i(\omega),\,Y_{ij}(\omega)\big)+\chi_T\big(Y_{ij}(\omega)\big)\big) \\ &= n^{-1}\sum_{i=1}^{n}\prod_{j=1}^{m(i)}\big[\chi_T\big(Y_{ij}(\omega)\big)\big]^{\rho(X_i(\omega),Y_{ij}(\omega))}. \end{aligned} \tag{3.14}$$

We shall try to demonstrate, in the rest of this paper, that, in a certain sense and degree, the value $B_{X_1,\rho}\big(T,n,\langle m(i)\rangle_{i=1}^{n},\omega\big)$ is a reasonable and good estimation of the value $\mathrm{bel}_{U_{X_1,\rho}}(T)$.

4. Asymptotic properties of the proposed estimation

FACT 4.1. *Let the systems $\{X_i\}_{i=1}^{n}$ and $\{Y_{ij}\}_{i=1,\ j=1}^{n\quad m(i)}$ of random variables defined above be mutually statistically independent, let $\langle E,\mathscr{E}\rangle$ be such that E is finite or countable and $\mathscr{E}=\mathscr{P}(E)$, let $\varepsilon>0$ be given. Then there exist a value $\lambda^*\in\big\langle\,\mathrm{bel}_{U_{X_1,\rho}}(T),\mathrm{bel}_{U_{X_1,\rho}}(T)+\varepsilon\big)$ and a value $m^*\in\mathbb{N}^+$ such that, if $m(i)\geq m^*$ for each $i\in\mathbb{N}^+$, then the relation*

$$P\Big(\big\{\omega\in\Omega\colon\ \lim_{n\to\infty}B_{X_1,\rho}\big(T,n,\langle m(i)\rangle_{i=1}^{n},\omega\big)=\lambda^*\big\}\Big)=1 \tag{4.1}$$

holds. □

Proof. First, let us prove that, given $T\subset S$, $\big\langle\lambda_i(T,\cdot)\big\rangle_{i=1}^{n}$ is a sequence of statistically independent random variables. So, we have to prove that the equality

$$P\Big(\bigcap_{i=1}^{n}\{\lambda_i(T,\omega)=\kappa_i\}\Big)=\prod_{i=1}^{n}P\big(\{\lambda_i(T,\omega)=\kappa_i\}\big) \tag{4.2}$$

holds for each n-tuple $\langle\kappa_1,\kappa_2,\dots,\kappa_n\rangle\in\{0,1\}^n$. The proof will be done by induction on the number of occurrences of zeros in $\langle\kappa_1,\kappa_2,\dots,\kappa_n\rangle$, so that the first step deals with the case when $k=0$, i.e., when $\kappa_i=1$ for each $i\leq n$.

Set $F_T(s)=\big\{x\in E\colon \rho(x,s)=0\big\}$ if $s\in S-T$ and $F_T(s)=E$ otherwise. An easy calculation using the supposed statistical independences among particular random variables X_i, $i\leq n$, Y_{ij}, $i\leq n$, $j\leq m(i)$, as well as between the systems $\{X_i\}_{i=1}^{n}$ and $\{Y_{ij}\}_{i=1,j=1}^{n\quad m(i)}$ of random variables, yield that

$$P\Big(\bigcap_{i=1}^{n}\Big\{\lambda_i(T,\omega)=1\Big\}\Big)=P\Big(\bigcap_{i=1}^{n}\bigcap_{j=1}^{m(i)}\Big\{\rho\big(X_i(\omega),\ Y_{ij}(\omega)\big)\le\chi_T\big(Y_{ij}(\omega)\big)\Big\}\Big)$$

$$P\Big(\bigcup_{\langle\langle s_1^1,\dots,s_{m(1)}^1\rangle,\langle s_1^2,\dots,s_{m(2)}^2\rangle,\dots,\langle s_1^n,\dots,s_{m(n)}^n\rangle\rangle\in\mathop{\mathsf{X}}_{i=1}^{n} S^{m(i)}}$$

$$\bigcap_{\langle j_1,\dots,j_n\rangle=\langle 1,1,\dots,1\rangle}^{\langle m(1),m(2),\dots,m(n)\rangle}\Big\{\rho\big(X_i(\omega),s_{j_i}^i\big)\le\chi_T(s_{j_i}^i)\,,\ Y_{ij_i}(\omega)=s_{j_i}^i\Big\}\Big)$$

$$=\sum_{\langle\langle s_1^1,\dots,s_{m(1)}^1\rangle,\langle s_1^2,\dots,s_{m(2)}^2\rangle,\dots,\langle s_1^n,\dots,s_{m(n)}^n\rangle\rangle\in\mathop{\mathsf{X}}_{i=1}^{n} S^{m(i)}}$$

$$P\Big(\bigcap_{i=1}^{n}\bigcap_{j=1}^{m(i)}\Big\{\rho\big(X_i(\omega),s_j^i\big)\le\chi_T(s_j^i),\ Y_{ij}(\omega)=s_j^i\Big\}\Big)$$

$$=\sum_{\langle\langle s_1^1,\dots,s_{m(1)}^1\rangle,\langle s_1^2,\dots,s_{m(2)}^2\rangle,\dots,\langle s_1^n,\dots,s_{m(n)}^n\rangle\rangle\in\mathop{\mathsf{X}}_{i=1}^{n} S^{m(i)}}$$

$$\prod_{i=1}^{n}\prod_{j=1}^{m(i)}P\Big(\big\{X_i(\omega)\in F_T(s_j^i)\,,\ Y_{ij}(\omega)=s_j^i\big\}\Big) \tag{4.3}$$

$$=\sum_{\langle s_1^n,\dots,s_{m(n)}^n\rangle\in S^{m(n)}}\sum_{\langle\langle s_1^1,\dots,s_{m(1)}^1\rangle,\dots,\langle s_1^{n-1},\dots,s_{m(n-1)}^{n-1}\rangle\rangle\in\mathop{\mathsf{X}}_{i=1}^{n-1} S^{m(i)}}$$

$$\left[\prod_{i=1}^{n-1}\prod_{j=1}^{m(i)}P\Big(\big\{X_i(\omega)\in F_T(s_j^i)\,,\ Y_{ij}(\omega)=s_j^i\big\}\Big)\right]\times$$

$$\times\left[\prod_{k=1}^{m(n)}P\Big(\big\{X_n(\omega)\in F_T(s_k^n)\,,\ Y_{nk}(\omega)=s_k^n\big\}\Big)\right]$$

$$=\left[\sum_{\langle s_1^n,\dots,s_{m(n)}^n\rangle\in S^{m(n)}}\prod_{k=1}^{m(n)}P\Big(\big\{X_n(\omega)\in F_T(s_k^n)\,,\ Y_{nk}(\omega)=s_k^n\big\}\Big)\right]\times$$

$$\times\Bigg[\sum_{\langle\langle s_1^1,\dots,s_{m(1)}^1\rangle,\langle s_1^2,\dots,s_{m(2)}^2\rangle,\langle s_1^{n-1},\dots,s_{m(n-1)}^{n-1}\rangle\rangle\in\mathop{\mathsf{X}}_{i=1}^{n-1} S^{m(i)}}$$

$$\prod_{i=1}^{n-1}\prod_{j=1}^{m(i)}P\Big(\big\{X_i(\omega)\in F_T(s_j^i),\ Y_{ij}(\omega)=s_j^i\big\}\Big)\Bigg].$$

Applying the same calculation to the last component in the last right-hand side expression in (4.3) we obtain, after $n-1$ further steps, that

$$\begin{aligned} & P\left(\bigcap_{i=1}^{n} \{\lambda_i(T,\omega)=1\}\right) \\ &= \prod_{i=1}^{n}\left[\sum\nolimits_{\langle s_1^i,\ldots,s_{m(i)}^i\rangle\in S^{m(i)}} \prod_{k=1}^{m(i)} P\Big(\{X_i(\omega)\in F_T(s_k^i), Y_{ik}(\omega)=s_k^i\}\Big)\right] \\ &= \prod_{i=1}^{n}\left[\sum\nolimits_{\langle s_1^i,\ldots,s_{m(i)}^i\rangle\in S^{m(i)}} \prod_{k=1}^{m(i)} P\Big(\{\rho(X_i(\omega),s_k^i)\le \chi_T(s_k^i), Y_{ik}(\omega)=s_k^i\}\Big)\right] \\ &= \prod_{i=1}^{n} P\left(\bigcap_{k=1}^{m(i)} \Big\{\rho\big(X_i(\omega),Y_{ik}(\omega)\le \chi_T\big(Y_{ik}(\omega)\big)\Big\}\right) \qquad (4.4) \\ &= \prod_{i=1}^{n} P\Big(\{\lambda_i(T,\omega)=1\}\Big)\,. \end{aligned}$$

Hence, the first step of our induction, i.e., the validity of the relation (4.2) for $\langle\kappa_1,\kappa_2,\ldots,\kappa_n\rangle=\langle 1,1,\ldots,1\rangle$, is proved.

Let (4.2) hold for $k\le K$, i.e., for each $\langle\kappa_1,\kappa_2,\ldots,\kappa_n\rangle\in\{0,1\}^n$ such that $\sum\limits_{i=1}^{n}\kappa_i\ge n-K$, let $\langle\kappa_1,\kappa_2,\ldots,\kappa_n\rangle$ be such that $\sum\limits_{i=1}^{n}\kappa_i=n-K-1$. Take arbitrarily an $i_0\le n$ such that $\kappa_{i_0}=0$, and set $\kappa_i^*=\kappa_i$ for all $i\ne i_0$, $\kappa_{i_0}^*=1$, hence, $\sum\limits_{i=1}^{n}\kappa_i^*=n-K$. Then

$$\begin{aligned} & P\left(\bigcap_{i=1}^{n} \{\lambda_i(T,\omega)=\kappa_i\}\right)+P\left(\bigcap_{i=1}^{n} \{\lambda_i(T,\omega)=\kappa_i^*\}\right) \\ &= P\left(\bigcap_{i=1}^{n} \{\lambda_i(T,\omega)=\kappa_i\}\right)+\prod_{i=1}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i^*\}\Big) \qquad (4.5) \\ &= P\left(\bigcap_{\substack{i=1\\ i\ne i_0}}^{n} \{\lambda_i(T,\omega)=\kappa_i^*\}\right)=\prod_{\substack{i=1\\ i\ne i_0}}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i^*\}\Big)\,, \end{aligned}$$

so that

$$
\begin{aligned}
&P\left(\bigcap_{i=1}^{n}\{\lambda_i(T,\omega)=\kappa_i\}\right)\\
&=\prod_{i=1\ i\neq i_0}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i^*\}\Big)-\prod_{i=1}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i^*\}\Big)\\
&=\prod_{i=1\,,\,i\neq i_0}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i^*\}\Big)\Big[1-P\Big(\{\lambda_{i_0}(T,\omega)=\kappa_{i_0}^*\}\Big)\Big]\\
&=\prod_{i=1\ i\neq i_0}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i\}\Big)\Big[1-P\Big(\{\lambda_{i_0}(T,\omega)=1\}\Big)\Big] \qquad (4.6)\\
&=\left[\prod_{i=1\ i\neq i_0}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i\}\Big)\right]P\Big(\{\lambda_{i_0}(T,\omega)=0\}\Big)\\
&=\left[\prod_{i=1\ i\neq i_0}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i\}\Big)\right]P\Big(\{\lambda_{i_0}(T,\omega)=\kappa_{i_0}\}\Big)\\
&=\prod_{i=1}^{n} P\Big(\{\lambda_i(T,\omega)=\kappa_i\}\Big)\,.
\end{aligned}
$$

So, (4.2) holds also for all $\langle\kappa_1,\kappa_2,\dots,\kappa_n\rangle$ with $K+1$ occurrences of zero, hence, by induction, (4.2) holds for each $\langle\kappa_1,\kappa_2,\dots,\kappa_n\rangle\in\{0,1\}^n$, and the random variables $\{\lambda_i(T,\omega)\}_{i=1}^{n}$ are proved to be statistically independent.

Our further goal will be to compute the value $P(\{\lambda_i(T,\omega)=1\})$, which is evidently identical with the expected value $E\lambda_i(T,\cdot)$. An easy calculation yields that

$$
\begin{aligned}
&P(\{\lambda_i(T,\omega)=1\})=P\left(\bigcap_{j=1}^{m(i)}\{\rho(X_i(\omega),\,Y_{ij}(\omega))\le\chi_T(Y_{ij}(\omega))\}\right)\\
&=P\left(\left[\bigcap_{j=1}^{m(i)}\Big\{\rho(X_i(\omega),\,Y_{ij}(\omega))\le\chi_T(Y_{ij}(\omega))\Big\}\right]\cap\{U(X_i(\omega))\subset T\}\right)\\
&\qquad+P\left(\left[\bigcap_{j=1}^{m(i)}\Big\{\rho(X_i(\omega)\,,\,Y_{ij}(\omega))\le\chi_T(Y_{ij}(\omega))\Big\}\right]\cap\{U(X_i(\omega))\not\subset T\}\right) \qquad (4.7)\\
&=P\Big(\{U(X_i(\omega))\subset T\}\Big)\\
&\qquad+P\left(\left[\bigcap_{j=1}^{m(i)}\Big\{\rho(X_i(\omega),\,Y_{ij}(\omega))\le\chi_T(Y_{ij}(\omega))\Big\}\right]\cap\{U(X_i(\omega))\not\subset T\}\right),
\end{aligned}
$$

as $U(X_i(\omega)) \subset T$ implies that if $Y_{ij}(\omega) \in U(X_i(\omega))$, i.e., if $\rho(X_i(\omega), Y_{ij}(\omega)) = 1$, then $Y_{ij}(\omega) \in T$, hence, $\chi_T(Y_{ij}(\omega)) = 1$, consequently, for all $j \le m(i)$ the inequality $\rho(X_i(\omega), Y_{ij}(\omega)) \le \chi_T(Y_{ij}(\omega))$ holds. Another easy calculation yields that

$$
\begin{aligned}
&P\left(\bigcap_{j=1}^{m(i)} \{\rho(X_i(\omega), Y_{ij}(\omega)) \le \chi_T(Y_{ij}(\omega))\} \cap \{U(X_i(\omega)) \not\subset T\}\right) \\
&= P\left(\left[\bigcap_{j=1}^{m(i)} \left\{Y_{ij}(\omega) \notin \Big(U(X_i(\omega)) - T\Big)\right\}\right] \cap \{U(X_i(\omega)) \not\subset T\}\right) \\
&= P\left(\left[\bigcap_{j=1}^{m(i)} \left\{Y_{ij}(\omega) \in S - \Big(U(X_i(\omega)) - T\Big)\right\}\right] \cap \{U(X_i(\omega)) \not\subset T\}\right) \\
&= \sum_{\substack{x\in E \\ U(x)\not\subset T}} P\left(\bigcap_{j=1}^{m(i)} \left\{Y_{ij}(\omega) \in S - \big(U(x) - T\big)\right\} \cap \{X_i(\omega) = x\}\right) \qquad (4.8) \\
&= \sum_{\substack{x\in E \\ U(x)\not\subset T}} \left\{\left[\prod_{j=1}^{m(i)} \Big[1 - P\Big(\{Y_{ij}(\omega) \in U(x) - T\}\Big)\Big]\right] P\Big(\{X_i(\omega) = x\}\Big)\right\} \\
&= \sum_{\substack{x\in E \\ U(x)\not\subset T}} \left\{\Big[1 - P\Big(\{Y_{11}(\omega) \in U(x) - T\}\Big)\Big]^{m(i)} P\Big(\{X_i(\omega) = x\}\Big)\right\},
\end{aligned}
$$

as the random variables Y_{ij} are supposed to be statistically independent and identically distributed. Random variables X_i are also identically distributed, so that we can replace X_i by X_1 in (4.8).

As the set E is finite or countable, there exists, given $\varepsilon > 0$, a finite subset $E_0 = E_0(\varepsilon) \subset E$ such that $P(\{X_1(\omega) \in E_0\}) > 1 - (\varepsilon/2)$ holds. If $x \in E_0$ is such that $U(x) \not\subset T$, i.e., such that $U(x) - T \neq \emptyset$, then $P(\{Y_{11}(\omega) \in U(x) - T\}) = \sum_{s\in U(x)-T} P(\{Y_{11}(\omega) = s\}) > 0$ holds due to the conditions imposed to random variables Y_{ij}. Hence, the value

$$m^*(x) = \min\Big\{m \in \mathbb{N}^+ : \Big(1 - P\big(\{Y_{11}(\omega) \in U(x) - T\}\big)\Big)^m < \varepsilon/2\Big\} < \infty \quad (4.9)$$

is uniquely defined for each $x \in E_0$ such that $U(x) \not\subset T$. Set

$$m^* = \max\{m^*(x) : x \in E_0\} \quad (4.10)$$

and define

$$\lambda^{**} = \sum_{\substack{x\in E \\ U(x)\not\subset T}} \left\{\Big[1 - P\Big(\{Y_{11}(\omega) \in U(x) - T\}\Big)\Big]^{m^*} P\Big(\{X_1(\omega) = x\}\Big)\right\}. \quad (4.11)$$

Obviously,

$$
\begin{aligned}
\lambda^{**} = & \sum_{\substack{x \in E_0 \\ U(x) \not\subset T}} \left\{ \left[1 - P\Big(\{ Y_{11}(\omega) \in U(x) - T \} \Big) \right]^{m^*} P\Big(\{ X_1(\omega) = x \} \Big) \right\} \\
& + \sum_{\substack{x \in E - E_0 \\ U(x) \not\subset T}} \left\{ \left[1 - P\Big(\{ Y_{11}(\omega) \in U(x) - T \} \Big) \right]^{m^*} P\Big(\{ X_1(\omega) = x \} \Big) \right\} \\
\leq & \sum_{\substack{x \in E_0 \\ U(x) \not\subset T}} \left\{ \left[1 - P\Big(\{ Y_{11}(\omega) \in U(x) - T \} \Big) \right]^{m^*} P\Big(\{ X_1(\omega) = x \} \Big) \right\} \\
& + \sum_{x \in E - E_0} P\Big(\{ X_1(\omega) = x \} \Big) \qquad (4.12) \\
< & \sum_{x \in E_0} (\varepsilon/2) P\Big(\{ X_1(\omega) = x \} \Big) + P\Big(\{ X_1(\omega) \in E - E_0 \} \Big) \\
< & \, (\varepsilon/2) P\Big(\{ X_1(\omega) \in E_0 \} \Big) + (\varepsilon/2) \leq \varepsilon \,.
\end{aligned}
$$

Consequently, setting $m(i) = m^*$ for each $i \in \mathbb{N}^+$, setting

$$
\lambda^* = P\Big(\{ U(X_1(\omega)) \subset T \} \Big) + \lambda^{**} = \mathrm{bel}_{U_{X_1,\rho}}(T) + \lambda^{**} \,, \tag{4.13}
$$

and combining together all the relations from (4.7) to (4.13) we obtain that, for each $i \in \mathbb{N}^+$,

$$
\begin{aligned}
& P\Big(\{ \lambda_i(T, \omega) = 1 \} \Big) = P\Big(\{ U(X_1(\omega)) \subset T \} \Big) + \lambda^{**} \\
& = \lambda^* \in \Big\langle P\Big(\{ U(X_1(\omega)) \subset T \} \Big), \, P\Big(\{ U(X_1(\omega)) \subset T \} \Big) + \varepsilon \Big\rangle \\
& = \big\langle \mathrm{bel}_{U_{X_1,\rho}}(T), \, \mathrm{bel}_{U_{X_1,\rho}}(T) + \varepsilon \big\rangle \,.
\end{aligned} \tag{4.14}
$$

Hence, $\{\lambda_i(T, \cdot)\}_{i=1}^{\infty}$ is a sequence of statistically independent and identically distributed binary-valued random variables so that, due to the strong law of large numbers, the relative frequency of occurrences of unit values taken by $\lambda_i(T, \omega)$, i.e., the value $B_{X_1,\rho}\big(T, \, n, \, \langle m(i) \rangle_{i=1}^{n}, \, \omega\big)$ according to (3.14), tends, with n increasing, almost surely to the probability λ^* with which each $\lambda_i(T, \cdot)$ takes the unit value. So, the assertion (4.1) is proved. □

Fact 4.2. *Let the notations and conditions of Fact 4.1 hold, let* $\lim_{i \to \infty} m(i) = \infty$. *Then the equality*

$$
P\Big(\{ \omega \in \Omega \colon \lim_{n \to \infty} B_{X_1,\rho}(T, n, \langle m(i) \rangle_{i=1}^{n}, \omega) = \mathrm{bel}_{U_{X_1,\rho}}(T) \} \Big) = 1 \tag{4.15}
$$

holds. □

P r o o f . Let $m(1), m(2), \dots \in \mathbb{N}^+$ be such that $\lim_{i \to \infty} m(i) = \infty$. Set, for each $i, k \in \mathbb{N}^+$,

$$\widetilde{m}^{(k)}(i) = m(k), \quad \text{if} \quad i \le k\,, \qquad \widetilde{m}^{(k)}(i) = m(i)\,, \quad \text{if} \quad i \ge k\,. \tag{4.16}$$

Define $\lambda_i^{(k)}(T, \omega)$ analogously to (3.13), but with $m(i)$ replaced by $\widetilde{m}^{(k)}(i)$, so that

$$\widetilde{\lambda}_i^{(k)}(T, \omega) = \prod_{j=1}^{\widetilde{m}^{(k)}(i)} \min\left\{1, 1 - \rho\big(X_i(\omega), Y_{ij}(\omega)\big) + \chi_T\big(Y_{ij}(\omega)\big)\right\}. \tag{4.17}$$

Consequently, for $i \ge k$ the equality $\widetilde{\lambda}_i^{(k)}(T, \omega) = \lambda_i(T, \omega)$ holds for each $\omega \in \Omega$. The values taken by the random variables $\widetilde{\lambda}_i^{(k)}(T, \cdot)$ are either 0 or 1, so that the limit values of the relative frequencies of both the results, if defined, do not depend on any initial segment of a sequence of such values. Hence,

$$\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \widetilde{\lambda}_i^{(k)}(T, \omega) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \lambda_i(T, \omega) \tag{4.18}$$

for each $k \in \mathbb{N}^+$ and each $\omega \in \Omega$ for which one of these limit values is defined. So, setting

$$\widetilde{B}_{X_1,\rho}^{(k)}\Big(T, n, \big\langle m(i) \big\rangle_{i=1}^{n}, \omega\Big) = n^{-1} \sum_{i=1}^{n} \widetilde{\lambda}_i^{(k)}(T, \omega) \tag{4.19}$$

for each $k \in \mathbb{N}^+$ and each $\omega \in \Omega$, we obtain that

$$\begin{aligned}
&\lim_{n \to \infty} B_{X_1,\rho}^{(k)}\Big(T, n, \big\langle \tilde{m}^{(k)}(i) \big\rangle_{i=1}^{n}, \omega\Big) \\
= &\lim_{n \to \infty} \widetilde{B}_{X_1,\rho}^{(k)}\Big(T, n, \big\langle \widetilde{m}(i) \big\rangle_{i=1}^{n}, \omega\Big) \\
= &\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \widetilde{\lambda}_i^{(k)}(T, \omega) \\
= &\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \lambda_i^{(k)}(T, \omega) \\
= &\lim_{n \to \infty} B_{X_1,\rho}\Big(T, n, \big\langle m(i) \big\rangle_{i=1}^{n}, \omega\Big)
\end{aligned} \tag{4.20}$$

by (3.14) for each $\omega \in \Omega$ for which at least one of the limit values in (4.20) is defined, and for each $k \in \mathbb{N}^+$.

Take arbitrarily $\varepsilon > 0$ and define, for $x \in E$, $m_\varepsilon^*(x)$ by (4.9); define also m_ε^* by (4.10). As $\lim_{k \to \infty} m(k) = \infty$, there exists $k_0 \in \mathbb{N}^+$ such that $m(k) \ge m_\varepsilon^*$ holds for each $k \geq k_0$, hence, $\widetilde{m}^{(k)}(j) \geq m_\varepsilon^*$ holds for each such k and each $j \in \mathbb{N}^+$. Fact 4.1 then yields that, for each $k \geq k_0$, the relation

$$\lim_{n \to \infty} B_{X_1,\rho}\Big(T, n, \langle \widetilde{m}^{(k)}(i) \rangle_{i=1}^n, \omega\Big) \in \big\langle \operatorname{bel}_{U_{X_1,\rho}}(T), \operatorname{bel}_{U_{X_1,\rho}}(T) + \varepsilon \big), \quad (4.21)$$

so, due to (4.20) also the relation

$$\lim_{n \to \infty} B_{X_1,\rho}\Big(T, n, \langle m(i) \rangle_{i=1}^n, \omega\Big) \in \big\langle \operatorname{bel}_{U_{X_1,\rho}}(T), \operatorname{bel}_{U_{X_1,\rho}}(T) + \varepsilon \big) \qquad (4.22)$$

hold for each $\omega \in \Omega_0 \subset \Omega$, where $\Omega_0 \in \mathscr{A}$ and $P(\Omega_0) = 1$. As $\varepsilon > 0$ was taken arbitrarily, (4.22) immediately implies that

$$P\Big(\big\{\omega \in \Omega\colon \lim_{n \to \infty} B_{X_1,\rho}(T, n, \langle m(i) \rangle_{i=1}^n, \omega) = \operatorname{bel}_{U_{X_1,\rho}}(T)\big\}\Big) = 1, \qquad (4.23)$$

and the assertion is proved. $\square$

5. A sketch of a Monte-Carlo algorithm for statistical estimation of the values of a belief function

Let us describe, very briefly, a Monte-Carlo algorithm based on the limit assertions presented and proved above and offering numerical values which can be taken, in the sense specified by Fact 5.1 below, as reasonable estimations of the values of the belief function in question. Let $S = \{s_1, s_2, \dots, s_N\}$ be a nonempty finite state space of an investigated system, let $E = \{e_1, e_2, \dots, e_K\}$ be a nonempty finite space of empirical or observational values, and let $\rho\colon E \times S \to \{0,1\}$ be a compatibility relation such that $\rho(e_i, s_j) = 1$ iff s_j cannot be eliminated from the set of possible internal states of the investigated system supposing that the empirical value e_i was obtained or observed as the result of some measurements, experiments or observations concerning the system in question and/or the environment where it is situated.

The role of the input of the proposed algorithm is played by a finite sequence $Y = \langle x_i, y_i, z_i \rangle_{i=1}^R$ of triples, where $x_i \in E$, $y_i \in S^* = \bigcup_{k=0}^{\infty} S^k$, and $z_i \in \{0,1\}^* = \bigcup_{i=0}^{\infty} \{0,1\}^k$. Namely, for each $1 \leq i \leq R$, $x_i \in E$ is an empirical value, i.e., $x_i = e_j$ for some $j \leq \mathbb{N}$, $y_i = \langle s_{i1}, s_{i2}, \dots, s_{im(i)} \rangle$ is a finite nonempty string of elements of the state space S, and $z_i = \big\langle \delta_{i1}, \delta_{i2}, \dots, \delta_{i\ell(i)} \big\rangle$ is a finite binary string such that $\ell(i) = m(i)$ and $\delta_{ij} = \rho(x_i, s_{ij})$ for each $1 \leq j \leq m(i)$. Informally, x_i is an element of E sampled at random by the random variable X_i; s_{ij} are the elements of the state space S sampled at random by the random variables Y_{ij} and tested whether they are compatible with x_i or not, and the results of these tests are encoded by z_i. Evidently, x_i may be the same for different triples and also the possibility that two or more triples in Y are identical is not excluded.

Set, for each $1 \leq i \leq R$ and for $\langle x_i, y_i, z_i \rangle$, $y_i = \langle s_{i1}, s_{i2}, \ldots, s_{im(i)} \rangle$, $z_i = \langle \delta_{i1}, \delta_{i2}, \ldots, \delta_{im(i)} \rangle$,

$$u(x_i) = \{ s_{ij} \colon 1 \leq j \leq m(i),\, \delta_{ij} = 1 \}, \tag{5.1}$$

set also, for each $1 \leq j \leq K$,

$$\mathscr{U}(e_j) = \bigcup_{\substack{1 \leqslant k \leqslant R \\ x_k = e_j}} u(x_k). \tag{5.2}$$

The set $\mathscr{U}(e_j)$ approximates, in the statistical sense, the set $U_\rho(e_j)$ of states compatible with the empirical value $e_j \in E$. Setting, moreover,

$$p(e_j) = R^{-1} \operatorname{card}\{ k \colon 1 \leq k \leq R,\, x_k = e_j \}, \tag{5.3}$$

we can see that $p(e_j)$ approximates the probability with which $X_i(\omega) = e_j$. Combining both these approximations together, we can set

$$\mathrm{bel}^0(T) = \sum_{\substack{1 \leqslant j \leqslant K \\ \mathscr{U}(e_j) \subset T}} p(e_j) \tag{5.4}$$

and we obtain that $\mathrm{bel}^0(T)$ is a reasonable statistical estimation of the value $\mathrm{bel}_{U_{X_1,\rho}}(T)$. The quality of this estimation is stated by the following assertion.

FACT 5.1. *Let the notations and conditions of Fact 4.1 hold, let $\varepsilon > 0$ be such that the inequality $\varepsilon > E\lambda_i(T,\cdot) - \mathrm{bel}_{U_{X_i,\rho}}(T)$ holds for a given subset $T \subset S$, where $\lambda_i(T,\cdot)$ is defined by (3.9). Then the inequality*

$$\begin{aligned} &P\Big(\big\{ \omega \in \Omega \colon \big| B_{X_1,\rho}(T, n, \langle m(i) \rangle_{i=1}^n, \omega) - \mathrm{bel}_{U_{X_1,\rho}}(T) \big| \geq \varepsilon \big\} \Big) \\ &\leq 1/4n\big(\varepsilon - \big(E\lambda_i(T,\cdot) - \mathrm{bel}_{U_{X_1,\rho}}(T)\big)\big)^2 \end{aligned} \tag{5.5}$$

holds. □

Proof. Due to the Chebyshev inequality for binary random variables we obtain that

$$P\Big(\big\{ \omega \in \Omega \colon \big| B_{X_1,\rho}(T, n, \langle m(i) \rangle_{i=1}^n, \omega) - E\lambda_i(T,\cdot) \big| \geq \varepsilon \big\} \Big) \leq \frac{1}{4n\varepsilon^2} \tag{5.6}$$

holds due to Fact 4.1. Evidently, if

$$\Big| B_{X_1,\rho}(T, n, \langle m(i) \rangle_{i=1}^n, \omega) - \mathrm{bel}_{U_{X_1,\rho}}(T) \Big| \geq \varepsilon \tag{5.7}$$

holds, then

$$\left|B_{X_1,\rho}(T,n,\langle m(i)\rangle_{i=1}^n,\omega) - E\lambda_i(T,\cdot)\right| \geq \varepsilon - \left|E\lambda_i(T,\cdot) - \mathrm{bel}_{U_{X_1,\rho}}(T)\right| \tag{5.8}$$

must hold as well. So,

$$\begin{aligned} &P\Big(\{\omega\in\Omega\colon \left|B_{X_1,\rho}(T,n,\langle m(i)\rangle_{i=1}^n,\omega) - \mathrm{bel}_{U_{X_1,\rho}}(T)\right| \geq \varepsilon\}\Big) \\ &\leq P\Big(\{\omega\in\Omega\colon \left|B_{X_1,\rho}(T,n,\langle m(i)\rangle_{i=1}^n,\omega) - E\lambda_i(T,\cdot)\right| \geq \\ &\qquad\qquad \varepsilon - \big(E\lambda_i(T,\cdot) - \mathrm{bel}_{U_{X_1,\rho}}(T)\big)\}\Big) \\ &\leq 1/4n\big(\varepsilon - \big(E\lambda_i(T,\cdot) - \mathrm{bel}_{U_{X_1,\rho}}(T)\big)\big)^2 \end{aligned} \tag{5.9}$$

follows immediately (from) 5.6 when replacing ε by $\varepsilon - \left|E\lambda_i(T,\cdot) - \mathrm{bel}\, U_{X_1,\rho}(T)\right|$ and taking into consideration that $E\lambda_i(T,\cdot) \geq \mathrm{bel}_{U_{X_1,\rho}}(T)$ always holds (cf. the proof of Fact 4.1). □

The dependence of the upper bound on the right-hand side of (5.5) on the value $\mathrm{bel}_{X_1,\rho}(T)$ can be eliminated, due to Fact 4,1, when choosing $m(i)$ uniformly large enough.

FACT 5.2. *Let the notations and conditions of Fact 4.1 hold, let $\varepsilon > 0$ be given. Then there exists $m^*(\varepsilon) \in \mathbb{N}^+$ such that, if $m(i) \geq m^*(\varepsilon)$ holds for each $i \leq n$, then*

$$P\left(\left\{\omega\in\Omega\colon \left|B_{X_1,\rho}(T,n,\langle m(i)\rangle_{i=1}^n,\omega) - \mathrm{bel}_{U_{X_1,\rho}}(T)\right| \geq \varepsilon\right\}\right) \leq 1/n\varepsilon^2 \tag{5.10}$$

holds. □

P r o o f. Due to Fact 4.1 there exists $m^*(\varepsilon)$ such that

$$\left|E\lambda_i(T,\cdot) - \mathrm{bel}_{U_{X_1,\rho}}(T)\right| \leq \varepsilon/2 \tag{5.11}$$

holds supposing that $m(i) \geq m^*(\varepsilon)$ holds for each $i \leq n$. Combining (5.5) and (5.11) we obtain immediately that (5.10) holds. □

As this paper seems to be almost self-explanatory, up to some elementary foundations of probability theory, let us limit ourselves to three references to textbooks dealing with this theory, namely to [2], [6] and [8]. A more general setting of probability theory as a particular case of measure theory can be found in [4]. Philosophical and methodological background concerning the strong law of large numbers is discussed in [3]. The most elementary ideas, notions and results of the Dempster-Shafer theory in the extent sufficient for the purposes of this paper are explained and discussed in [1] or [7]. We do not introduce here the already extremely numerous references dealing with Monte-Carlo algorithms and estimations applied to various computation and decision problems.

REFERENCES

[1] DUBOIS, D.—PRADE, H.: *Théorie de Possibilités – Applications à le Représentation de Connaissances en Informatique*, Mason, Paris, 1985.

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

[3] FINE, T. L.: *Theories of Probability – an Examination of Foundations*, Academic Press, New York, 1973.

[4] HALMOS, P. R.: *Measure Theory*, D. van Nostrand, New York–Toronto–London, 1950.

[5] KRAMOSIL, I.: *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.

[6] LOÈVE, M.: *Probability Theory*, D. van Nonstrand, New York, 1960.

[7] SHAFER, G.: *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, 1976.

[8] ŠTĚPÁN, J.: *Teorie pravděpodobnosti* (*Probability Theory – in Czech*), Academia, Prague, 1987.

Received February 22, 1998

Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2
182 07 Praha 8–Libeň
CZECH REPUBLIC
E-mail: kramosil@uivt.cas.cz