УДК 519.213

The Study of Discrete Probabilistic Distributions of Random Sets of Events Using Associative Function

Natalia A. Lukyanova*
Daria V. Semenova†
Institute of Mathematics and Computer Science
Siberian Federal University
Svobodny, 79, Krasnoyarsk, 660041
Russia

Received 10.05.2014, received in revised form 25.08.2014, accepted 04.09.2014

In this work the class of discrete probabilistic distributions of the II-nd type of random sets of event is investigated. As the tool for constructing of such probabilistic distributions it is offered to use associative functions. There is stated a new approach to define a discrete probabilistic distribution of the II-nd type of a random set on a finite set of N events on the basis of obtained recurrence relation and a given associative function. Advantage of the offered approach is that for definition of probabilistic distribution instead of a totality of 2^N probabilities it is enough to know N probabilities of events and a type of associative function. In this paper an $|X|$-ary covariance of a random set of events is considered. It is a measure of the additive deviation of the events from the independent situation. The process of recurrent constructing a probabilistic distribution II-nd type is demonstrated by the example of three associative functions. The proof of the legitimacy / illegitimacy the obtained distribution by passing to the probabilistic distribution of the I-st type by formulas of Möbius is given. Theorems that establish the form and conditions of the legitimacy of the resulting probabilistic distributions are proved. $|X|$-ary covariances of random sets of events are found.

Keywords: random set of events, discrete probabilistic distributions, associative function, $|X|$-ary covariance.

Introduction

The theory of random sets is considered as natural generalization of the theory of random vectors which play a key role in the multidimensional statistical analysis. Random sets of data can be considered as partial/incomplete observations which are often met in modern technological society [1]. The central object of our research is the specific random set, namely — a random finite set of events [2]. Random sets of events allow to reveal the general statistical regularities of distribution of events in various systems of objects of the non-numerical nature. Probabilistic distribution of a random set of events is a convenient mathematical apparatus for the description of all ways of interaction of elements this set among themselves. In work the problem of creation of probabilistic distributions of random sets of events is investigated and the approach of the solution of this problem by means of the device of associative functions is offered.

*nata00sfu@gmail.com
†dariasdv@gmail.com

Organization of article is the following. Basic information about the key objects of research are given in Section 1. They are random sets of events and their characterizing probabilistic distributions. Also in Section 1 the instrument of research is described, it is the mathematical apparatus of associative functions. In Section 2 the essence of the offered recurrent approach for constructing discrete probabilistic distribution of the II-nd type of random sets of events on the basis of a given associative functions is stated. The type of the corresponding recurrence relation is defined. The offered approach is shown on three known associative functions. The theorems, which establish a form and conditions of legitimacy of received probabilistic distributions, are provided. $|X|$-ary covariances of random sets of events are found. In Section 3 the proved theorems and their consequences are illustrated by examples of the triplet and 4-plet events.

1. Base concepts and designations

1. Random set of events

Random objects of any nature are considered at the same time with random variables in the probability theory and its appendices, for example, random points, vectors, functions, fields, sets and sets of sets. The concept of a random element is used to describe this type of objects [3].

Definition 1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $(U, \mathcal{A})$ be a measurable space, where U is any set, and $\mathcal{A}$ is some σ-algebra of its subsets. Let say that function $K = K(w)$, defined on Ω and accepting values in U, is a $\mathcal{F}/\mathcal{A}$-measurable function, or a random element (with values in U) if for any $A \in \mathcal{A}$ is right $\{\omega : K(\omega) \in A\} \in \mathcal{F}$.*

Note that in a case when U is a finite set, it is possible to restrict to the notion of an algebra of subsets 2^U.

Let fix a finite set of events $\mathfrak{X} \subset \mathcal{F}$. Random element which values are subsets of a finite set $\mathfrak{X}$, i.e. elements of $2^{\mathfrak{X}}$, will be called as the random set of events [1] defined on a finite set $\mathfrak{X} \subset \mathcal{F}$.

Definition 2. *Random set of events K on a finite set of events $\mathfrak{X} \subset \mathcal{F}$ is defined as a mapping $K : \Omega \to 2^{\mathfrak{X}}$ measurable with respect to the pair algebras $\left(\mathcal{F}, 2^{2^{\mathfrak{X}}}\right)$ in the sense that for every $X \in 2^{2^{\mathfrak{X}}}$ there is the preimage $K^{-1}(X) \in \mathcal{F}$ such that $\mathbf{P}(X) = \mathbf{P}(K^{-1}(X))$.*

Expression $K(\omega) = \{x \in \mathfrak{X} : \omega \in x\}$ can be interpreted as «random set of occurring events», since an elementary outcome of the experiment $\omega \in \Omega$ is associated with a subset of the events $X \subseteq \mathfrak{X}$ which contains all the events which occurred in this trial.

2. Probabilistic distribution of a random set of events

Random set of events K given on a finite set of events $\mathfrak{X}$ is defined by its discrete probabilistic distribution. If the power of the considered set of events $|\mathfrak{X}| = N < \infty$, then there are 2^N species probabilistic dependencies between the events of this set, i.e. as much as it has a set of subsets.

Discrete probabilistic distribution (further, the probabilistic distribution) of the random set of events K given on a finite set of selected events $\mathfrak{X} \subset \mathcal{F}$ is a complex of 2^N values of the probabilistic measure $\mathbf{P}$ on events from $2^{\mathfrak{X}}$. As it is known, such a distribution can be defined by six equivalent ways [2, 4, 5]. In this paper only two of them are studied.

PI. The probabilistic distribution of the I-st type of a random set of events K on $\mathfrak{X}$ is a complex of $\{p(X), X \subseteq \mathfrak{X}\}$ from 2^N probabilities of the form

$$p(X) = \mathbf{P}(K = X) = \mathbf{P}\left(\left(\bigcap_{x \in X} x\right) \bigcap \left(\bigcap_{x \in X^c} x^c\right)\right), \tag{1}$$

where $X^c = \mathfrak{X} \setminus X$, $x^c = \Omega \setminus x$.

The probabilistic distribution of the I-st type is always legitimate, i.e. has properties

$$0 \leqslant p(X) \leqslant 1, \quad X \subseteq \mathfrak{X}, \tag{2}$$

$$\sum_{X \subseteq \mathfrak{X}} p(X) = 1. \tag{3}$$

PII. The probabilistic distribution of the II-nd type of a random set of events K on $\mathfrak{X}$ is a complex of $\{p_X, X \subseteq \mathfrak{X}\}$ from 2^N probabilities of the form

$$p_X = \mathbf{P}(X \subseteq K) = \mathbf{P}\left(\bigcap_{x \in X} x\right), \quad X \subseteq \mathfrak{X}. \tag{4}$$

The probabilistic distribution of the II-nd type $\{p_X, X \subseteq \mathfrak{X}\}$ of a random set of events K on $\mathfrak{X}$ satisfies the system that consists of 2^N inequations of Fréchet-Hoeffding:

$$0 \leqslant p_X^- \leqslant p_X \leqslant p_X^+ \leqslant 1,$$

where $p_X^- = \max\left\{0, 1 - \sum_{x \in X}(1 - p_x)\right\}$ is Fréchet-Hoeffding's lower bound, $p_X^+ = \min_{x \in X} p_x$ is Fréchet-Hoeffding's upper bound.

Remark 1. In the theory of random events [2,4,5,7,8] designation $\varnothing$ is used for $\mathfrak{X}^c = \Omega \setminus \mathfrak{X}$. Event $\varnothing = \bigcap_{x \in \mathfrak{X}} x^c$ means that none event has occurred from $\mathfrak{X}$. In the probabilistic distribution of the II-nd type always $p_\varnothing = 1$.

Remark 2. Only the distribution of the I-st type has the property of the normalization that the usually for the probability distribution (3). That property normalization follows from the fact that the relevant events $\left(\bigcap_{x \in X} x\right) \bigcap \left(\bigcap_{x \in X^c} x^c\right)$ form a partition of the space of elementary events. And since the event $\bigcap_{x \in X} x$ do not form a partition, but only covering Ω, the ratio of normalization for the probabilistic distribution of the II-nd type is not satisfied, the sum of the probabilities of these events is always greater than unity.

Probabilistic distributions of the I-st and the II-nd types are connected by reciprocal formulas of Möbius inversion [2, 4]

$$p_X = \sum_{Y \in 2^{\mathfrak{X}}:\, X \subseteq Y} p(Y), \tag{5}$$

$$p(X) = \sum_{Y \in 2^{\mathfrak{X}}:\, X \subseteq Y} (-1)^{|Y|-|X|} p_Y, \tag{6}$$

for all $X \in 2^{\mathfrak{X}}$.

If the probabilistic distribution of the I-st type $\{p(X), X \subseteq \mathfrak{X}\}$ is given, then on the formula (5) we always get the probabilistic distribution of the II-nd type $\{p_X, X \subseteq \mathfrak{X}\}$. However, the transformation (6) a given set of 2^N numbers $\{p_X,\ X \subseteq \mathfrak{X}\}$ satisfying to Fréchet-Hoeffding bounds can lead to the probabilistic distribution of the I-st type with negative values. This fact is demonstrated by the following example.

Example 1. Consider a random set given by the triplet events $\mathfrak{X} = \{x, y, z\}$. Let his probabilistic distribution of the II-nd sort be known.

$$\{p_\varnothing, p_x, p_y, p_z, p_{xy}, p_{xz}, p_{yz}, p_{xyz}\} = \{1,\ 0.375,\ 0.75,\ 0.625,\ 0.243,\ 0.19,\ 0.429,\ 0.118\}. \quad (7)$$

Note that all probabilities of (7) satisfy to to Fréchet-Hoeffding bounds. Example, for $p_{xz} = \mathbf{P}(x \cap z) = 0.19$ have

$$\max\{p_x + p_z - 1, 0\} \leqslant p_{xz} \leqslant \min\{p_x, p_z\} \Rightarrow 0 \leqslant p_{xz} \leqslant 0.375.$$

Application of the formulas (6) leads to the probabilistic distribution of the I-st type

$$p(x) = \mathbf{P}(x \cap y^c \cap z^c) = p_x - p_{xy} - p_{xz} + p_{xyz} = 0.375 - 0.243 - 0.19 + 0.118 = 0.06;$$
$$p(y) = \mathbf{P}(x^c \cap y \cap z^c) = p_y - p_{xy} - p_{yz} + p_{xyz} = 0.75 - 0.243 - 0.429 + 0.118 = 0.196;$$
$$p(z) = \mathbf{P}(x^c \cap y^c \cap z) = p_z - p_{xz} - p_{yz} + p_{xyz} = 0.625 - 0.19 - 0.429 + 0.118 = 0.124;$$
$$p(xy) = \mathbf{P}(x \cap y \cap z^c) = p_{xy} - p_{xyz} = 0.243 - 0.118 = 0.125;$$
$$p(xz) = \mathbf{P}(x \cap y^c \cap z) = p_{xz} - p_{xyz} = 0.19 - 0.118 = 0.072;$$
$$p(yz) = \mathbf{P}(x^c \cap y \cap z) = p_{yz} - p_{xyz} = 0.429 - 0.118 = 0.311;$$
$$p(xyz) = \mathbf{P}(x \cap y \cap z) = p_{xyz} = 0.118;$$
$$p(\varnothing) = \mathbf{P}(x^c \cap y^c \cap z^c) = 1 - p_x - p_y - p_z + p_{xy} + p_{xz} + p_{yz} - p_{xyz} = -0.006.$$

Such a distribution is illegitimate because $p(\varnothing) < 0$.

Definition 3. *Let say that a random set of events K has a legitimate probabilistic distribution of the II-nd type, unless the probabilistic distribution of the I-st type obtained by formulas of Möbius inversion (6), is legitimate.*

In general, the number of parameters defining the probabilistic distribution of a random set of events depends on the power of $\mathfrak{X}$, since each set of N events characterized by a set of 2^N probabilities. The task of describing of probabilistic distributions of random sets of events through a smaller number of parameters is one of the main problems of the study of random sets of events. For example, to determine the probabilistic distribution of a uniform random set is required only one parameter $N = |\mathfrak{X}|$. Class of m-dependent distributions of a random set of events that are defined through a fixed probability p_X, where $|X| \leqslant m < N$, was studied earlier. General idea of modern approaches [2, 7, 8] is an expression of the probability of the intersection of the set of events functionally by means of probabilities of the events themselves which leads to a decrease in the number of parameters which are needed to construct the probabilistic distributions of random sets of events. In this paper, recurrent approach to the construction of probabilistic distributions of random sets of events is proposed. In the approach are used the associative functions which will bind the probability of the intersection of events $\mathbf{P}\left(\bigcap_{x \in X} x\right)$ with probabilities of the events themselves $\mathbf{P}(x)$, $x \in X$, $X \subseteq \mathfrak{X}$.

3. Associative functions

Definition 4. *Associative function in the theory of random set of events* $\mathrm{AF} : [0,1]^2 \to [0,1]$ *is defined as a two-place function satisfying to the following properties.*

A1. *Boundary conditions:* $\mathrm{AF}(a,0) = \mathrm{AF}(0,a) = 0$, $\mathrm{AF}(a,1) = \mathrm{AF}(1,a) = a$, $\forall a \in [0,1]$.

A2. *Monotonicity:* $\forall a_1, a_2, b_1, b_2 \in [0,1]$ *such that* $a_1 \leqslant a_2$, $b_1 \leqslant b_2$ *is right* $\mathrm{AF}(a_1,b_1) \leqslant \mathrm{AF}(a_2,b_2)$.

A3. *Commutativity:* $\mathrm{AF}(a,b) = \mathrm{AF}(b,a)$, $\forall a,b \in [0,1]$.

A4. *Associativity:* $\mathrm{AF}(\mathrm{AF}(a,b),c) = \mathrm{AF}(a,\mathrm{AF}(b,c))$, $\forall a,b,c \in [0,1]$.

A5. *Condition of Lipschitz's continuity:* $\mathrm{AF}(c,b) - \mathrm{AF}(a,b) \leqslant c - a$, $a \leqslant c$, $a,b,c \in [0,1]$.

Geometrically, the graph of associative function is a surface which is spanned by the quadrilateral whose vertices are $(0,0,0)$, $(0,1,0)$, $(1,0,0)$ and $(1,1,1)$. According to property **A2** associative function increases both vertically and horizontally [9, 10], and which is symmetric with respect to the plane $a = b$ by the property of **A3**. Properties of **A1**–**A3**, **A5** guarantee that the value of function $\mathrm{AF}(a,b)$ will possess properties of probability. Property **A4** allows to pass to n-place function recurrently.

Note that the properties of associative function **A1**–**A4** correspond to the properties of triangular norm (t -norm). The notion of t-norm is introduced and used in the theory of probabilistic metric spaces [11, 12]. At the present time t-norms are widely used in fuzzy logic as a conjunction operation. They are of interest for fuzzy logic because they retain the basic properties of ligaments «and» which are executed at the same time, namely, commutativity, monotonicity, associativity and boundedness, and thus, they serve as a natural generalization of the classical conjunction of systems for multi-valued arguments [10, 13, 14]. Class of binary operations which are associated with t-norms are copula [9], introduced by A. Sklar in 1959. Models based on copula are extremely important for modeling multivariate observations. In fact copula contains all the information about the nature of dependence between random variables, and does not contain information about the marginal distributions. Dependence between random variables does not place in the marginal distributions. As a result, information about the marginals and the dependence between them are separated from each other by the copula [9]. Thus, the model with the use of the copula defines function of the joint distribution of random variables in the form of two parts that allows to consider separately one-dimensional marginal distributions and the function which is responsible for their dependence [9]. In our case use copulas allows to present probability of product of sets of events through probabilities of events and the function AF that reflecting them interrelation. Some families of t-norms are known as family of copulas under different names. In [10] proved that **A5** gives the conditions under which t-norm is a copula. On the other hand [10], any commutative and associative copula is t-norm. Known [9, 10, 14] that any copula satisfies the Fréchet-Hoeffding's bounds that, as a result, also valid for associative functions.

Thus, taking into account [9, 10, 14] and the above stated, the definition 4 can be expressed as follows: associative function is a continuous t-norm satisfying the Lipschitz's condition, or, equivalently, is associative and commutative copula. Such treatment of concept of associative function is used below in this paper.

2. Recurrent formation of probabilistic distributions of random sets of events by associative functions

Let use the apparatus of associative functions to probabilistic distributions of random sets of events. It is proposed to consider probabilities of events as arguments of associative function; their number is equal with the power of a basic set. Thus, associative functions will connect probabilities of the intersection of a set of events $p_X = \mathbf{P}\left(\bigcap_{x\in X} x\right)$ with probabilities of events themselves $p_x = \mathbf{P}(x)$, $x \in X$, $X \subseteq \mathfrak{X}$. Let proceed to describe the approach.

Input:

- set of events $\mathfrak{X}$, $|\mathfrak{X}| = N$;
- N probabilities of events p_x, $x \in \mathfrak{X}$;
- associative function $\mathrm{AF}(a, b)$.

Output: probabilistic distribution of the II-nd type $\{p_X, X \subseteq \mathfrak{X}\}$ of random sets of events K on $\mathfrak{X}$.

Main idea: proceeding from known probabilities of events of $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$, the formation of probabilities of the intersection of sets of events p_X to carry out consistently according to a recurrent formula:

$$p_X = \mathbf{P}\left(\bigcap_{x\in X} x\right) = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y\in X\setminus\{x\}} y\right)\right), \qquad X \subseteq \mathfrak{X}, \qquad |X| > 1. \tag{8}$$

For example,

$$p_{xy} = \mathbf{P}\left(x \cap y\right) = \mathrm{AF}\left(p_x, p_y\right),$$
$$p_{xyz} = \mathbf{P}\left(x \cap y \cap z\right) = \mathrm{AF}\left(p_x, \mathrm{AF}(p_y, p_z)\right) = \mathrm{AF}\left(p_x, \mathbf{P}\left(y \cap z\right)\right),$$

and further similarly. Let remind that the probability of the II-nd type for an empty set of events is always known and equal 1.

Formula (8) allows us to construct a probabilistic distribution II-nd type of a random set of events where the input parameters are the N probabilities of events and a type of associative function. $2^N - N - 1$ probabilities of II-nd type is not enough to complete set. They are formed by the formula (8). All the probabilities satisfy to Fréchet-Hoeffding bounds. However, these distributions can turn out illegitimate in the sense of Definition 3. Hence for each family of associative functions it is necessary to determine the conditions of legitimacy of constructed probabilistic distributions.

In [2] a convenient tool for analyzing the structures of probabilistic dependency was offered. It is a $|X|$-ary covariance of any set of events $X \subset \mathcal{F}$, $|X| > 1$:

$$\mathrm{Kov}_X = \mathbf{P}\left(\bigcap_{x\in X} x\right) - \prod_{x\in X} \mathbf{P}(x) = p_X - \prod_{x\in X} p_x. \tag{9}$$

$|X|$-ary covariance Kov_X turns into 0 when the events from X are independent; becomes greater than 0 when the events of the set X are statistically attracted; and becomes less than 0 when the events of the set X are statistically repelled. It is known [2] that for the set $\mathfrak{X}$ consisting of N events there is a set of $2^N - N - 1$ of $|X|$-ary covariances which are determined by the probabilistic distribution of the II type of a random set of events at $\mathfrak{X}$.

Taking into account (8) formula (9) will take the form

$$\mathrm{Kov}_X = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y \in X \setminus \{x\}} y\right)\right) - \prod_{x \in X} p_x. \tag{10}$$

Let Illustrate the proposed recurrent approach on three associative functions studied in the works [9, 10, 13, 14]:

- $\mathrm{AF}(a, b) = a \cdot b$;
- $\mathrm{AF}(a, b) = \min\{a, b\}$;
- $\mathrm{AF}(a, b) = \max\{a + b - 1, 0\}$.

Show for each of them to which the probabilistic distribution of a random set of events it leads. Research the legitimacy of this distribution.

Theorem 1. *Let there be given the probabilities of events $p_x = \mathbf{P}(x) > 0$, $x \in \mathfrak{X}$. Then the associative function*

$$\mathrm{AF}(a, b) = a \cdot b \tag{11}$$

determines independently-point random set of events with a legitimate probability distribution of the II-nd type.

Proof. The proof follows directly from the definition of independently-point random set of events [2, 15, 16]. Based on this definition, the values of the independently-point random set K are subsets $X \subseteq \mathfrak{X}$ of occurring events which independent in the sum total. Since the events from $\mathfrak{X}$ are independent in the sum total then for all $X \subseteq \mathfrak{X}$ is right

$$p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) = \prod_{x \in X} \mathbf{P}(x) = \prod_{x \in X} p_x. \tag{12}$$

On the other hand, for $X \subseteq \mathfrak{X}$ by the formula (8) it will be obtained

$$p_X = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y \in X \setminus \{x\}} y\right)\right) = \prod_{x \in X} p_x. \tag{13}$$

From (12) and (13) follows that an associative function (11) determines independently-point random set of events. Known [2, 15, 16] that for an independently-point random set of events the probabilistic distribution of the I-st type has the form

$$p(X) = \prod_{x \in X} \mathbf{P}(x) \prod_{x \in X^c} \mathbf{P}(x^c) = \prod_{x \in X} p_x \prod_{x \in X^c} (1 - p_x), \quad X \subseteq \mathfrak{X},$$

and this distribution is always legitimate.

Thus, the formula (8) with associative function (11) always allows to construct one legitimate distribution which determines independently-point random set of events. Thus, the theorem is proved. □

Corollary 1. *For independent-point random set of events with a legitimate probabilistic distribution of the II-nd type defined associative function (11), all $|X|$-ary covariances are equal zero:* $\mathrm{Kov}_X = 0$.

Proof. Indeed [2,16], from (10) and (13) for all $X \subseteq \mathfrak{X}$, it will be obtained

$$\mathrm{Kov}_X = p_X - \prod_{x\in X} p_x = \prod_{x\in X} p_x - \prod_{x\in X} p_x = 0.$$

□

Theorem 2. *Let there be given the probabilities of events $p_x = \mathbf{P}(x) > 0$, $x \in \mathfrak{X}$. Then the associative function*

$$\mathrm{AF}(a,b) = \min\{a,b\} \tag{14}$$

defines a random set of embedded events with legitimate probabilistic distribution of the II-nd type.

Proof. Let put in order the events in $\mathfrak{X} = \{x_1,\ x_2,\ \dots, x_N\}$ in ascending order of their probabilities $p_{x_1} \leqslant p_{x_2} \leqslant \dots \leqslant p_{x_N}$. Application of the formula (8) gives:

$$\begin{aligned} p_{x_i x_j} &= \mathbf{P}\left(x_i \cap x_j\right) = \min\left\{p_{x_i},\ p_{x_j}\right\} = p_{x_i}, \quad \forall\ i < j; \\ p_{x_i x_j x_k} &= \mathbf{P}\left(x_i \cap x_j \cap x_k\right) = \min\left\{p_{x_i},\ p_{x_j},\ p_{x_k}\right\} = p_{x_i}, \quad \forall\ i < j < k. \end{aligned} \tag{15}$$

From (15) follows that $x_i \subseteq x_j \subseteq x_k$. Really, if $x \subseteq y$, then $\mathbf{P}(x\cap y) = \mathbf{P}(x)$; if $x \subseteq y$ and $y \subseteq z$, then $x \subseteq y \subseteq z$ и $\mathbf{P}(x\cap y\cap z) = \mathbf{P}(x)$.

Consider the ordered set of indices $I \subseteq \{1,\ 2, \dots,\ N\}$ corresponding to a subset $X \subseteq \mathfrak{X}$ which elements are ordered in ascending order probabilities of events. Then the formula (8) takes the form

$$p_X = \mathbf{P}\left(\bigcap_{i\in I} x_i\right) = \min_{i\in I}\{p_{x_i}\} = p_{x_m}, \text{ где } m = \min_{i\in I} i. \tag{16}$$

Formula (16) defines the probabilistic distribution of the random set K given on the set $\mathfrak{X} \subset \mathcal{F}$ from N embedded events $x_1 \subseteq x_2 \subseteq \dots \subseteq x_N$.

Let turn to the question of the legitimacy of the resulting probabilistic distribution of the II-nd type of a random set of embedded events (16). By formulas of Möbius inversion (6) move to the probabilistic distribution of the I-st type and obtain $N+1$ non-zero probabilities

$$\begin{gathered} p(\mathfrak{X}) = \mathbf{P}\left(\bigcap_{i=\overline{1,N}} x_i\right) = p_{x_1}; \quad p(\mathfrak{X}\setminus\{x_1\}) = \mathbf{P}\left(\bigcap_{i=\overline{2,N}} x_i\right) - \mathbf{P}\left(\bigcap_{i=\overline{1,N}} x_i\right) = p_{x_2} - p_{x_1}; \\ p(\mathfrak{X}\setminus\{x_1,x_2\}) = \mathbf{P}\left(\bigcap_{i=\overline{3,N}} x_i\right) - \mathbf{P}\left(\bigcap_{i=\overline{2,N}} x_i\right) = p_{x_3} - p_{x_2}; \\ \dots \\ p(\mathfrak{X}\setminus\{x_1,x_2,\dots,x_{k-1}\}) = \mathbf{P}\left(\bigcap_{i=\overline{k,N}} x_i\right) - \mathbf{P}\left(\bigcap_{i=\overline{k-1,N}} x_i\right) = p_{x_k} - p_{x_{k-1}}; \\ \dots \\ p(x_N) = \mathbf{P}\left(x_N\right) - \mathbf{P}\left(x_{N-1}\cap x_N\right) = p_{x_N} - p_{x_{N-1}}. \\ p(\oslash) = \mathbf{P}\left(\bigcap_{i=\overline{1,N}} x_i^c\right) = 1 - p_{x_N}. \end{gathered} \tag{17}$$

The remaining $2^N - N - 1$ probabilities are equal to zero.

From (17) can be seen that the distribution of the I-st type satisfies the conditions (2) and (3), hence the formula (8) with associative function (14) always allows to construct one legitimate distribution which defines a random set of embedded events. Thus, the theorem is proved. □

Corollary 2. *For a random set of embedded events with legitimate probabilistic distribution of the II-nd type which is defined by associative function (14), all $|X|$-ary covariances are positive:* $\mathrm{Kov}_X > 0$.

Proof. Let consider the ordered set of indices $I \subseteq \{1, 2, \dots, N\}$ corresponding to some subset $X \subseteq \mathfrak{X}$, whose elements are ordered in ascending order probabilities of events. Then it follows from (10) and (13) for all $X \subseteq \mathfrak{X}$, it will be obtained

$$\mathrm{Kov}_X = p_X - \prod_{x\in X} p_x = p_{x_m} - \prod_{i\in I} p_{x_i} = p_{x_m}\left(1 - \prod_{i\in I\setminus\{m\}} p_{x_i}\right) > 0, \text{ где } m = \min_{i\in I} i.$$

□

Theorem 3. *Let there be given the probabilities of events $p_x = \mathbf{P}(x) > 0$, $x \in \mathfrak{X}$. Associative function*

$$\mathrm{AF}(a,b) = \max\{a+b-1, 0\} \tag{18}$$

defines

1) *a random set with disjoint structure of dependency with a legitimate probabilistic distribution if the probabilities of the events $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$ satisfy the inequality*

$$\sum_{x\in\mathfrak{X}} p_x \leqslant 1, \tag{19}$$

2) *a random set of events with a legitimate probabilistic distribution if the probabilities of the events $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$ satisfy the inequalities*

$$|\mathfrak{X}| - 1 \leqslant \sum_{x\in\mathfrak{X}} p_x \leqslant |\mathfrak{X}|. \tag{20}$$

Proof. In [13,14] obtained form n-place function

$$\mathrm{AF}(a_1, \dots, a_n) = \max\left\{\sum_{i=1}^{n} a_i - n + 1, 0\right\}. \tag{21}$$

For all $X \subseteq \mathfrak{X}$, $|X| > 1$ from (8) and (21) follow

$$p_X = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y\in X\setminus\{x\}} y\right)\right) = \max\left\{\sum_{x\in X} p_x - |X| + 1, 0\right\}. \tag{22}$$

Consider the following situations.

1. Suppose that in (22) all the probabilities p_X equal to zero, i.e.

$$\max\left\{\sum_{x\in X} p_x - |X| + 1, 0\right\} = 0$$

for all $X \subseteq \mathfrak{X}$ such that $|X| > 1$. By formulas of Möbius inversion (6) move to the probabilistic distribution of the I-st type

$$\begin{aligned} &p(x) = p_x, \text{ for all } x \in \mathfrak{X}, \\ &p(X) = 0, \text{ for all } X \subseteq \mathfrak{X} \text{ such that } |X| > 1, \\ &p(\varnothing) = 1 - \sum_{x \in \mathfrak{X}} p_x. \end{aligned} \tag{23}$$

It is obvious that the received distribution (23) will satisfy to property (2), if $p(\varnothing) \geqslant 0$. Consequently, the distribution will be legitimate if and only if $\sum\limits_{x \in \mathfrak{X}} p_x \leqslant 1$.

2. Suppose that in (22) all $p_X \neq 0$, $X \subseteq \mathfrak{X}$, $|X| > 1$. Then from (22) and condition (2) it follows $0 < \sum\limits_{x \in X} p_x - |X| + 1 \leqslant 1$ or, that is equivalent, $|X| - 1 < \sum\limits_{x \in X} p_x \leqslant |X|$.

By formulas of Möbius inversion (6) with the use of the principles set-summation [4] will be obtained a legitimate probabilistic distribution of the I-st type:

$$\begin{aligned} &p(X) = 0, \text{ for all } X \subseteq \mathfrak{X} \text{ such that } |X| < |\mathfrak{X}| - 1, \\ &p(\mathfrak{X} \setminus \{x\}) = 1 - p_x, \text{ for all } x \in \mathfrak{X}, \\ &p(\mathfrak{X}) = \sum_{x \in \mathfrak{X}} p_x - |\mathfrak{X}| + 1. \end{aligned} \tag{24}$$

Thus, associative function (18) defines a random set of events with a legitimate probabilistic distribution (24) if the probabilities of the events $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$ satisfy the system of $2^{|\mathfrak{X}|} - 1$ inequalities

$$|X| - 1 < \sum_{x \in X} p_x \leqslant |X|, \ X \subseteq \mathfrak{X}. \tag{25}$$

Show that if the inequality (20) holds then the whole system is valid (25). Really, suppose that $|\mathfrak{X}| - 1 \leqslant \sum\limits_{x \in \mathfrak{X}} p_x \leqslant |\mathfrak{X}|$. Upper bound $\sum\limits_{x \in X} p_x \leqslant |X|$ for all $X \subseteq \mathfrak{X}$ follows from the properties of probability $p_x \leqslant 1$. Lower bound for $X \subset \mathfrak{X}$ is obtained as follows

$$\sum_{x \in \mathfrak{X}} p_x = \sum_{x \in X} p_x + \sum_{y \in \mathfrak{X} \setminus X} p_y \geqslant |\mathfrak{X}| - 1 \Rightarrow \sum_{x \in X} p_x \geqslant |\mathfrak{X}| - 1 - \sum_{y \in \mathfrak{X} \setminus X} p_y \Rightarrow$$

$$\Rightarrow \sum_{x \in X} p_x \geqslant |\mathfrak{X}| - 1 - (|\mathfrak{X}| - |X|) \Rightarrow \sum_{x \in X} p_x \geqslant |X| - 1.$$

From (25) also implies that for receiving the legitimate distribution it is required that all $p_X > 0$, except maybe $p_{\mathfrak{X}}$. Really, suppose that for some X such that $1 < |X| < |\mathfrak{X}|$, the probability p_X is equal to zero, and all other probabilities obtained with the use of associative function (18) are different from zero. From (22) and (25) it follows that $\sum\limits_{x \in X} p_x = |X| - 1$. Add to the set X any event y of $\mathfrak{X} \setminus X$. Then for the set $X \cup \{y\}$ it follows from (25)

$$|X| \leqslant \sum_{x \in X} p_x + p_y \leqslant |X| + 1 \Rightarrow |X| \leqslant |X| - 1 + p_y \leqslant |X| + 1 \Rightarrow 1 \leqslant p_y \leqslant 2,$$

that contradicts the property of probability $0 \leqslant p_y \leqslant 1$.

Consider the situation when $\sum_{x\in\mathfrak{X}} p_x = |\mathfrak{X}| - 1$, i.e. $p_{\mathfrak{X}} = 0$. From (25) implies that all $p_X > 0$ for $X \subset \mathfrak{X}$. By formulas of Möbius inversion (6) with the use of the principles set-summation [4] will be obtained a legitimate probabilistic distribution of the I-st type:

$$\begin{aligned} &p(X) = 0, \text{ for all } X \subseteq \mathfrak{X} \text{ such that } |X| < |\mathfrak{X}| - 1, \\ &p(\mathfrak{X} \setminus \{x\}) = \sum_{y\in\mathfrak{X}\setminus\{x\}} p_y - |\mathfrak{X}| + 2, \text{ for all } x \in \mathfrak{X}, \\ &p(\mathfrak{X}) = 0. \end{aligned}$$

Thus, the theorem is proved. □

From Theorem 3 it can be drawn the following conclusion: the recurrent construction of a legitimate probabilistic distribution of a random set of events with the use of associative function (18) is possible only when performing certain restrictions on input probabilities of events. Thus, there are only three types of resultant random sets of events with the corresponding probabilistic distributions:

1) a random set of disjoint events, if the condition is performed (19);

2) a random set of events which takes the values with non-zero probability only on a subsets of power of $|\mathfrak{X}| - 1$ and $|\mathfrak{X}|$, if it is right $|\mathfrak{X}| - 1 < \sum_{x\in\mathfrak{X}} p_x \leqslant |\mathfrak{X}|$;

3) a random set of events which takes the values with non-zero probability only on a subsets of power of $|\mathfrak{X}| - 1$, if it is right $\sum_{x\in\mathfrak{X}} p_x = |\mathfrak{X}| - 1$.

Corollary 3. *Let random set of events with a legitimate probabilistic distribution of the II-nd type is defined by associative function (18); the probabilities of the events* $0 < p_x = \mathbf{P}(x) < 1$, $x \in \mathfrak{X}$ *satisfy the inequalities*

$$|\mathfrak{X}| - 1 \leqslant \sum_{x\in\mathfrak{X}} p_x < |\mathfrak{X}|.$$

Then all $|X|$*-ary covariances are negative:* $\mathrm{Kov}_X < 0$, $X \subseteq \mathfrak{X}$.

Proof. Let write out the $|X|$-ary covariances recurrently for the probabilistic distribution of the II-nd type which is defined by associative function (18).

$$\mathrm{Kov}_{\{x,y\}} = p_x + p_y - 1 - p_x\, p_y = p_x(1 - p_y) - (1 - p_y) = (p_x - 1)(1 - p_y) < 0.$$

$$\mathrm{Kov}_{\{x,y,z\}} = p_x + p_y + p_z - 2 - p_x\, p_y\, p_z = (p_x + p_y - 1 - p_x\, p_y) + (p_z + p_x\, p_y - 1 - p_x\, p_y\, p_z) = \\ = \mathrm{Kov}_{\{x,y\}} + (p_z - 1)(1 - p_x\, p_y) < 0.$$

$$\mathrm{Kov}_{\{x,y,z,v\}} = p_x + p_y + p_z + p_v - 3 - p_x\, p_y\, p_z\, p_v = (p_x + p_y + p_z - 2 - p_x\, p_y\, p_z) + \\ +(p_v + p_x\, p_y\, p_z - 1 - p_x\, p_y\, p_z\, p_v) = \mathrm{Kov}_{\{x,y,z\}} + (p_v - 1)(1 - p_x\, p_y\, p_z) < 0.$$

$$\dots$$

$$\mathrm{Kov}_{\mathfrak{X}\setminus\{x\}} = \mathrm{Kov}_{\mathfrak{X}\setminus\{x,v\}} + p_v + \prod_{t\in\mathfrak{X}\setminus\{x,v\}} p_t - 1 - \prod_{t\in\mathfrak{X}\setminus\{x\}} p_t =$$

$$= \mathrm{Kov}_{\mathfrak{X}\setminus\{x,v\}} + (p_v - 1)\left(1 - \prod_{t\in\mathfrak{X}\setminus\{x,v\}} p_t\right) < 0, \text{ for all } v \in \mathfrak{X} \setminus \{x\}.$$

Finally,

$$\mathrm{Kov}_{\mathfrak{X}} = \mathrm{Kov}_{\mathfrak{X}\setminus\{v\}} + p_v + \prod_{t\in\mathfrak{X}\setminus\{v\}} p_t - 1 - \prod_{t\in\mathfrak{X}} p_t =$$

$$= \mathrm{Kov}_{\mathfrak{X}\setminus\{v\}} + (p_v - 1)\left(1 - \prod_{t\in\mathfrak{X}\setminus\{v\}} p_t\right) < 0, \text{ for all } v \in \mathfrak{X}.$$

□

Remark 3. Clearly, for random set of nonoverlapping events $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$ (19) all $|X|$-ary covariances are negative: $\mathrm{Kov}_X < 0$.

3. Examples

Let illustrate the proved theorems and their consequences with examples.

Example 2. Consider a random set given by the triplet events $\mathfrak{X} = \{x, y, z\}$ with known probabilities of events from Example 1 $p_x = 0.375$, $p_y = 0.75$, $p_z = 0.625$.

The results of the construction of probabilistic distributions of the II-nd type by the recurrent relation (8) with associative functions (11), (14), as well as the calculation of the corresponding probabilistic distributions of the I-st type and ary covariances are summarized in Tab. 1.

Table 1. Probabilistic distributions and ary covariances

$2^{\mathfrak{X}}$	AF $= ab$			AF $= \min\{a,b\}$		
	p_X	$p(X)$	Kov_X	p_X	$p(X)$	Kov_X
$\varnothing$	1	0.059	0	1	0.25	0
$\{x\}$	0.375	0.035	0	0.375	0	0
$\{y\}$	0.75	0.176	0	0.75	0.125	0
$\{z\}$	0.625	0.098	0	0.625	0	0
$\{x, y\}$	0.281	0.105	0	0.375	0	0.094
$\{x, z\}$	0.234	0.058	0	0.375	0	0.141
$\{y, z\}$	0.469	0.293	0	0.625	0.25	0.156
$\{x, y, z\}$	0.176	0.176	0	0.375	0.375	0.1992

Example 3. Consider a random set given by 4-plet events $\mathfrak{X} = \{x, y, z, v\}$, $|\mathfrak{X}| = 4$. And let the known probabilities of events p_x, p_y, p_z, p_v.

The results of the construction of probabilistic distributions of the II-nd type for the three possible situations by the recurrent relation (8) with associative function (18), as well as the calculation of the corresponding probabilistic distributions of the I-st type and ary covariances are summarized in Tab. 2. For disjoint dependent structure it has $\sum_{x\in\mathfrak{X}} p_x = 0.938 \leqslant 1$.

For the second situation the condition will be satisfied with: $3 < \left(\sum_{x\in\mathfrak{X}} p_x = 3.25\right) \leqslant 4$.

For the third situation it is satisfied $\sum_{x\in\mathfrak{X}} p_x = 3$.

Table 2. Probabilistic distributions for $\mathrm{AF} = \max\{a + b - 1, 0\}$

$2^{\mathfrak{X}}$	$\sum_{x\in\mathfrak{X}} p_x \leqslant 1$			$3 < \sum_{x\in\mathfrak{X}} p_x \leqslant 4$			$\sum_{x\in\mathfrak{X}} p_x = 3$		
	p_X	$p(X)$	Kov_X	p_X	$p(X)$	Kov_X	p_X	$p(X)$	Kov_X
$\varnothing$	1	0.0625	–	1	0	–	1	0	–
$\{x\}$	0.0625	0.0625	–	0.75	0	–	0.625	0	–
$\{y\}$	0.25	0.25	–	0.875	0	–	0.875	0	–
$\{z\}$	0.125	0.125	–	0.875	0	–	0.75	0	–
$\{v\}$	0.5	0.5	–	0.75	0	–	0.75	0	–
$\{x,y\}$	0	0	-0.016	0.625	0	-0.031	0.5	0	-0.047
$\{x,z\}$	0	0	-0.008	0.625	0	-0.031	0.375	0	-0.094
$\{x,v\}$	0	0	-0.031	0.5	0	-0.063	0.375	0	-0.094
$\{y,z\}$	0	0	-0.031	0.75	0	-0.016	0.625	0	-0.031
$\{y,v\}$	0	0	-0.125	0.625	0	-0.031	0.625	0	-0.031
$\{z,v\}$	0	0	-0.063	0.625	0	-0.031	0.5	0	-0.063
$\{x,y,z\}$	0	0	-0.002	0.5	0.25	-0.074	0.25	0.25	-0.16
$\{x,y,v\}$	0	0	-0.008	0.375	0.125	-0.117	0.25	0.25	-0.16
$\{x,z,v\}$	0	0	-0.004	0.375	0.125	-0.117	0.125	0.125	-0.226
$\{y,z,v\}$	0	0	-0.016	0.5	0.25	-0.074	0.375	0.375	-0.117
$\{x,y,z,v\}$	0	0	-0.001	0.25	0.25	-0.181	0	0	-0.308

Conclusion

In this paper, it is proposed a new approach to the definition of a discrete probabilistic distribution of the II-nd type of random set on a finite set of N events based on a given associative function. The advantage of this approach lies in the fact that the determination of the probabilistic distribution instead of a full set of 2^N is enough to know the probabilities of N probabilities of events and the type of the associative function. This approach is demonstrated on the example of the three associative functions. Considered in the work function are well known and widely used in fuzzy logic [10,13,14], and in probability theory [9,11,12]. Recurrent construction of a discrete probabilistic distribution on the basis of associative functions has led to the well-known probabilistic distributions of random sets of events with independently-point (12), embeddable (17) and disjoint (23) dependency structure which confirms the correctness of the proposed approach. Random sets with such structures dependencies play a key role in the theory of random sets of events, as described extreme situation [17]. It should be noted that the proposed recursive approach with the use of associative functions (11) and (14) always leads to one of the respective legitimate probability distributions of the random set of events. While the construction of a recurrence of a legitimate probability distribution of the random set of events with the use of associative function (18) is possible only under certain restrictions on the input probabilities of events. Thus, there are only three types of result sets of random events with the appropriate probabilistic distributions. Further investigating the characteristics of the constructed probabilistic distribution of random sets of events depending on the arguments of associative functions, are shown as promising (11), (14), (18), and the study of new classes of probability distributions constructed using well-known one-parameter families of associative functions, for example, Ali-Mikhail-Haq, Frank, and others [10].

References

[1] H.T.Nguyen, An Introduction to Random Sets, Taylor & Francis Group, LLC, 2006.

[2] O.Yu.Vorobyev, Eventology, Siberian Federal University, Krasnoyarsk, 2007 (in Russian).

[3] A.N.Shiryaev, Probability, vol. 1, Moscow Center for Continuous Mathematical Education, Moscow, 2004 (in Russian).

[4] O.Yu.Vorobyev, A.O.Vorobyev, Summation of the set-additive functions and Mobius inversion formula, *Reports of the Russian Academy of Sciences*, **336**(2009), no. 4, 417–420 (in Russian).

[5] D.V.Semenova, On New Notion of Quasi-Entropies of Eventological Distribution, Proc. of the Second IASTED International Multi-Conference on Automation Control And Information Technology, ACTA PRESS, Novosibirsk, Russia, 2005, 380–385.

[6] O.Yu.Vorobyev, N.A.Lukyanova, Properties of the entropy of multiplicative-truncated approximations of eventological distributions, *J. of Siberian Federal University. Math. & Physics*, **4**(2011), no. 1, 50–60.

[7] D.V.Semenova, N.A.Lukyanova, Random set decomposition of joint distribution of random variables of mixed type, *J. Proceedings of the Institute of Applied Mathematics*, Baku, Azerbaijan, **1**(2012), no. 2, 50–60.

[8] D.V.Semenova, L.Yu.Shangareeva, Associative Ali-Mikhail-Haq's random set, Proc. of scientific-applied conference «Statistics and its Applications», National University of Uzbekistan, Tashkent, Uzbekistan, 2013, 88–94.

[9] R.B.Nelsen, An Introduction to Copulas, Springer Science+Business Media, Inc., New York, 2006.

[10] S.Alsina, M.Frank, B.Schveizer, Associative functions: Triangular Norms and Copulas, World Scientific Publishing Co. Pte. Ltd., Singapore, 2006.

[11] K.Menger, Statistical Metrics. *Proc. Nat. Acad. Sci. USA.*, **8**(1942), 535–537.

[12] B.Schweizer, A.Sklar, Probabilistic metric spaces, New York, North Holland, 1983.

[13] E.P.Klement, R.Mesiar, E.Pap, Triangular norms, Kluwer Academic Pub., Boston, 2000.

[14] E.P.Klement, R.Mesiar, Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms, Elsevier, Amsterdam, 2005.

[15] A.I.Orlov, Non-numerical statistics, Moscow, MZ-Press, 2004 (in Russian).

[16] O.Yu.Vorobyev, A dependence of events, Proc. of the VII All-Russian FAM Conf. on Financial and Actuarial Mathametics and Related Fields, SFU, Krasnoyarsk, Russia, **1**(2008), 65–66 (in Russian).

[17] E.E.Goldenok, N.A.Lukyanova, D.V.Semenova, Applications of wide dependence theory in eventological scoring, Proc. of the IASTED Int. Conf. «Automation, Control and Information Technology» (ACIT-CDA 2010) in co-operation with the RAS, ACTA Press Anaheim–Calgary–Zurich, Novosibirsk, Russia, 2010, 316–322.

Исследование дискретных вероятностных распределений случайных множеств событий с помощью ассоциативных функций

Наталья А. Лукьянова
Дарья В. Семенова

В статье исследуется класс дискретных вероятностных распределений II-рода случайных множеств событий. В качестве инструмента построения таких распределений предлагается использовать ассоциативные функции. Излагается новый подход к определению дискретного вероятностного распределения II-рода случайного множества на конечном множестве из N событий на основе полученного рекуррентного соотношения и заданной ассоциативной функции. Преимущество предлагаемого подхода заключается в том, что для определения вероятностного распределения вместо полного набора 2^N вероятностей достаточно знать N вероятностей событий и вид ассоциативной функции. Рассмотрена $|X|$-арная ковариация случайного множества событий как мера аддитивного отклонения событий от независимой ситуации. На примере трех ассоциативных функций продемонстрирован процесс рекуррентного построения вероятностного распределения II-рода и приведено доказательство легитимности / нелегитимности полученного распределения с помощью перехода к вероятностному распределению I-рода по формулам Мёбиуса. Доказаны теоремы, устанавливающие вид и условия легитимности результирующих вероятностных распределений; найдены $|X|$-арные ковариации случайных множеств событий.

Ключевые слова: случайное множество событий, дискретное вероятностное распределение, ассоциативная функция, $|X|$-арная ковариация.