Proceedings of IAM, V.1, N.2, 2012, pp.188-202

RANDOM SET DECOMPOSITION OF JOINT DISTRIBUTION OF RANDOM VARIABLES OF MIXED TYPE*

Daria Semenova[1], Natalia Lukyanova[2]

[1]Siberian Federal University, Institute of Mathematics
[2]Siberian Federal University, Institute of Undergraduate Core Programs,
Krasnoyarsk, Russia
e-mail: nata00sfu@gmail.com, dariasdv@gmail.com

Abstract. Authors have studied the random set decomposition of random vectors whose components are constructed as a convex combination of any continuous and discrete random variables. For two-parametric random variables of mixed type are considered their characteristics and properties of the parameters. Authors have proved the theorem about decomposition of joint distribution on random set basis. This theorem is shown on a simple example for a doublet of events.

Keywords: random variables of mixed type, random set of events, eventological distributions, random set basis, quantitative superstructure.

AMS Subject Classification: 62H10, 62P05, 91B70, 91G10, 91G70, 91G99.

1. Introduction

Mathematical statistics almost always deals with either discrete or continuous random variables, but it is not so in real problems. Many functions of distribution used in various models (in particular, for modeling insurance payments or consumer choice) have both "continuously increasing" sites territory, and some positive jumps. In the present paper we consider random variables of mixed type:

$$\rho = I \cdot \xi + (1 - I) \cdot \nu, \tag{1}$$

where ξ is a continuous random variable (c.r.v.), ν is a discrete random variable (d.r.v.), and I is a Bernoulli random variable with parameter $p = \mathbf{P}(I = 1)$, $(1 - p) = \mathbf{P}(I = 0)$, such that I is stochastically independent on ξ and ν.

The distribution function $F_\rho(u)$ is a mix (a convex combination) of random variables ξ and ν

$$F_\rho(u) = p \cdot F_\xi(u) + (1 - p) \cdot F_\nu \tag{2}$$

and it is the function of the mixed, discrete-continuously type. Random variables of the form (1) are widely used in actuarial mathematics to model individual risks [1], in determining insurance rates and reserves, and also in reinsurance.

* Reported at the seminar of the Institute of Applied Mathematics in 11.09.2012

In this work authors consider a special case of variables (1) which have a positive jump at a given point *c*.

Let ξ be a c.r.v. with distribution function $F_\xi(u)$ and finite expectation, let d.r.v. ν have a degenerate distribution, i.e. $\mathbf{P}(\nu \equiv c) = 1$, $c \geq 0$, and let ν have a distribution function

$$F_\nu(u) = \mathbf{P}(\nu < u) = \begin{cases} 0, & u < c, \\ 1, & u \geq c. \end{cases}$$

Then (1) takes the form

$$\rho = I \cdot \xi + (1 - I) \cdot c \tag{3}$$

or it can be written equivalently

$$\rho = \begin{cases} \xi, & \textit{with probability } p, \\ c, & \textit{with probability } 1 - p. \end{cases}$$

Definition 1. A random variable of the form (3) is called a "two-parametric random variable of mixed type". Value of jump $(1 - p)$ and location of the jump *c* are parameters of jump and location accordingly. In considering the location *c* there are two situations:

1. Let *c* belong to a range of values of c.r.v. ξ, i.e. $c \in \mathrm{Ran}(\xi)$. In this case (2) can be written equivalently (fig.1)

$$F_\rho(u) = \begin{cases} p F_\xi(u), & u < c, \\ p F_\xi(u) + (1 - p), & u \geq c. \end{cases} \tag{4}$$

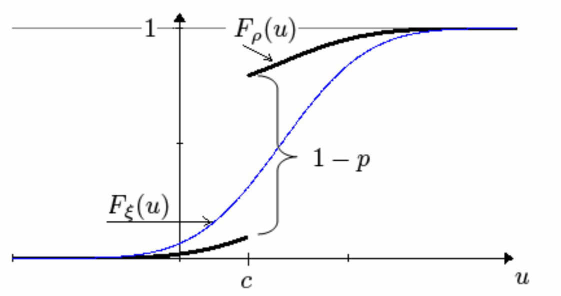

Figure 1. Distribution function of c.r.v. ξ and distribution function of random variable ρ for the case $c \in \mathrm{Ran}(\xi)$.

2. Let *c* do not belong to a range of values of c.r.v. ξ, i.e $c \notin \mathrm{Ran}(\xi) = [b, \infty)$. In this case (2) can be written equivalently (fig.2)

$$F_{\rho}(u)=\begin{cases}0, & u<c,\\ 1-p, & c\le u<b,\\ p\,F_{\xi}(u)+(1-p), & u\ge b.\end{cases} \tag{5}$$

The interval [c, b) is called "blind interval."

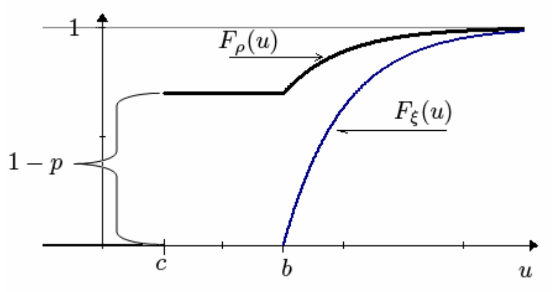

Figure 2. Distribution function of c.r.v. ξ and distribution function of random variable ρ for case $c\notin \text{Ran}(\xi)=[b,\infty)$.

In practice it is often assumed that ξ has an exponential distribution. Then the case $c\in \text{Ran}(\xi)$ describes, for example, a class of distributions which are used for modeling insurance payments [1]. Assuming $c\notin \text{Ran}(\xi)=[b,\infty)$ and $c=0$ we obtain a class of distributions of Gibbs random variables G_0 [2]. A key property of a Gibbs random variable is that it does not take values in the "blind interval" [0, b]. This property was pointed to by some interpretations of statistical theory of consumer choice [2], where a price of random purchase is a random variable of mixed type which has distribution significantly separated from zero (for most goods and services), while a zero value of the purchase price (corresponding to the absence of purchase) has a positive probability.

In the model of individual risks the insurance payments made by an insurance company, are represented as the sum of payments to many individuals [1]. The central limit theorem provides a method for finding the numerical values for the distribution of sums of independent random variables [3]. Here we offer mathematical tools allowing to work with joint distributions of random variables of the mixed type (3), and generalizes the results received earlier [2, 4, 5, 6].

2. Random set decomposition

Eventology [7] is a new direction of probability theory which is based on the Kolmogorov's axiomatics which is added by two eventological axioms: a sufficiency and a simpliciality [8]. The basic objects of researches are sets of random events and their eventological distributions. Eventology studies the structures of the dependences of the sets of events. It allows to include the mathematical model of a person, together with his/her persuasions, in a subject of the scientific research of in the form of eventological distribution (E- distribution) of the set of his/her own events, i.e. allows to consider any kinds of the set of events which are perceived and/or created by a person. New eventological language, new eventological methods and approaches allow formulating and solving various tasks in socio-economic areas which did not manage to formulate and solve earlier within the limits of traditional approaches.

Let's consider the *set of random events* $\mathbf{S} \subseteq A$ chosen by the algebra of the probabilistic space $(\Omega, A, \mathbf{P})$, or (that is equivalent), *random set of events* $K:(\Omega, A, \mathbf{P}) \to \left(2^{S}, 2^{2^{S}}\right)$ under a finite set of events $\mathbf{S}$, where $\mathbf{S} \subseteq A$ is a finite set of events (consisting of $N = |\mathbf{S}|$ events, i.e. $|\mathbf{S}|$ denote the cardinality of set ***S***), $2^{\mathbf{S}}$ is the power set of ***S***.

Each chosen event $x \in \mathbf{S}$ divides sample space into two disjoint events $\Omega = x + x^c$. These disjoint events are the event x and the event $x^c = \Omega - x$ which is its complement. In eventology the subsets or fragments of dividing **Ω** are called events-terraces generated by the finite set of events $\mathbf{S} \subseteq A$. All set of events ***S*** divides sample space into disjoint events-terraces of the following form:

$$\text{ter}\,(X) = \left(\bigcap_{x \in X} x\right) \cap \left(\bigcap_{x \in X^c} x^c\right) = \bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\,, \tag{6}$$

where $x^c = \Omega - x, \quad X^c = \mathbf{S} - X, \quad X \subseteq \mathbf{S}$.

Definition 2. Events (6) are called events-terraces of *I*-st sort, where

$$\Omega = \sum_{X \subseteq \mathbf{S}} \text{ter}(X) \text{ and } \text{ter}(X) \cap \text{ter}(Y) = \varnothing \Leftrightarrow X \neq Y.$$

Definition 3. Eventological distribution (E-distribution) of *I*-st sort of the set of random events ***S*** of the power of set $N = |\mathbf{S}|$ is a collection $p_{\text{I}} = \{p(X),\ X \subseteq \mathbf{S}\}$ of 2^N probabilities of event-terraces of *I*-st sort [7] generated by this set of events in which

$$p(X) = \mathbf{P}(K = X) = \mathbf{P}(\text{ter}(X)) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right). \tag{7}$$

The events-terraces (6) form the partition Ω in all $X \subseteq \mathbf{S}$: $\Omega = \sum_{X \subseteq \mathbf{S}} \text{ter}(X)$ and provide of probabilistic normalization $\sum_{X \subseteq \mathbf{S}} p(X) = 1$ for this sort of E-distribution of the set $\mathbf{S}$.

Definition 4. Events

$$\mathrm{ter}_X = \bigcap_{x \in X} x \tag{8}$$

are called events-terraces of *II*-nd sort.

Definition 5. E-distribution of *II*-nd sort of the set of random events $\boldsymbol{S}$ of the power $N = |\boldsymbol{S}|$ is a collection $p_{\mathrm{II}} = \{p_X,\ X \subseteq \boldsymbol{S}\}$ of 2^N probabilities of event-terraces of *II*-nd sort [7] generated by this set of events in which

$$p_X = \mathbf{P}(K \supseteq X) = \mathbf{P}(\mathrm{ter}_X) = \mathbf{P}\left(\bigcap_{x \in X} x\right),\ X \subseteq \boldsymbol{S}\,. \tag{9}$$

The event-terraces (8) form is not a partition, but only covering Ω, then the normalization relation for the E-distribution of *II*-nd sort is not satisfied. The sum of probabilities such event-terraces always there is more than unit:

$$\sum_{X \subseteq \boldsymbol{S}} p_X = \sum_{X \subseteq \boldsymbol{S}} 2^{|X|} p(X) > 1\,.$$

E-distribution of the *II*-nd sort is connected with E-distribution of the *I*-st sort by Mőbius inversion formulas

$$p_X = \sum_{X \subseteq Y} p(Y), \quad p(X) = \sum_{X \subseteq Y} (-1)^{|Y|-|X|} p_Y\,.$$

On the example of an arbitrary doublet of events $\boldsymbol{S} = \{x, y\}$ with E-distribution of *I*-st sort $\{p(\varnothing),\ p(\{x\}),\ p(\{y\}),\ p(\{x, y\})\}$ and E-distribution of *II*-nd sort $\{p_\varnothing,\ p_{\{x\}},\ p_{\{y\}},\ p_{\{x,y\}}\}$ we write out the Mőbius inversion formulas:

$$p_\varnothing = p(\varnothing) + p(\{x\}) + p(\{y\}) + p(\{x, y\}) = 1;$$
$$p_{\{x\}} = p(\{x\}) + p(\{x, y\}) = \mathbf{P}(x);\ \ p_{\{y\}} = p(\{y\}) + p(\{x, y\}) = \mathbf{P}(y);$$
$$p_{\{x,y\}} = p(\{x, y\}) = \mathbf{P}(x \cap y).$$

On the other hand,

$$p(\varnothing) = p_\varnothing - p_{\{x\}} - p_{\{y\}} + p_{\{x,y\}};$$
$$p(x) = p_{\{x\}} - p_{\{x,y\}};\ \ p(y) = p_{\{y\}} - p_{\{x,y\}};\ \ p(xy) = p_{\{x,y\}}.$$

A random set of events is a random element with values in a power set $\boldsymbol{S}$, where $\boldsymbol{S}$ is a finite set of selected events. The main idea of the contemporary theory of random sets† asserts that the structure of statistical interdependence of subsets of a finite set is completely determined by the distribution of the random set which is defined on the power set. Distribution of a random set is a convenient mathematical tool for description of all conceivable ways to combine elements in coalitions, in other words, all the ways of interaction among elements.

† Though the theory of random sets has well-traced connections with the multivariate statistical analysis, the subject of its researches is a random finite abstract set which essentially differs from a random vector that belongs to the abstract spaces which do not have habitual linear structures.

Enumerate the N random variables of the form (3) by the elements of the set $\boldsymbol{S}$ in order by first difference (lexicographical order). Let's consider N-dimensional joint distribution of random variables

$$\{\rho_x, x \in \boldsymbol{S}\} = \{I_x \cdot \xi_x + (1 - I_x) \cdot c_x, x \in \boldsymbol{S}\},$$

where for all $x \in \boldsymbol{S}$

- ξ_x is c.r.v.,
- $c_x \geq 0$ is an invariable,
- and we associate Bernoulli random variables I_x with indicators

$$I_x = 1_K(x) = \begin{cases} 1, & x \in K, \\ 0, & x \notin K. \end{cases}$$

The components of the random vector $\{\rho_x, x \in \boldsymbol{S}\}$ are random variables of the mixed type. Hence we can say that the random vector $\{\rho_x, x \in \boldsymbol{S}\}$ is constructed from a random vector $\{\xi_x, x \in \boldsymbol{S}\}$ composed of N continuous random variables ξ_x, $x \in \boldsymbol{S}$ by adding jumps at the points $\{c_x, x \in \boldsymbol{S}\}$.

Let's consider the events

$$\mathbf{I}_X = \left\{ \left(\bigcap_{x \in X} \{I_x = 1\} \right) \bigcap \left(\bigcap_{x \in X^c} \{I_x = 0\} \right) \right\} = \left\{ \bigcap_{x \in X} \{I_x = 1\} \bigcap_{x \in X^c} \{I_x = 0\} \right\}$$

for all $X \subseteq \boldsymbol{S}$, where $X^c = \boldsymbol{S} \setminus X$ is the complement of a subset of events X to $\boldsymbol{S}$, and $x^c = \Omega \setminus x$ is the complement of an event x. Note that an event $\mathbf{I}_X$ means that all the Bernoulli random variables indexed by elements x of the set X take a value of 1, i.e. $I_x = 1$ for all $x \in X$, and all the Bernoulli random variables from set X^c take zero value. Thus, the event $\mathbf{I}_X$ is a partition of the Bernoulli random vector into two parts:

$$\{I_x, x \in \boldsymbol{S}\} = \{1, x \in X\} + \{0, x \in X^c\}.$$

The number of such partitions coincides with the power of set $\boldsymbol{S}$. In [9] is proved the following statement:

Statement. The set of events $\{\mathbf{I}_X, X \subseteq \boldsymbol{S}\}$ forms an exhaustive event.

3. Theorem on decomposition of joint distribution on random set basis

Theorem 1. For a joint distribution of two-parametric random variables of mixed type $\{\rho_x, x \in \boldsymbol{S}\}$ the random set decomposition

$$F(u_x, x \in \boldsymbol{S}) = \mathbf{P}\left(\bigcap_{x \in \boldsymbol{S}} \{\rho_x < u_x\} \right) = \sum_{X \subseteq \boldsymbol{S}} F_X(u_x, x \in \boldsymbol{S}) \cdot p(X), \tag{10}$$

takes place, where

- the random set basis $\{p(X), \quad X \subseteq \boldsymbol{S}\}$

$$p(X) = \mathbf{P}\ (\mathbf{I}_X) = \mathbf{P}\ \left\{ \bigcap_{x\in X} \{I_x = 1\} \bigcap_{x \in X^c} \{I_x = 0\} \right\}; \tag{11}$$

- the quantitative superstructure is the collection of the conditional distribution functions $\{F_X(u_x, x \in \boldsymbol{S}),\ X \subseteq \boldsymbol{S}\}$:

$$F_X(u_x, x \in \boldsymbol{S}) = F_{\xi_X}(u_x, x \in X) \cdot \prod_{x \in X^c} F_{c_x}(u_x), \tag{12}$$

where

- $F_{\xi_X}(u_x, x \in X) = \mathbf{P}\ \left(\bigcap_{x \in X} \{\xi_x < u_x\} \middle| \mathbf{I}_X \right)$ is the conditional distributions of a continuous random vector $\{\xi_x, x \in \boldsymbol{S}\}$ conditioned on the event $\mathbf{I}_X$ occurrence;
- $F_{c_x}(u_x)$ is the distribution function of the degenerate random variables c_x, $x \in \boldsymbol{S}$, herewith $\prod_{x \in X^c} F_{c_x}(u_x) = \begin{cases} 1, & u_x \geq c_x, \quad \forall x \in X^c \\ 0, & \textit{otherwise.} \end{cases}$

Proof. Consider the joint distribution of two- parametric random variables of mixed type of the form (3) $\{\rho_x, x \in \boldsymbol{S}\}$

$$F(u_x, x \in \boldsymbol{S}) = \mathbf{P}\left(\bigcap_{x \in \boldsymbol{S}} \{\rho_x < u_x\} \right).$$

The values taken by random variables $\{\rho_x, x \in \boldsymbol{S}\}$ depend on the values which Bernoulli random variables $I_x, x \in \boldsymbol{S}$ take. Let's notice that the number of partitions of all the components of the vector $\{I_x, x \in \boldsymbol{S}\}$ equal to the power set **S**, then $\sum_{X \in 2^S} \mathbf{I}_X = \Omega$. Since, set of events $\{\mathbf{I}_X, X \subseteq \boldsymbol{S}\}$ form an exhaustive events (6), hence from (7) $\mathbf{P}\ (\mathbf{I}_X) = \mathbf{P}\ (K = X) = p(X),\ X \subseteq \boldsymbol{S}$.

Therefore, using the formula of total probability we have the following representation

$$F(u_x, x \in \boldsymbol{S}) = \mathbf{P}\left(\bigcap_{x \in \boldsymbol{S}} \{\rho_x < u_x\} \right) =$$

$$= \sum_{X \subseteq \boldsymbol{S}} \mathbf{P}\ \left(\bigcap_{x \in X} \{\rho_x \leq u_x\} \middle| \mathbf{I}_X \right) \cdot \mathbf{P}\ (\mathbf{I}_X) = \sum_{X \subseteq \boldsymbol{S}} F_X(u_x, x \in \boldsymbol{S}) \cdot p\ (X).$$

Consider the conditional distribution functions

$$F_X(u_x, x \in \boldsymbol{S}) = \mathbf{P}\ \left(\bigcap_{x \in X} \{\rho_x < u_x\} \middle| \mathbf{I}_X \right).$$

Since, random variables ρ_x of a set $x \in X$ accept continuous values ξ_x at approach of an event $\mathbf{I}_X$, and random variables from a set X^c accept values c_x, then

$$\mathbf{P}\left(\left(\bigcap_{x\in X}\{\xi_x < u_x\}\right)\cap\left(\bigcap_{x\in X^c}\{c_x < u_x\}\right)\right).$$

Note that random variables ξ_x and c_x are independent for all $x \in \boldsymbol{S}$, then

$$\mathbf{P}\left(\bigcap_{x\in X}\{\xi_x < u_x\}\right)\cdot\mathbf{P}\left(\bigcap_{x\in X^c}\{c_x < u_x\}\right) = F_{\xi_X}(u_x, x\in X)\cdot\prod_{x\in X^c}F_{c_x}(u_x),$$

.where

- $X^c = \boldsymbol{S} \setminus X$,
- $F_{\xi_X}(u_x, x \in X)$ is the joint distribution function of a continuous random variables $\{\xi_x, x \in \boldsymbol{S}\}$,
- $F_{c_x}(u_x)$ is the distribution function of degenerate random variables $c_x, x \in \boldsymbol{S}$.

Hence, we obtain the decomposition

$$F(u_x, x \in \boldsymbol{S}) = \sum_{X \subseteq \boldsymbol{S}}\left(F_{\xi_X}(u_x, x \in X)\cdot\prod_{x\in X^c}F_{c_x}(u_x)\right)\cdot p(X). \qquad (13)$$

Thus, the theorem is proved.

Note that the sum (10) contains 2^N summands each of which is representable as a product of an element of the quantitative superstructure (12) on the corresponding element of the basis (11).

Thus, it is possible to speak about two-level structure of dependences and interactions of the components of the random vector $\{\rho_x, x \in \boldsymbol{S}\}$. The first is the random set level which is responsible for full structure of statistical dependences and interactions of random events. It forms the random set basis $\mathrm{p}_{\mathrm{I}} = \{p(X), \quad X \subseteq \boldsymbol{S}\}$. The second is the quantitative level which is responsible for structure of dependences and interactions of components of the joint distribution of the two-parametric random variables of mixed type in a quantitative superstructure $\{F_X(u_x, x \in \boldsymbol{S}), \ X \subseteq \boldsymbol{S}\}$ as the set of the conditional distribution functions from the joint distribution of the continuous random vector $\{\xi_x, x \in \boldsymbol{S}\}$.

Let's consider independence of a superstructure. Let $N = |\boldsymbol{S}|$ marginal functions of distribution of c.r.v. F_{ξ_x}, $x \in \boldsymbol{S}$ are known and let c.r.v. ξ_x are total independence. Then

$$F_{\xi_X}(u_x, x \in X) = \prod_{x\in X}F_{\xi_x}(u_x).$$

Hence, sum (13) takes the form

$$F(u_x,\ x\in \boldsymbol{S})=\sum_{X\subseteq \boldsymbol{S}}\left(\prod_{x\in X} F_{\xi_x}(u_x)\right)\cdot\left(\prod_{x\in X^c} F_{c_x}(u_x)\right)\cdot p(X)\,. \qquad (14)$$

Further, we assume independence of the basis. E-distribution for total independence set of events $\boldsymbol{S}$ has to form [10]:

$$p(X)=\prod_{x\in X} p_x \prod_{x\in X^c}(1-p_x),\ \ X\subseteq \boldsymbol{S}\,.$$

Then sum (10) has following form

$$F(u_x,\ x\in \boldsymbol{S})=\sum_{X\subseteq \boldsymbol{S}}\left(\prod_{x\in X} F_{\xi_x}(u_x)\right)\cdot\left(\prod_{x\in X^c} F_{c_x}(u_x)\right)\cdot \prod_{x\in X} p_x \prod_{x\in X^c}(1-p_x)\,. \ (15)$$

4. Example for the doublet of events

Let $\boldsymbol{S}=\{x, y\}$ be an arbitrary doublet of events with E-distribution of *I*-st sort $\{p(\varnothing), p(\{x\}), p(\{y\}), p(\{x, y\})\}$.

Enumerate random variables (3) of elements of set $\boldsymbol{S}=\{x, y\}$

$$\{\rho_x, \rho_y\}=\{I_x\cdot\xi_x+(1-I_x)\cdot c_x,\ I_y\cdot\xi_y+(1-I_y)\cdot c_y\}.$$

Find the joint distribution of random variables

$$F(u_x, u_y)=\mathbf{P}(\{\rho_x<u_x\}\cap\{\rho_y<u_y\})\,.$$

The values of the random variables depend on the values I_x and I_y. In this example 4 variants are possible:

1. $\mathbf{I}_\varnothing=\{I_x=0,\, I_y=0\}$ for event $X=\varnothing$ with probability $p(\varnothing)$. In that case $\rho_x=c_x$, $\rho_y=c_y$.

2. $\mathbf{I}_{\{x\}}=\{I_x=1,\, I_y=0\}$ for event $X=\{x\}$ with probability $p(\{x\})$. In that case $\rho_x=\xi_x$, $\rho_y=c_y$.

3. $\mathbf{I}_{\{y\}}=\{I_x=0,\, I_y=1\}$ for event $X=\{y\}$ with probability $p(\{y\})$. In that case $\rho_x=c_x$, $\rho_y=\xi_y$.

4. $\mathbf{I}_{\{x,y\}}=\{I_x=1,\, I_y=1\}$ for event $X=\{x, y\}$ with probability $p(\{x\ y\})$. In that case $\rho_x=\xi_x$, $\rho_y=\xi_y$.

Note that

$$\mathbf{P}(\{I_x=1\})=\mathbf{P}(\mathbf{I}_{\{x\}})+\mathbf{P}(\mathbf{I}_{\{x,y\}})=p(\{x\})+p(\{x\ y\}).$$

The events $\{\mathbf{I}_\varnothing, \mathbf{I}_{\{x\}}, \mathbf{I}_{\{y\}}, \mathbf{I}_{\{x,y\}}\}$ form an exhaustive event, then on the formula of total probability we obtain

$$\mathbf{P}(\{\rho_x<u_x\}\cap\{\rho_y<u_y\})=$$

$$= \mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_\varnothing \right) \cdot p(\varnothing) +$$
$$+ \mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_{\{x\}} \right) \cdot p(\{x\}) +$$
$$+ \mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_{\{y\}} \right) \cdot p(\{y\}) +$$
$$+ \mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_{\{x,y\}} \right) \cdot p(\{x, y\}).$$

Consider each summand individually.

1. Since $\rho_x = c_x, \rho_y = c_y$; c_x in $\mathbf{I}_\varnothing$ and c_y has degenerate distributions, then

$$\mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_\varnothing \right) = \mathbf{P}\left(\{c_x < u_x\} \cap \{c_y < u_y\} \right) =$$
$$= \mathbf{P}\left(\{c_x < u_x\} \right) \cdot \mathbf{P}\left(\{c_y < u_y\} \right) = F_\varnothing(u_x, u_y) =$$
$$= \begin{cases} 1, & u_x \ge c_x, u_y \ge c_y, \\ 0, & \textit{otherwise}. \end{cases}$$

2. Since in $\mathbf{I}_{\{x\}}$ the random variables $\rho_x = \xi_x, \rho_y = c_y$, ξ_x and c_y are independent, because c_y has degenerate distribution, then

$$\mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_{\{x\}} \right) = \mathbf{P}\left(\{\xi_x < u_x\} \cap \{c_y < u_y\} \right) =$$
$$= \mathbf{P}\left(\{\xi_x < u_x\} \right) \cdot \mathbf{P}\left(\{c_y < u_y\} \right) = F_{\xi_x}(u_x) \cdot F_{c_y}(u_y) = F_{\{x\}}(u_x, u_y) =$$
$$= \begin{cases} F_{\xi_x}(u_x), & u_y \ge c_y, \\ 0, & u_y < c_y, \end{cases}$$

where $F_{\xi_x}(u_x)$ is the distribution function of c.r.v. ξ_x.

3. Similarly, we obtain

$$\mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_{\{y\}} \right) = \mathbf{P}\left(\{c_x < u_x\} \cap \{\xi_y < u_y\} \right) =$$
$$= F_{c_x}(u_x) \cdot F_{\xi_y}(u_y) = F_{\{y\}}(u_x, u_y) = \begin{cases} F_{\xi_y}(u_y), & u_x \ge c_x, \\ 0, & u_x < c_x, \end{cases}$$

where $F_{\xi_y}(u_y)$ is the distribution function of c.r.v. ξ_y.

4. Since $\rho_x = \xi_x, \rho_y = \xi_y$, in $\mathbf{I}_{\{x,y\}}$ then

$$\mathbf{P}\left(\{\rho_x < u_x\} \cap \{\rho_y < u_y\} \,\middle|\, \mathbf{I}_{\{x,y\}} \right) = \mathbf{P}\left(\{\xi_x < u_x\} \cap \{\xi_y < u_y\} \right) =$$
$$= F_{\{x,y\}}(u_x, u_y) = \begin{cases} F_{\xi_x \xi_y}(u_x, u_y), & u_x \ge c_x, u_y \ge c_y, \\ 0, & u_x < c_x, u_y < c_y, \end{cases}$$

where $F_{\xi_x \xi_y}(u_x, u_y)$ is the joint distribution function of c.r.v. ξ_x and ξ_y.

In summary, we obtain the decomposition (fig.3.)

$$F(u_x, u_y) = F_{\varnothing}(u_x, u_y)\cdot p(\varnothing) + F_{\{x\}}(u_x, u_y)\cdot p(\{x\}) +$$
$$+ F_{\{y\}}(u_x, u_y)\cdot p(\{y\}) + F_{\{x,y\}}(u_x, u_y)\cdot p(\{x, y\})$$

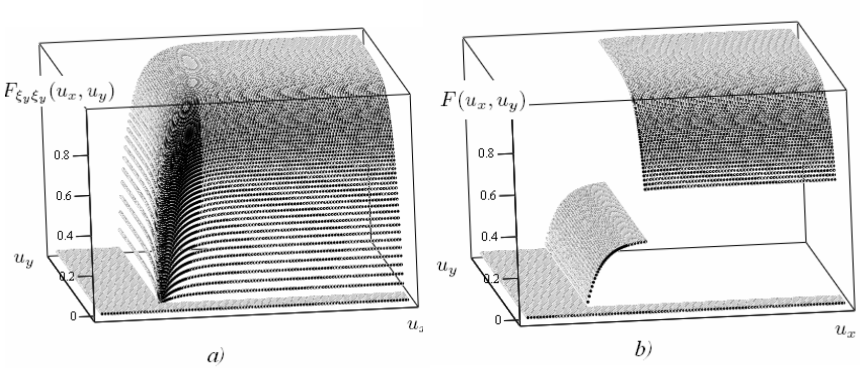

Figure 3. Example of the graphical representation of the joint distribution function of two-parametric discrete-continuous random variables ρ_x and ρ_y (*b*). It is constructed from the joint distribution of c.r.v. ξ_x and ξ_y (*a*).

We give the following interpretation for this example. Under an event in statistical system of a consumer choice we understand the purchases (sale) of these or those goods. These goods take part in trade turnover in the considered commodity market. And they form a finite set of names of the goods. We will use the notation $\boldsymbol{S}$ for the set of events as purchases of the goods, and for a set of names of the corresponding goods which are involved in the turnover in the market. It is thought that $x \in \boldsymbol{S}$ is an event that consists in purchase of the goods with the name of x.

Consider two-dimensional random vector

$$\{\rho_x, \rho_y\} = \{I_x \cdot \xi_x + (1 - I_x)\cdot c_x, I_y \cdot \xi_y + (1 - I_y)\cdot c_y\},$$

that describes the joint purchase of two goods x and y by the random buyer. Assume that there is a statistical sampling from the n observations of the values of the random vector. It can be, for example, statistics of sales of a supermarket on products $x = \{\text{bread}\}$ and $y = \{\text{milk}\}$. This usual statistics allows estimating the two-level structure of dependences and interactions of an observable random vector which offered in the work.

Thus, in this example, based on statistical data we need to estimate the joint distribution function $F(u_x, u_y)$ of two-dimensional random vector $\{\rho_x, \rho_y\}$ of revenues of goods.

Algorithm of estimation:

1. Statistical evaluation of random set basis is made on the first level. This is the estimate of the distribution of the random set of events. Since we are considering a two-dimensional random vector, its random set basis is the

corresponding random set of events K which is determined the probability distribution (11) $\{p(X), \quad X \subseteq \boldsymbol{S}\}$ under a doublet events $\boldsymbol{S} = \{x, y\}$. It contains 4 probabilities, which we denote accordingly:

- $p(\varnothing) = \mathbf{P}(K = \varnothing)$ is probability of not purchasing goods (or in other words, it is probability the "empty" purchases of goods under the doublet $\boldsymbol{S} = \{x, y\}$, i.e. the probability that the buyer retired from supermarket without the goods x and y).
- $p(\{x\}) = \mathbf{P}(K = \{x\})$ is probability of purchase only goods x.
- $p(\{y\}) = \mathbf{P}(K = \{y\})$ is probability of purchase only goods y.
- $p(\{x, y\}) = \mathbf{P}(K = \{x, y\})$ is probability of purchase two goods.

In practice, the statistical evaluation of the distribution of random set demand of the buyer is reduced to estimating the distribution of a random set of events K, i.e. to the purchasing goods. This estimate is made on the basis of an available sample of checks of purchases on a formula:

$$p(X) = \frac{n_X}{n},$$

where n_X is the number of checks for a subset of goods X, n is the total number of checks. Any market, first of all, is defined by interaction of the buyer and the seller, that is supply and demand.

2. At the second level, we make statistical estimates of the quantitative superstructure. Conditional distribution function (12) $\{F_X(u_x, x \in \mathbf{S}), \ X \subseteq \mathbf{S}\}$ for a sample using standard methods of mathematical statistics are estimated here. In this example of the two-dimensional random vector, this collection consists of four conditional distributions:

- $F_{\varnothing}(u_x, u_y)$ is degenerate distribution under condition of "empty" purchasing.
- $F_{\{x\}}(u_x, u_y)$ is one-dimensional distribution under condition of purchase only goods x.
- $F_{\{y\}}(u_x, u_y)$ is one-dimensional distribution under condition of purchase only goods y.
- $F_{\{x, y\}}(u_x, u_y)$ is two-dimensional distribution under condition of purchase two goods x and y.

3. In the last step according to Theorem 1 (theorem on decomposition of joint distribution on random set basis) we construct a overall statistical estimate of the distribution of the observed two-dimensional random vector of the value of purchases of the two goods:

$$F(u_x, u_y) = F_{\varnothing}(u_x, u_y) \cdot p(\varnothing) + F_{\{x\}}(u_x, u_y) \cdot p(\{x\}) +$$

$$+F_{\{y\}}(u_x, u_y)\cdot p(\{y\}) + F_{\{x,y\}}(u_x, u_y)\cdot p(\{x,y\})$$

We can introduce the simplifying assumption of independence conditional two-dimensional distribution $F_{\{x,y\}}(u_x, u_y)$ which will lead to the fact that to estimate the distribution of the original random vector will be sufficient estimates no more than one-dimensional conditional distributions:

$$F_{\{x,y\}}(u_x, u_y) = F_{\{x\}}(u_x, u_y)\cdot F_{\{y\}}(u_x, u_y).$$

Similar assumptions can be made by statistical estimates of random vectors of higher dimension when is assumed the independence of the conditional distributions of dimension greater than two, three, etc. depending on the application.

5. Conclusion

The joint distribution of the random vector $\{\rho_x, x \in \boldsymbol{S}\}$ is input data for a series of practical problems [1, 2, 4]. According to Theorem 1 to estimate the joint distribution of the random vector $\{\rho_x, x \in \boldsymbol{S}\}$, we need to be able to solve the following two problems:

Problem 1. Approximation of the E-distribution of the *I*-st sort which plays the role of the random set basis (2^N parameters) in our model.

Problem 2.Modeling the joint distribution of a continuous random vector

$$\{\xi_x, x \in \boldsymbol{S}\}.$$

Thus, the problem of modeling the joint distribution of discrete-continuous random vector moved from the domain of a multivariate distribution function to area of assessment of its parameters. As a rule, using real statistics we can estimate only 2*N* parameters: *N* probabilities of *II*-nd sort $\{p_X, \quad X \subseteq \boldsymbol{S}\}$ and *N* marginal distribution functions $\{F_{\xi_x}, \ x \in \boldsymbol{S}\}$.

For solving the first problem, we used methods that were considered in [2, 4, 7, 10]. If the random variables are independent, their joint distribution is determined through the product of the marginal (15). Otherwise, the second problem may be solved using the concept of copula to describe dependence between random variables that relates the marginal distributions to their joint distribution function [5, 6, 7, 11, 12, 13].

The statistical system can be defined as the random set of events which form an original structure of statistical interrelations of random events. Studying structures of statistical interrelations of random events means learning probability distributions corresponding to random sets of the events. Therefore it is necessary to study some fundamental structures of interdependence of systems of random events which generate many known structures of interdependence of random variables, random vectors, random processes and fields and demand special research by random set methods.

Acknowledgment

We would like to express special gratitude to our teacher prof. Oleg Vorobyev which stimulated in a current of many years of our researches in eventology and also we would like to express our sincere gratitude to prof. Arkady Novosyolov for his constructive criticism and the help.

References

1. Kaas R., Goovaerts M., Dhaene J., Denuit M. Modern Actuarity Risk Theory, Kluwer Academic Publishers, Boston/Dordrecht/London, 2001, 376 p.
2. Vorob'ov O.Yu., Goldenok E.E. Structural set-analysis of random events, Krasnoyarsk: KrasSU, 2004, 106 p. (in Russian).
3. Shiryaev A.N. Probability, M: MCCME, 2004, 576 p. (in Russian).
4. Vorob'ov O.Yu., Semenova D.V. Portfolio set-analysis of random events, Krasnoyarsk: KrasSU, 2005, 106 p. (in Russian).
5. Vorob'ov O.Yu., Goldenok E.E., Semenova D.V. Random-set decompositions of bipartite random vectors, Journal of Krasnoyarsk State University: Mathematics & Physics, vol. 1, 2003, pp.88–96 (in Russian).
6. Semenova D.V. About distributions of bipartite random vectors, Proc. of the IASTED Int. Conf. on ACIT - Control, Diagnostics, and Automation (ACIT-CDA 2010), v.692, Novosibirsk, Russia, 2010, pp.306-312.
7. Vorobyev O.Yu. Eventology, Krasnoyarsk, Russia, Siberian Federal University (SFU), 2007, 435 p. (in Russian).
8. Vorobyev O.Yu. On elements of axiomatizing eventology, J. Sib. Fed. Univ. Math. Phys., v.3, No.2, Krasnoyarsk, Russia, 2010, pp.157–164.
9. Semenova D.V., Lukyanova N.A. Distributions of the two-parameter discrete-continuous random variables, Proc. of the XI Int. FAMES'2012 Conf., Krasnoyarsk, Russia, 2012, pp.347–353 (in Russian).
10. Vorobyev O.Yu., Lukyanova N.A. Properties of the entropy of multiplicative-truncated approximations of eventological distributions, J. Sib. Fed. Univ. Math. Phys., vol.4, N.1, Krasnoyarsk, Russia, 2011, pp.50-60.
11. Korolev V.Yu., Benning V.E., Shorgin S.Ya. Mathematical bases of the theory of risk, Moscow, FIZMATLIT, 2007 (in Russian).
12. Nelsen R.B. An Introduction to Copulas, New York, Springer, 1999.
13. Embrechts P., Lindskog F., McNeil A.J. Modelling dependence with copulas and applications to risk management, in Rachev ST (Ed.) Handbook of heavy tailed distributions in finance. Amsterdam: Elsevier/North-Holland, 2003.
14. Nguyen Hung T. An Introduction to Random Sets. USA, Taylor & Francis Group, LLC, 2006, 264 p.

Qarışıq tip təsadüfi kəmiyyətlərin birgə paylanmasının təsadüfi-çoxluqlar üzrə ayrılışı

D. Semyonova, N. Lukyanova

XÜLASƏ

İşdə təsadüfi vektorların təsadüfi-çoxluqlar üzrə ayrılışı tədqiq olunur. Bu vektorların komponentləri ixtiyari kəsilməz və diskret təsadüfi kəmiyyətlərin qabarıq kombinasiyasıdır. Qarışıq tipli ikiparametrli təsadüfi kəmiyyətlərin xarakteristikaları və parametrlərin xassələri öyrənilir. Təsadüfi-çoxluqlar bazisi üzrə birgə paylanmanın ayrılışı ilə bağlı teorem isbat olunur. Bu teorem hadisələr cütü misalında nümayiş etdirilir.

Açar sözlər: qarışıq tipli təsadüfi kəmiyyət, hadisələrin tasadüfi çoxluğu, eventologi paylanma, təsadüfi-çoxluq bazisi, kəmiyyət üst strukturu.

Случайно-множественное разложение совместного распределения случайных величин смешанного типа

Д. Семёнова, Н. Лукьянова

РЕЗЮМЕ

В работе исследуется случайно-множественное разложение случайных векторов, компоненты которых конструируются как выпуклая комбинация произвольных непрерывных и дискретных случайных величин. Рассматриваются характеристики и свойства параметров для двухпараметрических случайных величин смешанного типа. Авторами доказывается теорема о разложении совместного распределения по случайно-множественному базису. Данная теорема демонстрируется на простом примере для дуплета событий.

Ключевые слова: случайная величина смешанного типа, случайное множество событий, эвентологическое распределение, случайно-множественный базис, количественная надстройка.