# Random Set Decomposition of Discrete-Continuous Random Variables

Daria Semenova[1], Natalia Lukyanova[2]
[1]Institute of Mathematics Siberian Federal University, Krasnoyarsk, Russia
[2]Institute of Undergraduate Core Programs Siberian Federal University, Krasnoyarsk, Russia
[1]*dariasdv@gmail.com*, [2]*nata00sfu@gmail.com*

***Abstract*— In work the definition of a two-parametric discrete-continuous random variable is offered. Authors have studied decomposition of joint distribution of two-parametric discrete-continuous random variable by the random set basis.**

***Keywords*— discrete-continuous random variables; random set of events**

## I. INTRUDUCTION

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 discrete-continuous random variables:

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

where $\xi$ is a continuous random variable (c.r.v.), $\nu$ 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 of $\xi$ and $\nu$. The distribution function $F_\rho(u)$ has the form

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

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, 2], in determining insurance rates and reserves, and also in reinsurance.

In this work authors consider a special case of variables (1) which have a positive jump at a given point $c$. Let $\xi$ be a c.r.v. with distribution function $F_\xi(u)$ and finite expectation, let d.r.v. $\nu$ have a degenerate distribution, i.e.

$$\mathbf{P}(\nu \equiv c) = 1, \quad c \ge 0,$$

and let $\nu$ have a distribution function

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

Then (1) takes the form

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

or it can be written equivalently

$$\rho = \begin{cases} \xi, & \text{with probability } p, \\ \text{c}, & \text{with probability } 1 - p. \end{cases} \tag{4}$$

A random variable of the form (3) is called a "two-parametric discrete-continuous random variable." Value of jump $1 - p$ and location of the jump $c$ are parameters of jump and location accordingly.

Let's write out characteristics of the random variable $\rho$.

- Expectation $\mathbf{E}\rho = p \cdot \mathbf{E}\xi + (1 - p) \cdot c.$
- Variance $\mathbf{D}\rho = p \cdot \mathbf{D}\xi + p \cdot (1 - p) \cdot (\mathbf{E}\xi - c)^2.$
- Moment generating function

  $$m(t) = p \cdot m_\xi(t) + (1 - p)\,\mathrm{e}^{tc},$$

  conditioned on existence of the moment generating function $m_\xi(t)$ of c.r.v. $\xi$.
- Characteristic function

  $$\varphi(t) = p \cdot \varphi_\xi(t) + (1 - p)\,\mathrm{e}^{itc},$$

  where $\varphi_\xi(t)$ is the characteristic function of c.r.v. $\xi$.

Up to this moment we didn't discuss the location of $c$. It is possible to allocate two situations.

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

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

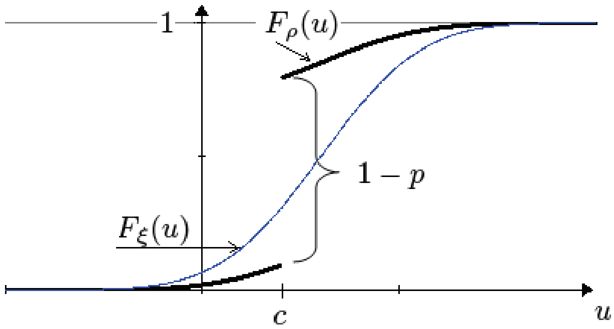



Figure 1. Distribution function of c.r.v. $\xi$ and distribution function of random variable $\rho$ for the case $c \in \mathrm{Ran}(\xi)$.

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

$$F_\rho(u)=\begin{cases} 0, & u<c, \\ 1-p, & c\le u<b, \\ pF_\xi(u)+(1-p), & u\ge b. \end{cases} \qquad (6)$$

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

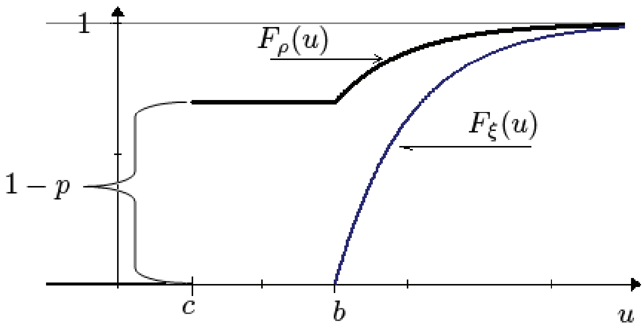



Figure 2. Distribution function of c.r.v. $\xi$ and distribution function of random variable $\rho$ for case $c \notin \mathrm{Ran}(\xi)=[b,\infty)$.

In practice it is often assumed that $\xi$ has an exponential distribution. Then the case $c \in \mathrm{Ran}(\xi)$ describes, for example, a class of distributions which are used for modeling insurance payments [1, 2]. Assuming $c \notin \mathrm{Ran}(\xi)=[b,\infty)$ and $c=0$ we obtain a class of distributions of Gibbs random variables $G_0$ [3]. 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 [3], where a price of random purchase is a discrete-continuous random variable 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 [4, 5]. 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 [3, 6, 7, 8, 9].

Consider a random set of events

$$K:(\Omega, \mathcal{A}, \mathbf{P}) \to \left(2^{\boldsymbol{S}}, 2^{2^{S}}\right)$$

under $\boldsymbol{S}$, where $\boldsymbol{S} \subseteq \mathcal{A}$ is a finite set of events (consisting of $N=|\boldsymbol{S}|$ of events), $2^{\boldsymbol{S}}$ is the power set of $\boldsymbol{S}$. A probability distribution of a random set $K$ is a collection of $2^N$ probabilities $p(X)=\mathbf{P}(K=X), \quad X \in 2^{\boldsymbol{S}}$ :

$$\mathrm{p_I}=\left\{p(X), \quad X\in 2^{\boldsymbol{S}}\right\}$$

– eventological distribution of I-st sort ; or

$$\mathrm{p_{II}}=\left\{p_X, \quad X\in 2^{\boldsymbol{S}}\right\}$$

– eventological distribution of II-nd sort [10, 11].

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

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

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[1] 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.

Enumerate the $N$ random variables of the form (3) by the elements of the set $S$ 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 $x\in\boldsymbol{S}$ $\xi_x$ is c.r.v., $c_x\ge 0$ is an invariable, and we associate Bernoulli random variables $I_x$ with indicators

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

[1] 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.

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 $\xi_x$, $x \in \boldsymbol{S}$ by adding jumps at the points $\{c_x, x \in \boldsymbol{S}\}$.

Note an interesting, from the viewpoint of practical applications, property [6] of the sum of random variables of the mixed type (3) for all $c_x = 0$, $x \in \boldsymbol{S}$:

$$\sum_{x \in K} \xi_x = \sum_{x \in \boldsymbol{S}} \rho_x .$$

- the sum of a random set of continuous random variables is equal to the sum of random variables of the mixed type (3).

Consider the joint distribution of two- parametric discrete-continuous random variables of the form (4) $\{\rho_x, x \in \boldsymbol{S}\}$.

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

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 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 $\boldsymbol{S}$.

***Statement.*** *The set of events* $\{\mathbf{I}_x, X \subseteq \boldsymbol{S}\}$ *form a complete set of events.*

***Theorem 1.*** *For a joint distribution of two-parametric discrete-continuous random variables*

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

*the random set decomposition*

$$F(u_x, x \in \boldsymbol{S}) = \sum_{X \subseteq \boldsymbol{S}} F_X(u_x, x \in \boldsymbol{S}) \cdot p(X), \qquad (7)$$

*takes place, where*

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

$$p(X) = \mathbf{P}(\mathbf{I}_X) = \mathbf{P}\left\{ \bigcap_{x \in X} \{I_x = 1\} \bigcap_{x \in X^C} \{I_x = 0\} \right\}; \qquad (8)$$

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

$$F_X(u_x, x \in \boldsymbol{S}) = F_{\xi_X}(u_x, x \in X) \cdot \prod_{x \in X^c} F_{c_x}(u_x), \qquad (9)$$

*where* $F_{\xi_X}(u_x, x \in X) = \mathbf{P}\left( \bigcap_{x \in X} \{\xi_x \le u_x\} | \mathbf{I}_X \right)$ *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 constant random variables* $c_x, x \in \boldsymbol{S}$, *herewith*

$$\prod_{x \in X^c} F_{c_x}(u_x) = \begin{cases} 1, & u_x \ge c_x, \quad \forall x \in X^c \\ 0, & \textit{otherwise.} \end{cases}.$$

Note that the sum (7) contains $2^N$ summands each of which is representable as a product of an element of the quantitative superstructure (9) on the corresponding element of the basis (8).

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 $p_{\mathrm{I}} = \{p(X), \quad X \subseteq \boldsymbol{S}\}$. 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 discrete-continuous random variables 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 $\{\xi_x, x \in \boldsymbol{S}\}$ (fig.3).

The joint distribution of the random vector $\{\rho_x, x \in \boldsymbol{S}\}$ is input data for a series of practical problems [1, 2, 3, 6]. 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 eventological distribution of the I-st sort which plays the role of the random set basis ($2^N$ parameters) in our model.

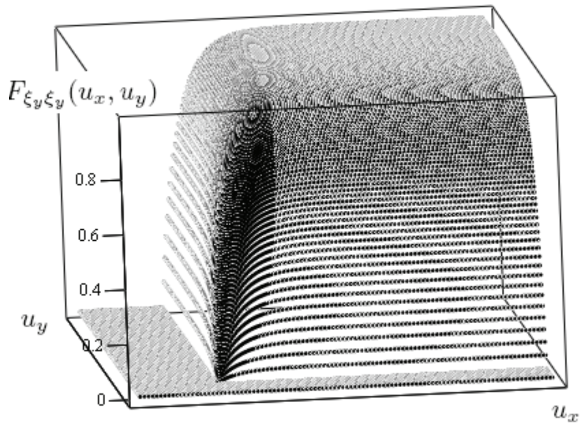
Problem 2. Modeling the joint distribution of a continuous random vector $\{\xi_x, x \in \boldsymbol{S}\}$.

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.

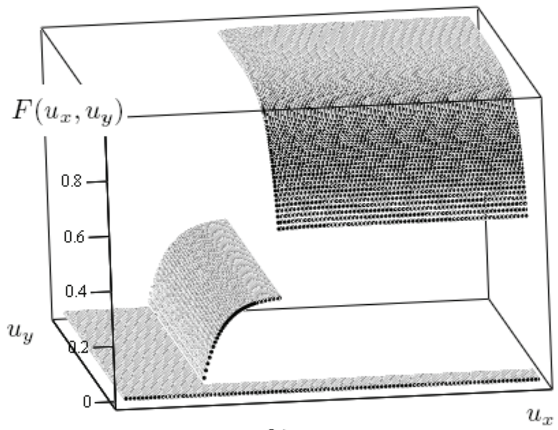
As a rule, using real statistics we can estimate only $2N$ parameters:

1. $N$ probabilities of II-nd sort $\{p_X, \quad X \subseteq \boldsymbol{S}\}$;
2. $N$ marginal distribution functions $\{F_{\xi_x}, \quad x \in \boldsymbol{S}\}$.

For solving the first problem, we used methods that were considered in [3, 6, 10, 12, 13].




*a)*




*b)*

Figure 3. Example of the graphical representation of the joint distribution function of two-parametric discrete-continuous random variables $\rho_X$ and $\rho_Y$ (*b*). It is constructed from the joint distribution of c.r.v. $\xi_X$ and $\xi_Y$ (*a*). Detailed consideration of this example in the work [14, 15].

If the random variables are independent, their joint distribution is determined through the product of the marginals. 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 [7, 8, 9, 10, 16, 17, 18].

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.

## References

[1] R. Kaas, M. Goovaerts, J. Dhaene and M. Denuit, "Modern Actuarity Risk Theory," Kluwer Academic Publishers, Boston/Dordrecht/London, 2001.

[2] "Models of individual risks on a short interval of time," StatSoft Russia, www.statsoft.ru/statportal/tabID__103/MId__464/ModeID__0/PageID__374/DesktopDefault.aspx, (in Russian).

[3] O. Yu. Vorob'ov and E.E. Goldenok, "Structural set-analysis of random events," Krasnoyarsk: KrasSU, 2004. (in Russian).

[4] A.N. Shiryaev, "Probability," M: MCCME, 2004. (in Russian).

[5] Yu.A. Rozanov, "Theory of probability, random processes and mathematical statistics," Moscow, Science, 1985. (in Russian).

[6] O. Yu. Vorob'ov and D.V. Semenova, "Portfolio set-analysis of random events, " Krasnoyarsk: KrasSU, 2005. (in Russian).

[7] O.Yu. Vorob'ov, E.E. Goldenok and D.V. Semenova, "Random-set decompositions of bipartite random vectors," J. of Krasnoyarsk state university: mathematics & physics, vol. 1, 2003, pp. 88–96. (in Russian).

[8] D.V. Semenova, "About distributions of bipartite random vectors," Proc. of the IASTED Int. Conf. on ACIT - Control, Diagnostics, and Automation (ACIT-CDA 2010), Vol. 692, Novosibirsk, Russia, 2010, pp.306–312.

[9] D.V. Semenova, "Distributions of bipartite random vectors and its appendix," Proc. of the III Int. Conf. "Problems of Cybernetics and Informatics" (PCI'2010), vol. III, Baku, Azerbaijan, Azerbaijan National Academy of Sciences, 2010, pp.205–208.

[10] O.Yu.Vorobyev, "Eventology," Krasnoyarsk, Russia, Siberian Federal University (SFU), 2007. (in Russian).

[11] O.Yu.Vorobyev, "On elements of axiomatizing eventology," J. of Siberian Federal University. Mathematics & Physics, Vol. 3, No. 2, Krasnoyarsk, Russia, 2010, pp.157–164.

[12] O.Yu. Vorobyev and N.A. Lukyanova, "Properties of the entropy of multiplicative-truncated approximations of eventological distributions," J. of Siberian Federal University. Mathematics & Physics, vol. 4, No 1, Krasnoyarsk, Russia, 2011, pp.50–60.

[13] N.A.Lukyanova, "Eventology sight at entropy," Proc. of the III Int. Conf. "Problems of Cybernetics and Informatics" (PCI'2010), vol. III, Baku, Azerbaijan, Azerbaijan National Academy of Sciences, 2010, pp.209–212.

[14] D.V. Semenova and N.A. Lukyanova, "Distributions of the two-parameter discrete-continuous random variables," Proc. of the XI Int. FAMES'2012 Conf. , Krasnoyarsk, Russia, 2012, pp.347–353.

[15] D.V. Semenova and N.A. Lukyanova, "Research of joint distributions of discrete-continuous random variables," Int. Conf. "Probability Theory and its Applications" in Commemoration of the Centennial of B.V. Gnedenko, Moscow, 2012, pp.298–299. (in Russian).

[16] V.Yu. Korolev, V.E. Benning, S.Ya. Shorgin, "Mathematical bases of the theory of risk," Moscow, FIZMATLIT, 2007. (in Russian).

[17] R.B. Nelsen, "An Introduction to Copulas, " New York, Springer, 1999.

[18] P. Embrechts, F. Lindskog and A.J. McNeil, "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.