DISTRIBUTIONS OF BIPARTITE RANDOM VECTORS AND ITS APPENDIX

Daria Semenova

Siberian Federal University, Institute of Mathematics
Krasnoyarsk, Russia, *dariasdv@aport2000.ru*

The statistical system can be defined as random set of events which form a original structure of statistical interrelations of random events with each other. The studying of structures of statistical interrelations of random events means the learning of probability distributions of corresponding random sets of 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. Practical interest to studying of these problems speaks aspiration to raise efficiency of functioning of some statistical systems of the nature and a society. Hence, there is a necessity for application mathematical well-founded models which allow to estimate consequences of this or that decision in the course of the control and management of statistical systems.

Any random finite set K can be considered as a special case of the stochastic element, which value are subsets of finite set x with power of set N=|x|. Therefore, it is possible to make following definition of random finite abstract set [1], [2]. Random finite abstract set (RFAS) K under x is a measurable mapping of certain probabilistic space (Ω, **A**, **P**) in the finite measurable space $(2^{\mathrm{X}}, 2^{2^{\mathrm{X}}})$. Key concept is random set of events under x which is defined as the random finite abstract set under x, where x$\subseteq$**A** is the selected finite set of events (consisting from N=|x|.); the set 2^{X} is the power set of. x As distribution of probabilities of the random set K (or, that is equivalent, a eventological distribution of selected random events set x$\subseteq$**A**) is called the set from 2^N probabilities: $p(X)=\mathbf{P}(K=X)$, $X\subseteq$ x .

Let's consider any random variable η with function of distribution F_η, continuous in zero, and a finite expectation $E\eta = \int_{-\infty}^{\infty} u dF_\eta(u) < \infty$. Let's add to a random variable "the atom" in zero with probability *1-p*, having kept it former distribution with probability *p*. Also we will designate the received random variable $\xi(p)$. Thus, it is possible to tell that the random variable $\xi(p)$ accepts two generalized values, one of which is usual zero with probability *1-p* and the second is the random variable with probability *p*:

$$\xi(p) = \begin{cases} 0, & with \quad probability \ 1-p, \\ 1, & with \quad probability \ \ p. \end{cases}$$

As a matter of fact it is a question of a mix of distributions. The received random variable will have the distribution function, keeping property of a continuity at the left in each point of the real axis:

$$F_\xi(u) = \begin{cases} pF_\eta(u), & u < 0, \\ 1 - p + pF_\eta(u), & otherwise. \end{cases}$$

*The bipartite random variable (or the generalized Bernoulli's random variable) with param*eters (p,η) we will name a random variable $\xi(p)$ having a continuous in zero distribution function with probability p and the atom in zero with probability of *1-p*. Random variable η with function of distribution continuous in zero we will name nonzero value of a bipartite random variable $\xi(p)$.

Let's consider finite set of events x and the random variables "numbered" by these events, which form a bipartite random vector $\overrightarrow{\xi(p)} = [\xi_x(p_x), x \in \mathbf{X}]$ with joint function of distribution

$F_p(u_x, x \in \mathbf{X}) = \mathbf{P}\left\{\bigcap_{x\in\mathbf{X}}\{\xi_x(p_x) < u_x\}\right\}$, where $\mathbf{p} = \{p_x,\ x \in \mathbf{X}\}$, and individual functions of distribution are $F_{\xi_x}(u_x) = \mathbf{P}\{\xi_x(p_x) < u_x\}$, $x \in \mathbf{X}$, accordingly.

Let's consider distribution of probabilities of random set K under x: $p(X)=\mathbf{P}(K=X)$, $X \subseteq$ x which is defined by joint function of distribution under formulas $p(X) = \mathbf{P}\left[\bigcap_{x\in X}\{\xi_x(p_x) \neq 0\} \bigcap_{x\in X^c}\{\xi_x(p_x) = 0\}\right]$, where $X\subseteq$ x. Distribution parameters p_x have obvious interpretation $p_x = \mathbf{P}\{\xi_x(p_x) \neq 0\}$, $x \in \mathbf{X}$, as probability of that $\xi_x(p_x) \neq 0$, and the probability $p(X)$ are the probability of that there was an event $\boldsymbol{e}_X = \left\{\bigcap_{x\in X}\{\xi_x(p_x) \neq 0\} \bigcap_{x\in X^c}\{\xi_x(p_x) = 0\}\right\}$, which consists that nonzero components of a bipartite random vector $\overrightarrow{\xi(p)}$ form a subset $\{\xi_x(p_x), x \in X\}$. Early [2] the following theorem has been proved.

Theorem about decompositions of a bipartite random vectors distribution on random-set basis. *Let $\overrightarrow{\xi(p)} = [\xi_x(p_x), x \in \mathbf{X}]$ is the bipartite random vector made from N=|x| of bipartite random variables with the joint N-dimensional function of distribution $F_p(u_x, x \in \mathbf{X}) = \mathbf{P}\left\{\bigcap_{x\in\mathbf{X}}\{\xi_x(p_x) < u_x\}\right\}$. Then for function $F_p(u_x, x \in \mathbf{X})$ decomposition is fair*

$$F_p(u_x, x \in \mathbf{X}) = \sum_{X\in 2^{\mathbf{X}}} F_X(u_X, x \in \mathbf{X} \mid \boldsymbol{e}_X) p(X), \qquad (1)$$

*where $F_X(u_X, x \in \mathbf{X} \mid \boldsymbol{e}_X) = \mathbf{P}\left\{\bigcap_{x\in\mathbf{X}}\{\xi_x(p_x) < u_x \mid \boldsymbol{e}_X\}\right\}$ are conditional functions of distribution of the bipartite random vector $\overrightarrow{\xi(p)}$ under condition of the random event $\boldsymbol{e}_X = \left\{\bigcap_{x\in X}\{\xi_x(p_x) \neq 0\} \bigcap_{x\in X^c}\{\xi_x(p_x) = 0\}\right\}$, and p(X)=**P**(**e**$_X$) X⊆x, - probability of such event.*

Zero-dimensional function of distribution $F_\varnothing(\bullet) = \mathbf{P}\{\bullet \mid_\varnothing\} = \prod_{x\in\mathbf{X}} \mathbf{1}_{\{u_x \geq 0\}}$ is a distribution function of a degenerate random vector which accepts unique zero value: {0,0, …, 0}. It is equal to unit every time, when all u_x are non-negative.

Set of the conditional distribution functions $\{F_X(u_X, x \in \mathbf{X}), X \subseteq \mathbf{X}\}$ form the quantitative superstructure.

Thus, it is possible to speak about two-level structure of dependencies. The first is the random-set level which is responsible for full structure of statistical dependencies and interactions of random events. It forms random-set basis. The second is the quantitative level which is responsible for structure of dependencies and interactions a component of a bipartite random vector in a quantitative superstructure.

Generally the quantity of disconnected ranges of definition of conditional density of distribution at decomposition of bipartite random vector distribution on random-set basis is equal $3^{|\mathrm{x}|}$, where |x| coincides with dimension of a random vector [2]. The theorem about a bipartite random vector distribution decomposition on random-set basis allows to describe two-level structure of dependencies and interactions a component of a bipartite random vector. On the basis of the given theorem for calculation of joint function of bipartite random vector distribution it is necessary to set $2^{|\mathrm{x}|+1}$ parameters. The task so a great number of parameters strongly complicates the decision of a problem of construction of distribution of a bipartite random vector. To take advantage of the given theorem, it is often convenient to make the assumption of conditional independence of a quantitative superstructure. Then joint distribution is completely defined by distribution of is random-set basis and individual conditional distributions. And various distributions of is random-set basis at the same quantitative superstructure lead to various joint distributions of a bipartite random vector and on the contrary. The bipartite random vector ξ is designed from a random vector η. Thus a quantitative superstructure are the conditional functions of distribution received from joint distribution of a random vector η. On practice the joint distribution of the random vector it is convenient to receive by means of a copula.

A copula is a function which joins or "couples" a multivariate distribution function to its one-dimensional marginal distribution functions. The word "copula" was first used in a mathematical or statistical sense by Sklar (1959) in the theorem which bears his name [4].

Copulas can be defined informally as follows [4]: let η_x and η_y be continuous random variables with distribution functions $F(u_x) = \mathbf{P}(\eta_x \le u_x)$ and $G(u_y) = \mathbf{P}(\eta_y \le u_y)$, and joint function of distribution $H(u_x, u_y) = \mathbf{P}(\eta_x \le u_x, \eta_y \le u_x)$ For every (u_x, u_y) in $[-\infty,\infty]^2$ consider the point in I^3 (I=[0,1]) with coordinates ($F(u_x)$, $G(u_y)$, $H(u_x, u_y)$). This mapping from I^2 to I is copula. Formally we have.

A two-dimensional copula is a function $C: I^2 \to I$ such that $C(0, u_x) = C(u_x, 0) = 0$ and $C(1, u_x) = C(u_x, 1) = u_x$ for all $x \in I$; C is 2-increasing: for $a,b,c,d \in I$ with $a \le b$ and $c \le d$: $V_C([a,b] \times [c,d]) = C(b,d) - C(a,d) - C(b,c) + C(a,c) \ge 0$

The informal and formal definitions are connected by the following theorem (Sklar, 1959), which also partially explains the importance of copulas in statistical modeling [4]. *N*-dimensional copula is a multidimensional joint function of distribution with uniformly distributed on [0,1] marginals [5].

Copula contains all information on the dependence nature between random variables which is not present in individual distributions, but does not contain the information about individual distributions. As a result the information on marginal and the information on dependence between them are separated by copula from each other. Let's formulate the theorem about decompositions of a bipartite random vectors distribution on random-set basis in language of copula.

Theorem about decompositions of a bipartite random vector distribution on random-set basis in language of copula. *Let* $\overrightarrow{\xi(p)} = [\xi_x(p_x), x \in \mathbf{X}]$ *is the bipartite random vector made from* $N=|\mathbf{X}|$ *of bipartite random variables* $\xi_x(p_x)$ *with parameters* (p_x, η_x) *accordingly and with joint N-dimensional function of distribution* $F_p(u_x, x \in \mathbf{X}) = \mathbf{P}\left\{ \bigcap_{x \in \mathbf{X}} \{\xi_x(p_x) < u_x\} \right\}$. *Then for function* $F_p(u_x, x \in \mathbf{X})$ *decomposition is fair* $F_p(u_x, x \in \mathbf{X}) = \sum_{X \in 2^{\mathbf{X}}} C_X(G_x \mid \boldsymbol{e}_X), x \in \mathbf{X}) p(X)$, *where* $C_X(G_x \mid \boldsymbol{e}_X)$ *are conditional copulas of bipartite random vectors* $\overrightarrow{\xi(p)}$ *under condition of random event* $\boldsymbol{e}_X = \left\{ \bigcap_{x \in X} \{\xi_x(p_x) \ne 0\} \bigcap_{x \in X^c} \{\xi_x(p_x) = 0\} \right\}$, *and* $p(X)=\mathbf{P}(\boldsymbol{e}_X)$ $X \subseteq \mathbf{X}$, *- probability of such event;* $G_x = G(u_x) = \mathbf{P}(\eta_x \le u_x)$, $x \in \mathbf{X}$, *are individual, continuous in zero distribution functions with finite expectation.*

The following theorem has the important practical value in financial appendices.

Theorem. *The sum of a random set of random variables* η_x, $x \in \mathbf{X}$ *is equal to the sum of the bipartite random variables* $\xi_x(p_x)$, $x \in \mathbf{X}$ *with parameters* (p_x, η_x): $\sum_{x \in K} \eta_x = \sum_{x \in \mathbf{X}} \xi_x(p_x)$, *where K is the random set of nonzero components;* $p_x = \mathbf{P}(x \in K)$ *is the probability of a covering of an element x by random set K.*

The offered instrument of bipartite random vectors can be used in financial appendices, namely in the portfolio analysis. Within the limits of the classical portfolio theory of Markowitz the problem of management by distribution of the capital is considered for the fixed portfolio of operations, i.e. management of a portfolio is reduced to effective distribution of the capital between in advance chosen fixed count of operations. However usually aside there is the most labour-consuming problem of the portfolio analysis: a management of a choice of structure of a portfolio, i.e. the management of set of the market transactions forming a portfolio. The classical problem of Markowitz gives only quantitative decision which considers a demand of the buyer for separate operations. While it is necessary to consider as well demand of the buyer for sets of operations (fig. 1). The theorem about the decompositions of bipartite random vectors distribution on random-set basis allows to analyze the two-level structure of dependencies. The quantitative superstructure is responsible for dependencies between profitabilities of operations. In works a insufficiency of multi-normal distribution which is usually applied to modeling of joint distribution of profitabilities of operations is shown. Functions are the flexible tool which is used at construction of effective algorithms for the best modeling of these distributions.

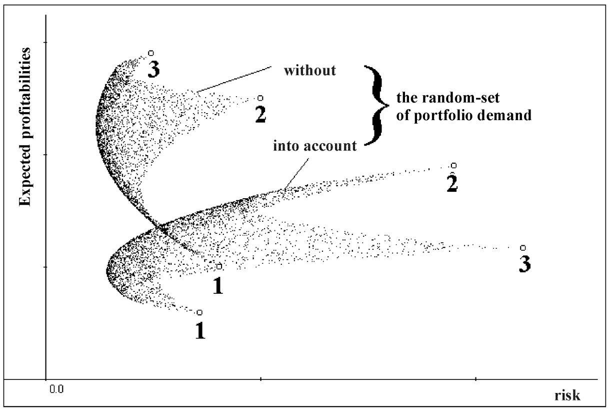

Fig 1: The numerical decision of a problem of management of a random portfolio of operations: at the left above - the classical bullet of Markowitz constructed without the random-set portfolio demand; on the right below – a bullet constructed taking into account random-set of portfolio demand of the buyer on the basis of the theorem of decompositions of a bipartite random vector distribution on random-set basis.

The results received in work are recommended to use at the analysis of the commodity markets, at the decision of problems of distribution of resources in various market systems [2], in problems of financing of the market of scientific researches, in bank and insurance business [6].

References

1. O.Yu. Vorobyev, E.E. Goldenok and D.V. Semenova. Random-set decompositions of bipartite random vectors, J. of Krasnoyarsk state university: mathematics & physics, 1, 2003, 88 - 96. (in Russian).
2. O. Yu. Vorobyev and D.V. Semenova. Portfolio set-analysis of random events. Krasnoyarsk: KrasSU, 2005. (in Russian).
3. K. Romano. Calibrating and simulating copula functions: an application to the Italian stock market, Centro Interdipartimentale sul Diritto e l'Economia dei Mercati, Working paper n. 12/2002.
4. R.B. Nelsen. An Introduction to Copulas. New York: Springer, 1999.
5. P. Embrechts, F. Lindskog and A.J. McNeil. Modelling dependence with copulas nd applications to risk management, in Rachev ST (Ed.) Handbook of heavy tailed distributions in finance. Amsterdam: Elsevier/North-Holland, 2003.
6. R. Kaas, M. Goovaerts, J. Dhaene and M. Denuit. Modern Actuarial Risk Theory. Boston/Dordrecht/London: Kluwer Academic Publishers, 2001.