REGRESSION BETWEEN THE POLYTYPIC DATA, DESCRIBING BEHAVIOR OF COMPLEX SYSTEM

Irina Baranova

Siberian Federal University, Krasnoyarsk, Russia
irinabar@yandex.ru

The purpose of the given work is dependence research between the polytypic data describing behavior of complex system, by means of a finding the regression between them. As it is known, regression is one of the main concepts of probability theory and the mathematical statistics and expresses dependence of average value of a random variable on value of other random variable or on several random variables.

In some fields of science and practical activities different researches result to solving the problems of system analysis. In paper it is considered a situation, when the one part of the events describing the complex system's behavior is numerical and the second part is sets. Eventology suggests analyzing the random events, whose describe of system's behavior. The basic difficulty in analysis of such complex systems lies in the fact that number of all possible events is big and the data describing system's behavior is polytypic. This problem especially actual for applied fields of science, whose are bound up with analysis of social, economic and natural systems. There are medicine, ecology, biology, actuary, finances, insurance, sociology and others.

In work [1, 2] was suggested the bipartite set of random events method, in which each system's element represents as bipartite set of random events. The first part of this set corresponds to the random variables, and second part – to the sets. The basic idea of this method concludes in reduction an analysis of system's elements to analysis of corresponding bipartite sets of events. It was offered to use the probability of Minkovsky symmetry difference set-operation as distance between bipartite sets of random events.

For dependence research between the polytypic data underlying behavior of system, the classical concept of regression does not approach, as it is a question of link revealing not only between random variables, but between sets which can consist of any quantity of random variables and random sets. Therefore in this paper is offered the new concept of eventological regression between bipartite sets of events. In work [3] O. Yu. Vorob'ev suggested the formula of the eventological regression of one set of events on other set of events. We generalize the concept of eventological regressions for more difficult case of regression of two bipartite sets of random events and find the appropriate formula for expression of this dependence. Also in work are considered the notions of the conditional probability of the bipartite set of events and the terrace for bipartite set of events.

Bipartite set of random elements

In paper we survey complex system, whose behavior is describing by numeric and set data. Then the results of observation for the behavior of researching object is a set, which consists of random variables and random sets. As it is known, each random variable and each random finite set, follow Kolmogorov, can be considered as random element's realization [2].

Let (Ω, F, P) is probabilistic space, where *F* is set of events, $x, y \in F$ — random events, $\boldsymbol{P}$ — probability measure. The measurable reflection

$$K : (\Omega, F, P) \to \left(2^{X}, 2^{2^{X}}\right),$$

which possible values are subsets of the finite set of events F, is called a random set of events *K*, defined under the given finite set of events $X \in F$ by probability distribution

$$p(X) = \mathrm{P}(K = X),\ X \in 2^{X}.$$

As stated above, our general situation, when the one part of the results of observation for behavior of researching object are numerical and the second part ─ sets, can be described as the set of the random elements.

The set of the random elements $\{\xi, \mathrm{K}\}$ we will call the *bipartite set of random elements*. It can be defined in that way:

$$\{\xi, \mathrm{K}\} = \xi \cup K = \{\xi_a, a \in A, K_\beta, \beta \in B\}.$$

This set of random elements separates on the two parts. The first part is the random variables ξ, the second part is the random sets *K*. Let *A* is the indices set of a random variables, *B* ─ indices set of a random sets.

Bipartite set of random events

Now we considered the bipartite set of random events, which corresponds the bipartite set of random elements, defines above.

Let $\{\xi_a, a \in A\}$ are random variables with finite set of possible values

$$\Re_a = \{r_{a_1}, \dots, r_{a_{Na}}\} \subset \mathrm{R},\ \ \mathrm{a} \in \mathrm{A}.$$

To each random variables can be put in correspondence set of events

$$\xi_a \Rightarrow Y_a = \{Y_a(r_a), r_a \in \Re_a\}.$$

The event $Y_a(r_a) = \{\omega : \xi_a(\omega) \le r_a\}$ is the event from definition of distribution function of random variable and a set with inserting structure of dependences.

To first part ─ random variables can be put in correspondence set of events:

$$\xi \Rightarrow Y = \bigcup_{a \in A} Y_a.$$

To each random set of events K_β can be put in correspondence the finite set of events X_β:

$$K_\beta \Leftrightarrow X_\beta.$$

To second part ─ random sets of events *K* can be put in correspondence common set of events X:

$$K \Leftrightarrow \mathrm{X} = \sum_{\beta \in B} X_\beta,\ \beta \in B.$$

Bipartite set of random events can be defined in that way:

$$\{Y, \mathrm{X}\} = \{Y_a, X_\beta, a \in A, \beta \in B\}. \tag{2}$$

It is suggested for the bipartite set of random elements $\{\xi, \mathrm{K}\}$ put in correspondence the bipartite set of random events $\{Y, \mathrm{X}\}$.

Conditional bipartite set of events and conditional probabilities of the bipartite set of events

Let (Ω, F, P) is probabilistic space, where $\mathrm{X} \in F$ is set of events, $x \in \mathrm{X}$ ─ random event, $s \subseteq \Omega$ ─ event with nonzero and not unit probability: $0 < \mathrm{P}(s) < 1$.

Let's spend the redenotation of bipartite set of random events presented by formula (2) in that way:

$$\mathrm{Z} = \{Y, \mathrm{X}\} = \{Y_a, X_\beta, a \in A, \beta \in B\}. \tag{3}$$

Let *s* is bipartite set of random events, which is a subset of bipartite set of random events Z (i.e. $s \subseteq \mathrm{Z}$):

$$s = \{Y_{s_A}, \mathrm{X}_{s_B}, s_A \subseteq A, s_B \subseteq B\}. \tag{4}$$

Terrace for bipartite set of events s represents as a set of not intersected events where each event is a subset of appropriate set of events Y_a or X_β:

$$\begin{aligned} ter(s) &= \{Y_{s_A}, X_{s_B}\} = \\ &= \bigcap_{a \in s_A} ter(Y_a) \bigcap_{\beta \in s_B} ter(X_\beta) = \\ &= \bigcap_{a \in s_A} Y_a(r_a) \bigcap_{\beta \in s_B} \left(\bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c \right), \end{aligned} \tag{5}$$

where $s_A \subseteq A,\ s_B \subseteq B,\ r_a \in \Re_a,\ X \in X_\beta$. As it is told above set $\Re_a$ is the finite set of possible values of random variable ξ_a: $\Re_a = \{r_{a_1}, \ldots, r_{a_{Na}}\} \subset \mathrm{R}$.

Conditional bipartite set of events $s \mid x$ is the event meaning that approach of events from bipartite set s occurs according to an occurrence of an event x (event $x \in Y_a$ or $x \in X_\beta$).

Conditional probability of approach the bipartite set of events (probability of approach the conditional bipartite set of events $s \mid x$) can be defined by following formula:

$$\begin{aligned} \mathrm{P}(ter(s) \mid x) &= p_{|x}(ter(s)) = \\ &= \frac{\mathrm{P}(ter(s) \cap x)}{\mathrm{P}(x)}, \end{aligned} \tag{6}$$

where $ter(s)$ finds under the above-stated formula (5).

Let s, t are the bipartite sets of random events and the subsets of bipartite set of random events Z (i.e. $s, t \subseteq \mathrm{Z}$):

$$s = \{Y_{s_A}, \mathrm{X}_{s_B}, s_A \subseteq A,\ s_B \subseteq B\},\ t = \{Y_{t_A}, \mathrm{X}_{t_B}, t_A \subseteq A,\ t_B \subseteq B\}.$$

Conditional bipartite set of events $t \mid s$ is the event meaning that approach of events from bipartite set t occurs according to an approach of events from bipartite set s.

Conditional probability of approach the bipartite set of events (probability of approach the conditional bipartite set of events $t \mid s$) can be defined by that way:

$$\begin{aligned} \mathrm{P}(ter(t) \mid ter(s)) &= p_{|ter(s)}(ter(t)) = \\ &= \frac{\mathrm{P}(ter(t) \cap ter(s))}{p(ter(s))}, \end{aligned} \tag{7}$$

where $p(ter(s)) = \sum_{ter(s) \subseteq \hat{\mathrm{Z}} \subseteq \mathrm{Z}} \mathrm{P}(\mathrm{Z})$, a $ter(t)$ and $ter(s)$ finds by formula (5).

Regression of bipartite sets of random events

In work [3] O. Yu. Vorob'ev suggested the formula of the eventological regression of one set of events on other set of events.

Eventological regression of one set of events X on another set of events Y, linking terrace-events $ter(Y)$ with terrace-events $ter(X)$ looks like the following:

$$ter(Y) = \varphi(ter(X)). \tag{8}$$

Let's generalize the given concept for more difficult case of regression of two bipartite sets of random events.

Eventological regression of one bipartite set of random events $s \subseteq \mathrm{Z}$ on another set of events $t \subseteq Z$ linking terrace-events $ter(s)$, $s \subseteq Z$ with terrace-events $ter(t)$, $t \subseteq Z$ determines by that way:

$$ter(t) = \varphi(ter(s)),\ s \subseteq Z. \tag{9}$$

Let's set the problem of a statistical estimation of the function of the theoretical eventological regression at level α, which value on it is defined as conditional α-kvantil of the bipartite set t.

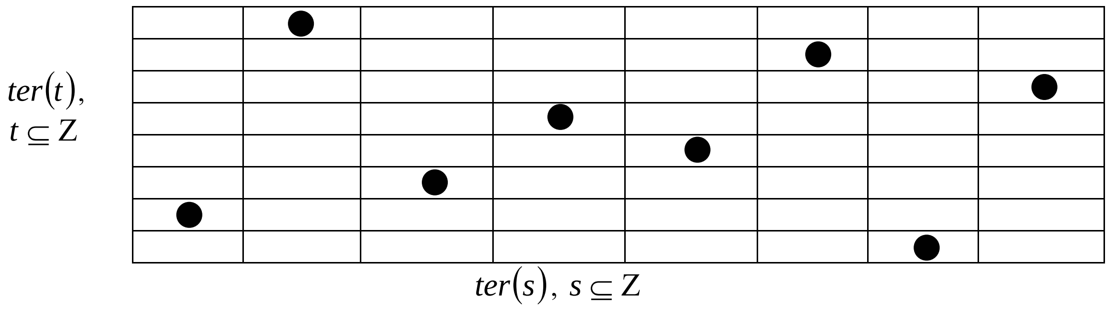
By the conditional α-kvantil of the bipartite set of random events t the occurrence of events from bipartite set s we mean the set of events from the bipartite set t (i.e. terrace-event $ter(t)$), which conditional probabilities is more than given value α:

$$Q_\alpha(t \mid ter(s)) = \{ter(t) \subseteq Z : \mathrm{P}(ter(t) \mid ter(s)) \geq \alpha\}. \tag{10}$$

Thus, function of the theoretical eventological regression of one bipartite sets of events $s \subseteq Z$ to another bipartite set of events $t \subseteq Z$ represents as a reflection of events of set s on set t. To define what elements of $ter(t)$ belong to value of function of the eventological regression it is necessary to define the events entering in conditional α-kvantil of the bipartite set of random events t under the above-stated formula (10). Conditional probabilities $\mathrm{P}(ter(t) \mid ter(s))$ from the definition of the conditional α-kvantil of the bipartite set of random events t will be calculated by the formula (7).

For construction of estimation of values of function theoretical eventological regression of the bipartite sets of events, it is necessary to have statistics for estimation of probabilities of crossings of events (from numerator of formula (7)) and probabilities of the eventological distribution $p(ter(s))$ (from denominator of formula (7)). After comparison of the calculated estimations of conditional probabilities with order of the conditional α-kvantil we receive estimations of sets-values of function eventological regression for each event $ter(t)$, $t \subseteq Z$.

On picture 1 is shown the schedule of the function of the theoretical eventological regression of the bipartite sets of events for the elementary case when the bipartite set of events consists only of one set plural. Here mugs designate the events which have got to the schedule of function of the theoretical eventological regression.



Picture 1. The schedule of function of the theoretical eventological regression

References

1. I.V. Baranova. The bipartite set of events method in eventological analysis of social-economic system (In Russian) / Bulletin of Krasnoyarsk State University, Krasnoyarsk (2006) — № 1 — pp. 142-152.
2. O. Yu. Vorob'ov, I.V. Baranova. The bipartite set method application in eventological analysis of complex system (In Russian) Siberian Federal University, Krasnoyarsk (2007) 132 p.
3. O. Yu. Vorob'ov. Eventology (In Russian) Siberian Federal University, Krasnoyarsk (2007) 435 p.