

## СРЕДНЕВЕРОЯТНОЕ СОБЫТИЕ КАК МЕРА ЗАВИСИМОСТИ МНОЖЕСТВА СОБЫТИЙ

Ендаурова А.В.,

научный руководитель канд. физ.-мат. наук, доцент **Голденок Е.Е.**

*Сибирский федеральный университет*

Рассматриваются понятия сет-расстояние множества событий и средневероятное событие для множества событий имеющее аналогию с понятием среднемерного множества.

### Введение

Зависимость событий - одно из фундаментальных понятий теории вероятностей и эвентологии. Традиционной мерой зависимости событий как в теории вероятностей, так и в эвентологии является ковариация. О.Ю. Воробьевым было введено понятие сет-расстояния множества событий, которое можно интерпретировать как ещё один из способов измерения зависимости множеств событий.

Среднемерное множество - это средняя множественная характеристика случайного множества, значениями которого служат подмножества измеримого пространства с мерой, играющая для случайного множества ту же роль, которую для случайного элемента со значениями из линейного пространства играет математическое ожидание, или среднее значение.

### Парные ковариации событий и их метрические интерпретации

Рассмотрим множество из двух событий  $\mathcal{X} = \{x, y\}$  — дуплет событий. Пусть эти события  $x$  и  $y$  имеют индивидуальные вероятности  $\mathbf{P}(x)$  и  $\mathbf{P}(y)$ . В эвентологии *парная ковариация событий по пересечению* определяется величиной:

$$\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y). \quad (1.1)$$

Для кратности будем называть эту величину *парной ковариацией* событий  $x$  и  $y$ , поскольку она совпадает с парной ковариацией случайных величин  $1_x$  и  $1_y$  — индикаторов этих событий:

$$\text{Cov}(1_x, 1_y) = \mathbf{E}(1_x - \mathbf{P}(x))(1_y - \mathbf{P}(y))$$

- центральным смешанным моментом второго порядка с.в.  $1_x$  и  $1_y$ .

Парная ковариация событий имеет метрическую интерпретацию:

$$2\text{Kov}_{xy} = \mathbf{P}^i(x \Delta y) - \mathbf{P}(x \Delta y), \quad (1.2)$$

где

$$\mathbf{P}(x \Delta y) = \mathbf{P}(x \cup y) - \mathbf{P}(x \cap y)$$

— *вероятностное расстояние* между событиями  $x$  и  $y$  в общем случае, а

$$\mathbf{P}^i(x \Delta y) = \mathbf{P}(x \cup y) - \mathbf{P}(x)\mathbf{P}(y)$$

—*вероятностное расстояние* между независимыми событиями, имеющими те же вероятности, что и события  $x$  и  $y$ .

Соотношение (1.1) означает, что удвоенная парная ковариация событий — это отличие вероятностного расстояния между этими событиями от вероятностного расстояния между независимыми событиями с такими же вероятностями. Причем если вероятностное расстояние между событиями больше вероятностного расстояния между независимыми событиями с такими же вероятностями (отрицательная парная ковариация), то значит события «вероятностно отталкиваются», в противном случае (положительная парная ковариация) — «вероятностно притягиваются».

### Случайное множество событий

В эвентологии с каждым *множеством событий*  $\mathfrak{M} \subset \mathcal{A}$  традиционно связывается эквивалентное понятие *случайного множества событий*, которое определяется на всеобщем вероятностном пространстве  $(\Omega, \mathcal{A}, \mathbf{P})$  как случайный элемент

$$K_{\mathfrak{M}}: (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (2^{\mathfrak{M}}, 2^{2^{\mathfrak{M}}})$$

со значениями

$$K_{\mathfrak{M}}(\omega) = \{\mu \in \mathfrak{M}: \omega \in \mu\} \subseteq \mathfrak{M}$$

—подмножествами событий из  $\mathfrak{M}$ , наступающих при наступлении всеобщего элементарного исхода  $\omega \in \Omega$ .

### Средневероятное событие

В дополнении к стандартному определению *среднемерного множества* событий, дадим другое определение *среднемерного множества*. Речь идет о новом понятии *средневероятного события* для данного множества событий  $\mathfrak{M} \subset \mathcal{A}$ , которое обозначается

$$\hat{\mu}_{\mathfrak{M}} \subseteq \Omega,$$

наступает с вероятностью, равной

$$\mathbf{P}(\hat{\mu}_{\mathfrak{M}}) = \frac{1}{|\mathfrak{M}|} \sum_{\mu \in \mathfrak{M}} \mathbf{P}(\mu)$$

— средней вероятности событий  $\mu \in \mathfrak{M}$ , и играет роль средней множественной характеристики событий из  $\mathfrak{M}$ , как подмножество  $\Omega$ .

Определение (средневероятное событие). *Средневероятным событием* для множества событий  $\mathfrak{M} \subset \mathcal{A}$  называется такое всеобщее событие  $\hat{\mu}_{\mathfrak{M}} \in \mathcal{A}$ :

$$\sum_{|Y| > m} \text{ter}(Y // \mathfrak{M}) \subseteq \hat{\mu}_{\mathfrak{M}} \subseteq \sum_{|Y| \geq m} \text{ter}(Y // \mathfrak{M}),$$

которое наступает с вероятностью

$$P(\hat{\mu}_{\mathfrak{M}}) = \frac{1}{|\mathfrak{M}|} \sum_{\mu \in \mathfrak{M}} P(\mu)$$

всякий раз, когда среди событий из  $\mathfrak{M}$  наступает не менее, чем  $m$  событий, где  $m \in \{0, 1, \dots, |\mathfrak{M}|\}$  удовлетворяет неравенствам

$$\sum_{|Y| > m} p(Y // \mathfrak{M}) < P(\hat{\mu}_{\mathfrak{M}}) \leq \sum_{|Y| \geq m} p(Y // \mathfrak{M}).$$

### Сет-расстояния множеств событий

Обычно расстояние определяется лишь для пар элементов, а не для произвольных множеств элементов.

Определим на  $2^{\mathfrak{X}}$  такую *сет-функцию*  $\Delta_X$  – функцию множества  $X \subseteq \mathfrak{X}$ , которая бы, во-первых, обобщала понятие *вероятностного расстояния* между двумя событиями  $x, y \in \mathfrak{X}$ , т.е. была пропорциональна вероятности их симметрической разности, когда  $X = \{x, y\}$  – дуплет:

$$\Delta_{xy} = \frac{1}{2} P(x \Delta y)$$

а, во-вторых, для произвольного  $X \subseteq \mathfrak{X}$  обладала бы свойствами, напоминающими привычные свойства расстояния. Этому условию удовлетворяет сет-функция  $\Delta_X$  определяемая для  $X \subseteq \mathfrak{X}$  как

$$\Delta_X = P(\cup_{x \in X} x) - \frac{1}{|X|} \sum_{x \in X} P(x),$$

которая называется *сет-расстоянием множества событий*  $X \subseteq \mathfrak{X}$ .

### Сет-метрика

Два события. Рассмотрим два события  $x, y \in \mathcal{F}$ , образующее дуплет событий  $X = \{x, y\} \subset \mathcal{F}$ . *Вероятностным расстоянием* между двумя событиями служит полу-вероятность их симметрической разности  $\Delta_{xy} = \frac{1}{2} P(x \Delta y)$ , которую можно представить в виде следующей взвешенной суммы вероятностей террасных событий:

$$\Delta_{xy} = 1/2 \cdot P(x \cap y^c) + 1/2 \cdot P(x^c \cap y) + 0 \cdot P(x \cap y) + 0 \cdot P(x^c \cap y^c).$$

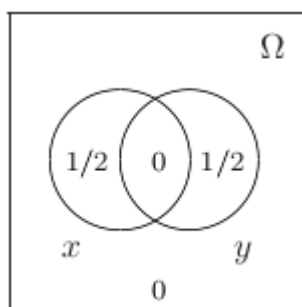


рис.1

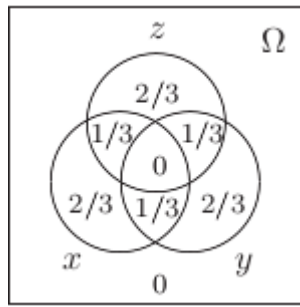


рис.2

Террасные «штрафы за несовпадение» множества  $\{x, y\} \subseteq \mathfrak{X}$  из двух событий  $x, y \subseteq \Omega$  (слева) и множества  $\{x, y, z\} \subseteq \mathfrak{X}$  из трёх событий  $x, y, z \subseteq \Omega$  (справа) для сет-расстояний этих множеств по объединению  $\Delta_{xy}, \Delta_{xyz}$  соответственно.

Веса вероятностей террасных событий имеют смысл «штрафов за несовпадение» событий. Наименьший «штраф» 0 имеет террасное событие  $x \cap y$ , так как его наступление означает одновременное наступление двух событий (т.е. события  $x$  и  $y$  совпадают на этом террасном событии); террасные события  $x \cap y^c$  и  $x^c \cap y$  штрафуются  $1/2$ , т.к. наступает только одно событие из двух (рис1).

Таким образом, *вес террасного события* – это доля не наступающих событий при его наступлении.

Три события.Применим ту же систему «террасных штрафов» для террасных событий, образованных пересечением трех событий  $x, y, z \in \mathcal{F}$  (рис 2). Тогда аналогичная взвешенная сумма, которую обозначим  $\Delta_{xyz}$ , имеет вид:

$$\Delta_{xyz} = 0 \cdot P(x \cap y \cap z) + 0 \cdot P(x^c \cap y^c \cap z^c) + 1/3 \cdot P(x \cap y \cap z^c) + 1/3 \cdot P(x \cap y^c \cap z) + 1/3 \cdot P(x^c \cap y \cap z) + 2/3 \cdot P(x \cap y^c \cap z^c) + 2/3 \cdot P(x^c \cap y \cap z^c) + 2/3 \cdot P(x^c \cap y^c \cap z).$$

Произвольное множество событий.Для произвольного множества событий  $X \subseteq \mathcal{F}$  аналогичная взвешенная сумма равна

$$\Delta_X = \sum_{\emptyset \subset Y \subseteq X} \left(1 - \frac{|Y|}{|X|}\right) P\left(\bigcap_{x \in Y} x \bigcap_{x \in X \setminus Y} x^c\right),$$

где суммирование распространяется на все непустые подмножества  $X$ . Сначала заметим, что

$$P\left(\bigcap_{x \in Y} x \bigcap_{x \in X \setminus Y} x^c\right) = \rho(X, Y) = P(K_X = Y)$$

– вероятность того, что среди множества событий  $X$  наступят только события из подмножества  $Y \subseteq X$ , т.е. вероятность того, что случайное множество наступивших событий  $K_X$ , определенное под  $X$ , примет значение  $Y$ .

Замечание 1. Под случайными множествами событий  $K_X$  понимаются проекции

$$K_X = K \cap X, X \subseteq \mathfrak{X}$$

случайного множества

$$K: (\Omega, \mathcal{F}, P) \rightarrow (2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}),$$

определенного под  $\mathfrak{X}$  распределением вероятностей

$$\rho(X) = P(K = X) = P\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right),$$

где  $X^c = \mathfrak{X} \setminus X$ .

Замечание 2. Сет-расстояние  $\Delta_X$  зависит только от распределения мощности случайного множества событий  $K_X$ :

$$\Delta_X = \sum_{\alpha=1}^{|\mathfrak{X}|} \left(1 - \frac{\alpha}{|\mathfrak{X}|}\right) P(|K_X| = \alpha).$$

А теперь перепишем формулу для сет-расстояния в виде

$$\Delta_X = \sum_{\emptyset \subset Y \subseteq X} \left(1 - \frac{|Y|}{|X|}\right) \rho(X, Y) - \rho(X, \emptyset) = 1 - \rho(X, \emptyset) - \frac{1}{|X|} \sum_{\emptyset \subset Y \subseteq X} |Y| \rho(X, Y)$$

и заметим, что, во-первых,

$$1 - \rho(X, \emptyset) = \mathbf{P}(K_X \neq \emptyset) = \mathbf{P}(K \cap X \neq \emptyset) = \mathbf{P}\left(\bigcup_{x \in X} x\right) = u_X,$$

во-вторых,

$$\sum_{\emptyset \subset Y \subseteq X} |Y| \rho(X, Y) = \sum_{\emptyset \subset Y \subseteq X} |Y| \mathbf{P}(K_X = Y) = \mathbf{E}|K_X|.$$

В итоге взвешенная сумма для произвольного множества событий  $X$  принимает вид:

$$\Delta_X = u_X - \frac{1}{|X|} \mathbf{E}|K_X|,$$

который можно выбрать в качестве определения новой характеристики  $\Delta_X$  множества событий  $X \subset \mathcal{F}$  - *сет-расстояния множества событий  $X$* . Прежде чем дать определение сет-расстояния, заметим только, что в силу «старой» теоремы Роббинса:

$$\mathbf{E}|K_X| = \sum_{x \in X} \mathbf{P}(x).$$

Определение (сет-расстояние множества событий).

*Сет-расстояние множества событий* – это сет-функция, определяемая для каждого множества событий  $X \subset \mathfrak{X} \subseteq \mathcal{F}$  как величина

$$\Delta_X = \mathbf{P}\left(\bigcup_{x \in X} x\right) - \frac{1}{|X|} \sum_{x \in X} \mathbf{P}(x).$$