ТЕОРИЯ ВЕРОЯТНОСТЕЙ

УДК 519.2

МЕТРИЧЕСКИЕ ЭВЕНТОЛОГИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ

О.Ю.Воробьёв, Е.Е.Голденок*

Рассматривается способ задания эвентологических распределений, основанный на новом понятии сет-расстояния, *определяемого для произвольного множества случайных событий.* Эвентологическое распределение *— это любое распределение вероятностей, определяющее множество случайных событий $\mathfrak{X} \subseteq \mathcal{F}$, выбранных из алгебры $\mathcal{F}$ некоторого вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$. Одно и то же множество случайных событий можно эквивалентным образом определить, задавая любое одно из следующих распределений вероятностей пересечений или объединений случайных событий [3]: $p(X)$, p_X, p^X, $u(X)$, u_X, u^X, $\quad X \subseteq \mathfrak{x}$, где, например, $p_X = \mathbf{P}\left(\bigcap_{x\in X} x\right)$, $u_X = \mathbf{P}\left(\bigcup_{x\in X} x\right)$, $\quad X \subseteq \mathfrak{X}$, — распределения вероятностей пересечений и объединений соответственно. Поскольку $p_X = \prod_{x\in X}\mathbf{P}(x) - \mathsf{Kov}_X$, $\quad X \subseteq \mathfrak{X}$, то распределение* вероятностей пересечений *p_X, можно определить, задав вероятности событий $\mathbf{P}(x)$, $x \in \mathfrak{X}$, и распределение арных ковариаций Kov_X, $X \subseteq \mathfrak{X}$. В некоторых задачах, например в задаче N случайных событий [5], для задания эвентологических распределений удобнее и естественнее вместо арной ковариации множества событий — меры статистических взаимозависимостей случайных событий, использовать другую характеристику множества событий, которая бы имела свойства расстояния между ними. Однако, расстояние обычно определяется только для пар элементов, а не для произвольных множеств элементов. В работе вводится новое понятие* ***сет-расстояния*** *$\Delta_X = |X| \cdot u_X - \sum_{x\in X}\mathbf{P}(x)$* ***множества случайных событий*** *$X \subseteq \mathfrak{X}$. Так определенное* ***сет-расстояние*** *Δ_X — это функция множества $X \subseteq \mathfrak{X}$, обобщающая расстояние между двумя событиями $x, y \in \mathfrak{X}$ (которое в новой терминологии "превращается" в* ***сет-расстояние дуплета событий*** *$X = \{x, y\}$), обычно определяемое как вероятность их симметрической разности: $\Delta_{xy} = \mathbf{P}(x\triangle y)$. В работе рассматриваются свойства сет-расстояния и доказывается*

Теорема (метрическое задание эвентологических распределений). *Любое эвентологическое распределение множества случайных событий $\mathfrak{X}$ определяется заданием их вероятностей $\mathbf{P}(x)$, $x \in \mathfrak{X}$, и сет-расстояний Δ_X всех подмножеств случайных событий $X \subseteq \mathfrak{X}$.*

Эвентологическое распределение — это любое распределение вероятностей, определяющее множество случайных событий $\mathfrak{X} \subseteq \mathcal{F}$, выбранных из алгебры $\mathcal{F}$ некоторого вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$. В ряде задач, например в задаче N случайных событий [5], для задания эвентологических распределений удобнее и естественнее использовать такую характеристику множества случайных событий, которая имеет свойства расстояния между ними. Однако расстояние обычно определяется лишь для пар элементов, а не для произвольных множеств элементов.

Поставим задачу определить на $2^{\mathfrak{X}}$ такую *сет-функцию* Δ_X — функцию множества $X \subseteq \mathfrak{X}$, которая бы, во-первых, обобщала понятие обычного расстояния между двумя случайными событиями $x, y \in \mathfrak{X}$, т.е. совпадала с вероятностью их симметрической разности, когда $X = \{x, y\}$ — дуплет: $\Delta_{xy} = \mathbf{P}(x\triangle y)$, а, во-вторых, для произвольного $X \subseteq \mathfrak{X}$ обладала бы свойствами, напоминающими привычные свойства расстояния. Как выяснится ниже, этим условиям вполне удовлетворяет сет-функция Δ_X, определяемая для $X \subseteq \mathfrak{X}$ как $\Delta_X = |X| \cdot u_X - \sum_{x\in X}\mathbf{P}(x)$, где $u_X = \mathbf{P}\left(\bigcup_{x\in X} x\right)$ — вероятности объединения случайных событий. Сет-функция Δ_X называется *сет-расстоянием множества случайных событий* $X \subseteq \mathfrak{X}$.

1. Метрика

Определение (метрика). *Метрика* на множестве $\mathfrak{X}$ (*расстояние между двумя элементами* x, y *множества* $\mathfrak{X}$) — это действительная числовая функция $\rho(x, y)$, определенная на $\mathfrak{X} \times \mathfrak{X}$ и удовлетворяющая для любых $x, y, z \in \mathfrak{X}$ двум аксиомам:

1) $\rho(x, y) = 0 \iff x = y \quad$ (*аксиома нуля*),

2) $\rho(x,z)+\rho(y,z) \geqslant \rho(x,y)$ *(аксиома треугольника).*

Замечание 1. Из определения вытекают два свойства расстояния:

- $\rho(x,y) \geqslant 0$ *(неотрицательность),*
- $\rho(x,y) = \rho(y,x)$ *(симметричность).*

Действительно, если в аксиоме треугольника, во-первых, положить $y = x$, то получим *неотрицательность*: $2\rho(x,y) \geqslant 0$, а во-вторых, положить $z = x$, то получим *симметричность*, так как из возникающего неравенства $\rho(y,x) \geqslant \rho(x,y)$ сразу же следует ему противоположное: $\rho(x,y) \geqslant \rho(y,x)$.
Замечание 2. Если для функции ρ выполнена только одна аксиома треугольника, то она называется *псевдометрикой (квазиметрикой)*, или *псевдорасстоянием (квазирасстоянием).*
Замечание 3. Когда $\mathfrak{X} \subseteq \mathcal{F}$ — конечное множество случайных событий, то действительная числовая функция $\Delta_{xy} = \mathbf{P}(x \triangle y)$, $\quad x,y \in \mathfrak{X}$, — *вероятность симметрической разности двух событий x и y*, будет метрикой на $\mathfrak{X}$ всякий раз, когда вероятность $\mathbf{P}$ такова, что $\mathbf{P}(x) = 0 \iff x = \emptyset$. Иначе $\mathbf{P}(x \triangle y)$ — псевдометрика.

2. Сет-метрика по объединению

Два события. Рассмотрим два случайных события $x, y \in \mathcal{F}$, образующие множество (дуплет) $X = \{x, y\} \subset \mathcal{F}$. Обычным расстоянием между двумя событиями служит вероятность их симметрической разности $\mathbf{P}(x \triangle y)$, которую можно представить в виде следующей взвешенной суммы вероятностей террасок: $\mathbf{P}(x \triangle y) = 0 \cdot \mathbf{P}(x \cap y) + 1 \cdot \mathbf{P}(x \cap y^c) + 1 \cdot \mathbf{P}(x^c \cap y)$, содержащихся в объединении двух событий: $x \cup y = x \cap y + x \cap y^c + x^c \cap y$.

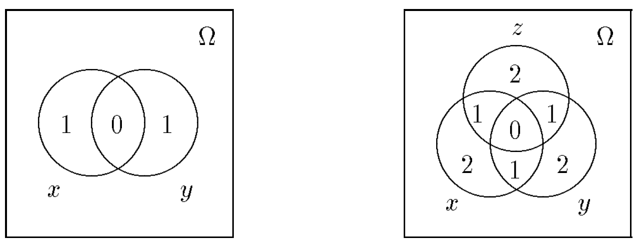

Рис. 1. Террасные "штрафы за несовпадение" множества $\{x,y\} \subseteq \mathfrak{X}$ из двух случайных событий $x, y \subseteq \Omega$ (слева) и множества $\{x,y,z\} \subseteq \mathfrak{X}$ из трех случайных событий $x,y,z \subseteq \Omega$ (справа) для сет-расстояний этих множеств по объединению: $\Delta^{\cup}_{xy}$ и $\Delta^{\cup}_{xyz}$

Веса вероятностей террасок имеют смысл "штрафов за несовпадение" случайных событий. Наименьший "штраф" 0 имеет терраска $x \cap y$, так как наступление этой терраски означает одновременное наступление двух событий (т.е. события x и y совпадают на этой терраске); терраски $x \cap y^c$ и $x^c \cap y$ штрафуются единицей, так как наступает только одно событие из двух (рис. 1 слева). Таким образом, *вес терраски — это число ненаступивших случайных событий* при наступлении данной терраски.

Три события. Применим ту же систему "террасных штрафов" для террасок, образованных пересечением трех событий $x, y, z \in \mathcal{F}$ (рис. 1 справа). Тогда аналогичная взвешенная сумма, которую мы обозначим Δ_{xyz}, имеет вид $\Delta_{xyz} = 0 \cdot \mathbf{P}(x \cap y \cap z) + 1 \cdot \mathbf{P}(x \cap y \cap z^c) + 1 \cdot \mathbf{P}(x \cap y^c \cap z) + 1 \cdot \mathbf{P}(x^c \cap y \cap z)$ $+2 \cdot \mathbf{P}(x \cap y^c \cap z^c) + 2 \cdot \mathbf{P}(x^c \cap y \cap z^c) + 2 \cdot \mathbf{P}(x^c \cap y^c \cap z)$.

Произвольное множество событий. Для произвольного множества событий $X \subset \mathcal{F}$ аналогичная взвешенная сумма равна $\Delta_X = \sum_{\emptyset \subset Y \subseteq X} (|X| - |Y|) \mathbf{P} \left(\bigcap_{x \in Y} x \bigcap_{x \in X \setminus Y} x^c \right)$, где суммирование распространяется на все непустые подмножества X. Сначала заметим, что $\mathbf{P} \left(\bigcap_{x \in Y} x \bigcap_{x \in X \setminus Y} x^c \right) = p(X, Y) = \mathbf{P}(K_X = Y)$ — 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 4. Под случайными множествами событий K_X понимаются проекции $K_X = K \cap X$, $\quad X \subseteq \mathfrak{X}$ случайного множества $K : (\Omega, \mathcal{F}, \mathbf{P}) \to \left(2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}} \right)$, определенного под $\mathfrak{X}$ распределением вероятностей $p(X) = \mathbf{P}(K = X) = \mathbf{P} \left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c \right)$, где $X^c = \mathfrak{X} \setminus X$.
Замечание 5. Сет-расстояние Δ_X зависит только от распределения мощности случайного множества событий K_X: $\Delta_X = \sum_{a=1}^{|X|} (|X| - a) \mathbf{P}(|K_X| = a)$.

А теперь перепишем формулу для сет-расстояния в виде

$$\Delta_X = \sum_{\emptyset \subseteq Y \subseteq X} (|X| - |Y|) p(X,Y) - |X| p(X, \emptyset) =$$

$$= |X| (1 - p(X,\emptyset)) - \sum_{\emptyset \subseteq Y \subseteq X} |Y| p(X,Y)$$

и заметим, что, во-первых,

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

во-вторых,

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

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

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

который можно выбрать в качестве определения новой характеристики Δ_X множества случайных событий $X \subset \mathcal{F}$ — *сет-расстояния множества случайных 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 случайных событий). *Сет-расстояние множества случайных событий* — это сет-функция, определяемая для каждого множества случайных событий $X \subset \mathcal{F}$ как величина

$$\Delta_X = |X| \cdot \mathbf{P}\left(\bigcup_{x\in X} x\right) - \sum_{x\in X} \mathbf{P}(x). \tag{2}$$

Определение (расстояние события до множества событий). *Расстоянием случайного события $x \in \mathfrak{X}$ до множества событий $X \subseteq \mathfrak{X}$* называется действительная числовая функция $d : \mathfrak{X} \times 2^{\mathfrak{X}} \to \mathbb{R}$, которая на $(x,X) \in \mathfrak{X} \times 2^{\mathfrak{X}}$ принимает значение $d(x,X) = \mathbf{P}\left\{x \bigtriangleup \left(\bigcup_{y\in X} y\right)\right\}$.

Определение (расстояние между множествами событий). *Расстоянием между множествами событий $X \subseteq \mathfrak{X}$ и $Y \subseteq \mathfrak{X}$* называется действительная числовая функция $d : 2^{\mathfrak{X}} \times 2^{\mathfrak{X}} \to \mathbb{R}$, которая на $(X,Y) \in 2^{\mathfrak{X}} \times 2^{\mathfrak{X}}$ принимает значение, равное вероятности симметрической разности объединений этих двух множеств событий: $d(X,Y) = \mathbf{P}\left\{\left(\bigcup_{x\in X} x\right) \bigtriangleup \left(\bigcup_{x\in Y} x\right)\right\}$.

Замечание 6. Для расстояний события до множества событий и расстояния между множествами событий не выполняется аксиома нуля. Поэтому, строго говоря, речь идет о сет-псевдометриках (сет-квазиметриках) и сет-псевдорасстояниях (сет-квазирасстояниях). Но, чтобы не отягощать терминологию будем продолжать опускать "псевдо-" или "квази-", всегда помня, однако, что аксиома нуля для этих сет-метрик не выполняется.

Замечание 7. Расстояние события до множества событий — это частный случай расстояния между множествами событий, когда одно из множеств — моноплет: $d(x,X) = d(\{x\},X)$.

Замечание 8. Только что введенные расстояния проще всего выразить через распределение вероятностей объединений: $u_X = \mathbf{P}\left(\bigcup_{x\in X} x\right)$. Расстояние между множествами событий X и Y равно $d(X,Y) = 2u_{X\cup Y} - u_X - u_Y$. Если же $Y \subseteq X$, то $d(X,Y) = u_X - u_Y$, так как тогда $X \cup Y = X$. В частности, для расстояния события x до множества событий X формула имеет вид $d(x,X) = 2u_{X\cup\{x\}} - u_X - u_{\{x\}}$. Если же $x \in X$, то $d(x,X) = u_X - u_{\{x\}}$.

Лемма 1. *Сет-расстояние Δ_X множества событий $X \subseteq \mathfrak{X}$ представимо в виде суммы расстояний всех событий из X до самого множества событий X: $\Delta_X = \sum_{x\in X} d(x,X)$.*

Доказательство. Из (2) следует, что $\Delta_X = \sum_{x\in X}\left\{\mathbf{P}\left(\bigcup_{y\in X} y\right) - \mathbf{P}(x)\right\}$. Кроме того, для $x \in X$ $x \bigtriangleup \left(\bigcup_{y\in X} y\right) = \left(\bigcup_{y\in X} y\right) \setminus x$, так как $x \subseteq \left(\bigcup_{y\in X} y\right)$. Это означает, что

$$\Delta_X = \sum_{x\in X} \mathbf{P}\left\{x \bigtriangleup \left(\bigcup_{y\in X} y\right)\right\},$$

что доказывает лемму. □

3. Сет-метрика по пересечению

Можно ввести еще одну характеристику расстояния множества случайных событий, двойственную только что рассмотренной. В ее определении операция объединения заменяется двойственной ей операцией пересечения: $\Delta^{\cap}_X = \sum_{x\in X} \mathbf{P}(x) - |X| \cdot \mathbf{P}\left(\bigcap_{x\in X} x\right)$. Называется эта характеристика *сет-расстоянием множества случайных событий X по пересечению.* Схема "террасных штрафов за несовпадение" для этого случая показана на рис. 2.

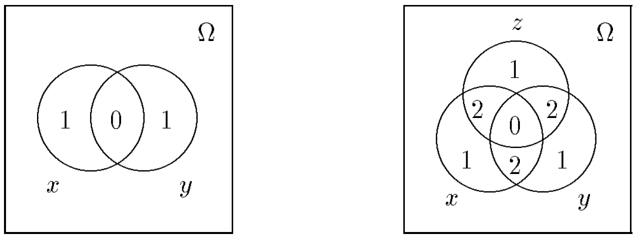

Рис. 2. Террасные "штрафы за несовпадение" множества $\{x,y\} \subseteq \mathfrak{X}$ из двух случайных событий $x,y \subseteq \Omega$ (слева) и множества $\{x,y,z\} \subseteq \mathfrak{X}$ из трех случайных событий $x,y,z \subseteq \Omega$ (справа) для сет-расстояний этих множеств по пересечению: $\Delta^{\cap}_{xy}$ и $\Delta^{\cap}_{xyz}$

Приведем еще три формулы, справедливые для сет-расстояния множества событий (по пересечению):

$$\Delta^{\cap}_X = \sum_{\emptyset \subseteq Y \subset X} |Y| p(Y), \quad \Delta^{\cap}_X = \sum_{m=0}^{|X|-1} m\mathbf{P}(|K_X| = m), \quad \Delta^{\cap}_X = \sum_{m=0}^{|X|-1} m\mathbf{P}(C^m_X).$$

Отметим еще, что в основе определения сет-расстояния по пересечению лежат соответствующие понятия *расстояния случайного события от множества событий (по пересечению)*: $d^{\cap}(x,X) = \mathbf{P}\left\{x \,\triangle \left(\bigcap_{y\in X} y\right)\right\}$, и более общего понятия — *расстояния между множествами событий (по пересечению)*: $d^{\cap}(X,Y) = \mathbf{P}\left\{\left(\bigcap_{x\in X} x\right) \triangle \left(\bigcap_{x\in Y} x\right)\right\}$. Аналогичное этому понятие расстояния между подмножествами абстрактного случайного множества было введено в [6] под названием вероятностной псевдометрики, правда, в менее отчетливой форме и в излишне громоздких и потому неудобных обозначениях.

Свойства данных *расстояний относительно пересечения* аналогичны соответствующим свойствам *расстояний относительно объединения.*

Лемма 2. *Сет-расстояние $\Delta^{\cap}_X$ множества событий $X \subseteq \mathfrak{X}$ представимо в виде суммы расстояний всех событий из X до самого множества событий X: $\Delta^{\cap}_X = \sum_{x\in X} d^{\cap}(x,X)$.*

Замечание 9. Рассмотренные расстояния по пересечению проще всего выразить через распределение вероятностей пересечений: $p_X = \mathbf{P}\left(\bigcap_{x\in X} x\right)$. Расстояние между множествами событий X и Y равно $d^{\cap}(X,Y) = p_X + p_Y - 2p_{X\cup Y}$. Если же $Y \subseteq X$, то $d^{\cap}(X,Y) = p_Y - p_X$, так как тогда $X \cup Y = X$. В частности, для расстояния события x до множества событий X формула имеет вид $d^{\cap}(x,X) = d^{\cap}(\{x\},X) = p_{\{x\}} + p_X - 2p_{X\cup\{x\}}$, а если $x \in X$, то $d^{\cap}(x,X) = p_{\{x\}} - p_X$.

Заметим еще, что $\Delta^{\cup}_X + \Delta^{\cap}_X = |X| \left\{\mathbf{P}\left(\bigcup_{x\in X} x\right) - \mathbf{P}\left(\bigcap_{x\in X} x\right)\right\} = |X| \cdot (u_X - p_X)$.

4. Свойства сет-расстояния

1. Неотрицательность. Сет-расстояние — неотрицательная сет-функция: $\Delta_X \geqslant 0, \quad X \subseteq \mathfrak{X}$. Это следует из того, что $\Delta_X = \sum_{x\in X} \left\{\mathbf{P}\left(\bigcup_{x\in X} x\right) - \mathbf{P}(x)\right\}$, а для любого $X \subseteq \mathfrak{X}$ и каждого $x \in X$ $\mathbf{P}\left(\bigcup_{x\in X} x\right) - \mathbf{P}(x) \geqslant 0$.

2. Неравенство треугольника. Сет-расстояние должно удовлетворять неравенству треугольника, которое можно пытаться записать несколькими способами. Нами выбрано неравенство следующего вида:

$$\Delta_{X\cap\{x\}^c} + \Delta_{X\cap\{y\}^c} \geqslant \Delta_{X\cap\{z\}^c}, \tag{3}$$

которое выполняется для любого $X \subseteq \mathfrak{X}$ любых $x,y,z \in X$. Заметим, что условие $x,y,z \in X$, скорее всего, можно ослабить: $x,y,z \in \mathcal{F}$.

В частности, когда X — триплет: $X = \{x,y,z\}$, тогда $X \cap \{z\}^c = \{x,y\}$, $X \cap \{y\}^c = \{x,z\}$, $X \cap \{x\}^c = \{y,z\}$, и неравенство (3) превращается в обычное неравенство треугольника: $\Delta_{xz} + \Delta_{yz} \geqslant \Delta_{xy}$, которое выполняется и означает, что $\mathbf{P}(x\triangle z) + \mathbf{P}(y\triangle z) \geqslant \mathbf{P}(x\triangle y)$.

Если договориться обозначать $X - x = X \cap \{x\}^c$, то неравенство (3) можно записать лаконичнее:

$$\Delta_{X-x} + \Delta_{X-y} \geqslant \Delta_{X-z}. \tag{3'}$$

Заметим, что в этом неравенстве множество X, вроде бы, должно содержать триплет $\{x, y, z\}$. Хотя это условие, скорее всего, можно убрать.

5. Доказательство неравенства треугольника для сет-расстояния

Будем доказывать неравенство треугольника $\Delta_{X'-x}+\Delta_{X'-y} \geqslant \Delta_{X'-z}$, где $X' = X+\{x, y, z\} \subseteq \mathfrak{X}$, которое перепишем в эквивалентном виде

$$\Delta_{X+\{y,z\}} + \Delta_{X+\{x,z\}} \geqslant \Delta_{X+\{x,y\}}, \tag{3''}$$

где $X \subseteq \{x, y, z\}^c = \mathfrak{X} \setminus \{x, y, z\}$.

Назовем события $C_X^m = \left\{\omega \in \Omega : \sum_{x\in X} \mathbf{1}_x(\omega) = m\right\} \subseteq \Omega, \quad m = 0, \ldots, |X|$, *$m$-слоями множества событий* $X \subseteq \mathfrak{X}$. Событие C_X^m наступает тогда и только тогда, когда наступают точно m событий из множества событий X. Ясно, что для каждого $X \subseteq \mathfrak{X}$ все соответствующие ему m-слои, $m = 0, \ldots, |X|$, образуют так называемое X-разбиение пространства элементарных событий Ω, т.е. $\sum_{m=0}^{|X|} C_X^m = \Omega$, и $C_X^m \cap C_X^{m'} = \emptyset \quad \iff \quad m \neq m'$.

Лемма 3. $\Delta_X = \sum_{m=1}^{|X|}(|X| - m)\mathbf{P}\left(C_X^m\right).$

Доказательство. В силу определения сет-расстояния имеем

$$\Delta_X = \sum_{\emptyset\subset Y\subseteq X} (|X| - |Y|)\mathbf{P}\left(\bigcap_{x\in X} x \bigcap_{x\in X\setminus Y} x\right) =$$

$$= \sum_{m=0}^{|X|}(|X| - m) \sum_{|Y|=m} \mathbf{P}\left(\bigcap_{x\in X} x \bigcap_{x\in X\setminus Y} x\right).$$

А поскольку $\sum_{|Y|=m} \mathbf{P}\left(\bigcap_{x\in X} x\bigcap_{x\in X\setminus Y} x\right) = \mathbf{P}\left\{\omega \in \Omega : \sum_{x\in X} \mathbf{1}_x(\omega) = m\right\} = \mathbf{P}(C_X^m)$, то лемма доказана. □

Лемма 4 (неравенство треугольника для сет-расстояния по объединению). *Для каждого $X \subseteq \mathcal{F}$ и любых $x, y, z \in \mathcal{F}$*

$$\Delta_{X+\{y,z\}} + \Delta_{X+\{x,z\}} \geqslant \Delta_{X+\{x,y\}}. \tag{3''}$$

Доказательство. Поскольку слои образуют разбиение Ω, то они разбивают также и любое другое событие из $\mathcal{F}$, в том числе и любое событие $x \in \mathfrak{X}$: $x = x \cap C_X^0 + x \cap C_X^1 + \ldots + x \cap C_X^{|X|}$. Данное разбиение будем называть *слойным разбиением события x относительно X*. Зафиксируем X и обозначим $x^m = x \cap C_X^m$, $m = 0, \ldots, |X|$, тогда $x = x^0 + x^1 + \ldots + x^{|X|}$. Выразим сет-расстояние $\Delta_{X+\{x,y\}} = \sum_{m=1}^{|X|+2}(|X| + 2 - m)\mathbf{P}\left(C_{X+\{x,y\}}^m\right)$ через вероятности m-слоев C_X^m и вероятности фрагментов соответствующих слойных разбиений событий x и y относительно множества событий X. Обратим внимание, что аналогичным слойным разбиением обладает также и объединение событий x и y, а фрагменты этого разбиения являются объединениями соответствующих фрагментов слойных разбиений каждого события: $x \cup y = x^0 \cup y^0 + x^1 \cup y^1 + \ldots + x^{|X|} \cup y^{|X|}$. Искомая формула имеет вид $\Delta_{X+\{x,y\}} = \Delta_{x^0y^0} + |X| \cdot \mathbf{P}(x^0 \cup y^0) + \sum_{m=1}^{|X|} \Big\{(|X| - m + 2)\mathbf{P}(C_X^m) + \Delta_{x^my^m} - 2\mathbf{P}(x^m \cup y^m)\Big\}$. Слегка преобразуем эту формулу: $\Delta_{X+\{x,y\}} = \Delta_{x^0y^0} + |X| \cdot \mathbf{P}(x^0 \cup y^0) + \sum_{m=1}^{|X|} \Big\{(|X| - m)\mathbf{P}(C_X^m) + \Delta_{x^my^m} + 2(\mathbf{P}(C_X^m) - \mathbf{P}(x^m \cup y^m))\Big\}$ и запишем ее в преобразованном виде для остальных двух членов неравенства треугольника: $\Delta_{X+\{x,z\}} = \Delta_{x^0z^0} + |X| \cdot \mathbf{P}(x^0 \cup z^0) + \sum_{m=1}^{|X|} \Big\{(|X| - m)\mathbf{P}(C_X^m) + \Delta_{x^mz^m} + 2(\mathbf{P}(C_X^m) - \mathbf{P}(x^m \cup z^m))\Big\}$, $\Delta_{X+\{y,z\}} = \Delta_{y^0z^0} + |X| \cdot \mathbf{P}(y^0 \cup z^0) + \sum_{m=1}^{|X|} \Big\{(|X| - m)\mathbf{P}(C_X^m) + \Delta_{y^mz^m} + 2(\mathbf{P}(C_X^m) - \mathbf{P}(y^m \cup z^m))\Big\}$. Заметим, что одно и то же неотрицательное слагаемое $\sum_{m=1}^{|X|}(|X| - m)\mathbf{P}(C_X^m)$ присутствует во всех трех членах неравенства, поэтому если показать, что $\Delta_{x^0z^0}+|X|\cdot\mathbf{P}(x^0\cup z^0)+\sum_{m=1}^{|X|}\left\{\Delta_{x^mz^m}+2(\mathbf{P}(C_X^m)-\mathbf{P}(x^m\cup z^m))\right\}$ $+\Delta_{y^0z^0}+|X|\cdot\mathbf{P}(y^0\cup z^0)+\sum_{m=1}^{|X|}\left\{\Delta_{y^mz^m}+2(\mathbf{P}(C_X^m)-\mathbf{P}(y^m\cup z^m))\right\}\geqslant\Delta_{x^0y^0}+|X|\cdot\mathbf{P}(x^0\cup y^0)+$ $\sum_{m=1}^{|X|}\left\{\Delta_{x^my^m}+2(\mathbf{P}(C_X^m)-\mathbf{P}(x^m\cup y^m))\right\}$, то неравенство треугольника для сет-расстояния будет доказано. Сначала покажем, что $\Delta_{x^0z^0}+|X|\cdot\mathbf{P}(x^0\cup z^0)+\Delta_{y^0z^0}+|X|\cdot\mathbf{P}(y^0\cup z^0)\geqslant\Delta_{x^0y^0}+|X|\cdot\mathbf{P}(x^0\cup y^0)$. Это следует из того, что, во-первых, $\Delta_{x^0z^0}+\Delta_{y^0z^0}\geqslant\Delta_{x^0y^0}$, так как сет-расстояние дуплета — это вероятностное расстояние между двумя событиями, для которого выполнено неравенство треугольника: $\mathbf{P}(x^0\Delta z^0)+\mathbf{P}(y^0\Delta z^0)\geqslant\mathbf{P}(x^0\Delta y^0)$, и во-вторых, $\mathbf{P}(x^0\cup z^0)+\mathbf{P}(y^0\cup z^0)\geqslant\mathbf{P}(x^0\cup y^0)$, так как $\mathbf{P}(x^0\cup z^0)+\mathbf{P}(y^0\cup z^0)\geqslant\mathbf{P}(x^0\cup y^0\cup z^0)\geqslant\mathbf{P}(x^0\cup y^0)$.

Теперь покажем, что для каждого $m=1,\ldots,|X|$ выполнено $\Delta_{x^mz^m}+2(\mathbf{P}(C_X^m)-\mathbf{P}(x^m\cup z^m))+$ $\Delta_{y^mz^m}+2(\mathbf{P}(C_X^m)-\mathbf{P}(y^m\cup z^m))\geqslant\Delta_{x^my^m}+2(\mathbf{P}(C_X^m)-\mathbf{P}(x^m\cup y^m))$. Заменим $\mathbf{P}(C_X^m)$ в каждом из трех членов неравенства равным выражением: $\mathbf{P}(C_X^m)-\mathbf{P}(x^m\cup y^m\cup z^m)+\mathbf{P}(x^m\cup y^m\cup z^m)$. Поскольку $x^m\cup y^m\cup z^m\subseteq C_X^m$, то $\mathbf{P}(C_X^m)-\mathbf{P}(x^m\cup y^m\cup z^m)\geqslant 0$. Удалим эту неотрицательную разность из всех трех членов неравенства (тем самым из левой части неравенства удаляется в 2 раза больше, чем из правой). Теперь если для оставшихся слагаемых выполнены неравенства: $\Delta_{x^mz^m}+2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(x^m\cup z^m))+\Delta_{y^mz^m}+2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(y^m\cup z^m))\geqslant\Delta_{x^my^m}+$ $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(x^m\cup y^m))$, то исходные неравенства выполнены для каждого $m=1,\ldots,|X|$.

Преобразуем отдельные слагаемые в неравенствах, выражая их через вероятности террасок трех событий x, y и z. Слагаемые правой части: $\Delta_{x^my^m}=\mathbf{P}(x^m\Delta y^m)=\mathbf{P}(x^m\cap(y^m)^c\cap(z^m)^c)+\mathbf{P}((x^m)^c\cap y^m\cap(z^m)^c)+\mathbf{P}(x^m\cap(y^m)^c\cap z^m)+\mathbf{P}((x^m)^c\cap y^m\cap z^m)$, $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(x^m\cup y^m))=2\mathbf{P}((x^m)^c\cap(y^m)^c\cap z^m)$. Слагаемые первого члена левой части: $\Delta_{x^mz^m}=\mathbf{P}(x^m\Delta z^m)=\mathbf{P}(x^m\cap(y^m)^c\cap(z^m)^c)+\mathbf{P}((x^m)^c\cap(y^m)^c\cap z^m)+\mathbf{P}(x^m\cap y^m\cap(z^m)^c)+\mathbf{P}((x^m)^c\cap y^m\cap z^m)$, $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(x^m\cup z^m))=2\mathbf{P}((x^m)^c\cap y^m\cap(z^m)^c)$. Слагаемые второго члена левой части: $\Delta_{y^mz^m}=\mathbf{P}(y^m\Delta z^m)=\mathbf{P}((x^m)^c\cap y^m\cap(z^m)^c)+\mathbf{P}((x^m)^c\cap(y^m)^c\cap z^m)+\mathbf{P}(x^m\cap y^m\cap(z^m)^c)+\mathbf{P}(x^m\cap(y^m)^c\cap z^m)$, $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(y^m\cup z^m))=2\mathbf{P}(x^m\cap(y^m)^c\cap(z^m)^c)$. Перепишем полученные громоздкие формулы в понятно сокращенных обозначениях, где 1 соответствует событию, а 0 — его дополнению, так что, например, $\mathbf{P}_{100}=\mathbf{P}(x^m\cap(y^m)^c\cap(z^m)^c)$. Слагаемые правой части принимают вид: $\Delta_{x^my^m}=\mathbf{P}_{100}+\mathbf{P}_{010}+\mathbf{P}_{101}+\mathbf{P}_{011}$, $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(x^m\cup y^m))=2\mathbf{P}_{001}$. Слагаемые первого члена левой части: $\Delta_{x^mz^m}=\mathbf{P}_{100}+\mathbf{P}_{001}+\mathbf{P}_{110}+\mathbf{P}_{011}$, $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(x^m\cup z^m))=2\mathbf{P}_{010}$. Слагаемые второго члена левой части: $\Delta_{y^mz^m}=\mathbf{P}_{010}+\mathbf{P}_{001}+\mathbf{P}_{110}+\mathbf{P}_{101}$, $2(\mathbf{P}(x^m\cup y^m\cup z^m)-\mathbf{P}(y^m\cup z^m))=2\mathbf{P}_{100}$. В этих обозначениях вся левая часть неравенств имеет вид: $3\mathbf{P}_{100}+3\mathbf{P}_{010}+2\mathbf{P}_{001}+2\mathbf{P}_{110}+\mathbf{P}_{011}+\mathbf{P}_{101}$, а правая часть равна $\mathbf{P}_{100}+\mathbf{P}_{010}+2\mathbf{P}_{001}+\mathbf{P}_{011}+\mathbf{P}_{101}$. Таким образом, разность левой и правой частей — это неотрицательная величина: $2\mathbf{P}_{100}+2\mathbf{P}_{010}+\mathbf{P}_{110}\geqslant 0$. Неравенство треугольника для сет-расстояния по объединению доказано. □

Лемма 5 (неравенство треугольника для сет-расстояния по пересечению). *Для каждого $X\subseteq\mathcal{F}$ и любых $x,y,z\in\mathcal{F}$*

$$\Delta^{\cap}_{X+\{y,z\}}+\Delta^{\cap}_{X+\{x,z\}}\geqslant\Delta^{\cap}_{X+\{x,y\}}. \tag{3''}$$

Доказательство. Пусть $x=x^0+x^1+\ldots+x^{|X|}$ слойное разбиение события x относительно фиксированного X. Выразим сет-расстояние $\Delta^{\cap}_{X+\{x,y\}}=\sum_{m=0}^{|X|+1}m\mathbf{P}\left(C^m_{X+\{x,y\}}\right)$ через вероятности m-слоев C_X^m и вероятности фрагментов соответствующих слойных разбиений событий x и y относительно множества событий X. Обратим внимание, что аналогичным слойным разбиением обладает также и объединение событий x и y, а фрагменты этого разбиения являются объединениями соответствующих фрагментов слойных разбиений каждого события: $x\cup y=x^0\cup y^0+x^1\cup y^1+\ldots+x^{|X|}\cup y^{|X|}$. Искомая формула имеет вид

$$\Delta^{\cap}_{X+\{x,y\}}=\sum_{m=0}^{|X|}\left\{m\mathbf{P}\left(C_X^m\right)+\mathbf{P}(x^m)+\mathbf{P}(y^m)\right\}-(|X|+2)\mathbf{P}\left(x^{|X|}\cap y^{|X|}\right).$$

Перепишем эту формулу в виде:

$$\Delta^{\cap}_{X+\{x,y\}}=\sum_{m=0}^{|X|-1}\left\{m\mathbf{P}\left(C_X^m\right)+\mathbf{P}(x^m)+\mathbf{P}(y^m)\right\}+|X|\left(\mathbf{P}\left(C_X^{|X|}\right)-\right.$$

$$\left.-\mathbf{P}\left(x^{|X|}\cup y^{|X|}\cup z^{|X|}\right)\right)+|X|\mathbf{P}\left(x^{|X|}\cup y^{|X|}\cup z^{|X|}\right)+$$

$$+\mathbf{P}\left(x^{|X|}\right)+\mathbf{P}\left(y^{|X|}\right)-(|X|+2)\mathbf{P}\left(x^{|X|}\cap\right)$$

и запишем ее в таком же виде для остальных двух членов неравенства треугольника:

$$\Delta^{\cap}_{X+\{x,z\}} = \sum_{m=0}^{|X|-1} \Big\{ m\mathbf{P}\left(C_X^m\right) + \mathbf{P}(x^m) + \mathbf{P}(z^m) \Big\} +$$

$$+|X| \left(\mathbf{P}\left(C_X^{|X|}\right) - \mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right)\right) + |X|\mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right) +$$

$$+\mathbf{P}\left(x^{|X|}\right) + \mathbf{P}\left(z^{|X|}\right) - (|X|+2)\mathbf{P}\left(x^{|X|} \cap z^{|X|}\right),$$

$$\Delta^{\cap}_{X+\{y,z\}} = \sum_{m=0}^{|X|-1} \Big\{ m\mathbf{P}\left(C_X^m\right) + \mathbf{P}(y^m) + \mathbf{P}(z^m) \Big\} +$$

$$+|X| \left(\mathbf{P}\left(C_X^{|X|}\right) - \mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right)\right) + |X|\mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right) +$$

$$+\mathbf{P}\left(y^{|X|}\right) + \mathbf{P}\left(z^{|X|}\right) - (|X|+2)\mathbf{P}\left(y^{|X|} \cap z^{|X|}\right).$$

Заметим, что одно и то же неотрицательное, так как из $x^{|X|} \cup y^{|X|} \cup z^{|X|} \subseteq C_X^{|X|}$ следует $\mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right) \leqslant \mathbf{P}\left(C_X^{|X|}\right)$.

Слагаемое $\sum_{m=0}^{|X|-1} m\mathbf{P}(C_X^m) + |X| \left(\mathbf{P}\left(C_X^{|X|}\right) - \mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right)\right)$ присутствует во всех трех членах неравенства, так что в левой части неравенства два таких слагаемых, а в правой одно.

Кроме того, $\sum_{m=0}^{|X|-1} \Big\{ \mathbf{P}\left(x^m\right) + \mathbf{P}\left(z^m\right) + \mathbf{P}\left(y^m\right) + \mathbf{P}\left(z^m\right) \Big\} \geqslant \sum_{m=0}^{|X|-1} \Big\{ \mathbf{P}\left(x^m\right) + \mathbf{P}\left(y^m\right) \Big\}$, так как для каждого $m = 0, \ldots, |X|$ $\mathbf{P}\left(x^m\right) + \mathbf{P}\left(z^m\right) + \mathbf{P}\left(y^m\right) + \mathbf{P}\left(z^m\right) \geqslant \mathbf{P}\left(x^m\right) + \mathbf{P}\left(y^m\right)$. Поэтому если показать, что для оставшихся слагаемых выполнено: $|X|\mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup\ z^{|X|}\right) + \mathbf{P}\left(x^{|X|}\right) + \mathbf{P}(\left(z^{|X|}\right) - (|X|+2)\mathbf{P}\left(x^{|X|} \cap z^{|X|}\right)\ + |X|\mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right) + \mathbf{P}\left(y^{|X|}\right) + \mathbf{P}(\left(z^{|X|}\right) - (|X|+2)\mathbf{P}\left(y^{|X|} \cap z^{|X|}\right) \geqslant |X|\mathbf{P}\left(x^{|X|} \cup y^{|X|} \cup z^{|X|}\right) + \mathbf{P}\left(x^{|X|}\right) + \mathbf{P}(\left(y^{|X|}\right) - (|X|+2)\mathbf{P}\left(x^{|X|} \cap y^{|X|}\right)$, то неравенство треугольника для сет-расстояния по пересечению будет доказано. Преобразуем отдельные слагаемые в неравенствах, выражая их через вероятности террасок трех событий x, y и z. Запишем формулы в тех же сокращенных обозначениях, что и при доказательстве предыдущей леммы, где 1 соответствует событию, а 0 — его дополнению, так что, например, $\mathbf{P}_{100} = \mathbf{P}(x^m \cap (y^m)^c \cap (z^m)^c)$.

Правая часть неравенства в этих обозначениях записывается в виде: $(|X|+1)(\mathbf{P}_{100}+\mathbf{P}_{010}+\mathbf{P}_{101}+\mathbf{P}_{011})+|X|\mathbf{P}_{001}+(|X|+2)\mathbf{P}_{110}$. Первый член левой части: $(|X|+1)(\mathbf{P}_{100}+\mathbf{P}_{001}+\mathbf{P}_{110}+\mathbf{P}_{011})+|X|\mathbf{P}_{010}+(|X|+2)\mathbf{P}_{101}$. Второй член левой части: $(|X|+1)(\mathbf{P}_{010}+\mathbf{P}_{001}+\mathbf{P}_{110}+\mathbf{P}_{101})+|X|\mathbf{P}_{100}+(|X|+2)\mathbf{P}_{011}$. Вся левая часть неравенств имеет вид: $(2|X|+1)(\mathbf{P}_{100}+\mathbf{P}_{010}) + (2|X|+3)(\mathbf{P}_{101}+\mathbf{P}_{011}) + (2|X|+2)(\mathbf{P}_{110}+\mathbf{P}_{001})$. Таким образом, разность левой и правой частей — это неотрицательная величина: $(|X|+2)(\mathbf{P}_{001}+\mathbf{P}_{011}+\mathbf{P}_{011}) + |X|(\mathbf{P}_{100}+\mathbf{P}_{010}+\mathbf{P}_{110}) \geqslant 0$. Неравенство треугольника для сет-расстояния по пересечению доказано. □

6. Метрическое задание эвентологических распределений

Теорема (метрическое задание эвентологических распределений). *Любое эвентологическое распределение множества случайных событий $\mathfrak{X}$ определяется заданием всех вероятностей событий $\mathbf{P}(x)$, $x \in \mathfrak{X}$ и всех сет-расстояний Δ_X, $X \subseteq \mathfrak{X}$.*

Доказательство. Из (1) легко получить распределение вероятностей объединений:

$$u_X = \frac{1}{|X|}\left(\Delta_X + \sum_{x\in X}\mathbf{P}(x)\right), \quad X \subseteq \mathfrak{X},$$

т.е. одно из эвентологических распределений множества случайных событий $\mathfrak{X}$, которое в условиях теоремы (когда заданы все вероятности событий и все сет-расстояния) определено для всех $X \subseteq \mathfrak{X}$. □

Данная теорема, справедливая для любого эвентологического распределения, рискует остаться довольно тривиальным утверждением, если с ее помощью не удастся выделить специальные классы *метрических эвентологических распределений*, определяемые той или иной сет-метрикой. Разумеется, как только один из таких классов будет построен, станут известны и другие способы задания распределений из "вновь обнаруженного" класса, в том числе и классические — при помощи распределений вероятностей пересечений или объединений. Но можно с уверенностью заявить, что эвентологическое распределение будет иметь полное право называться *метрическим*, если оно "всего лишь" обнаружено при помощи сет-метрики.

Список литературы

[1] Воробьёв О.Ю. *Случайные события: неравенства Фреше и корреляции Фреше* / О.Ю.Воробьёв // Вестник КрасГУ: физико-математические науки. – Красноярск: КрасГУ, 2003. – Вып. 1. – С. 80–87

[2] Воробьёв О.Ю. *Случайно-множественные разложения двудольных случайных векторов* / О.Ю.Воробьёв, Е.Е.Голденок, Д.В.Семёнова // Вестник КрасГУ: физико-математические науки. – Красноярск: КрасГУ, 2003. Вып. 1. – С. 88–96

[3] Воробьёв О.Ю. *Теория случайных событий и её применения* / О.Ю.Воробьёв, Е.Е.Голденок, Т.В.Куприянова, Д.В.Семёнова, А.Ю.Фомин – Красноярск: ИВМ СО РАН, 2002.

[4] Воробьёв А.О. *Прямые и обратные задачи для моделей распространения пространственных рисков*: Автореф. дис. ... канд. физ.-мат. наук / А.О.Воробьёв. – Красноярск: ИВМ СО РАН, 1998. – 23 с.

[5] Голденок Е.Е. *Моделирование структур зависимостей и взаимодействий случайных событий в статистических системах*: Автореф. дис. ... канд. физ.-мат. наук / Е.Е.Голденок. – Красноярск: КГТЭИ, 2002. – 24 с.

[6] Куприянова Т.В. *Задача классификации подмножеств случайного множества и ее применение*: Автореф. дис. ... канд. физ.-мат. наук / Т.В.Куприянова. – Красноярск: КрасГУ, 2002. – 20 с.

[7] Семёнова Д.В. *Методы построения статистических зависимостей портфельных операций в рыночных системах*: Автореф. дис. ... канд. физ.-мат. наук / Д.В.Семёнова. – Красноярск: ИВМ СО РАН, 2002. – 24 с.

[8] Фомин А.Ю. *Сет-регрессионный анализ зависимостей случайных событий в статистических системах*: Автореф. дис. ... канд. физ.-мат. наук / А.Ю.Фомин – Красноярск: ИВМ СО РАН, 2002. – 21 с.

METRIC EVENTOLOGICAL DISTRIBUTIONS

O.Vorob'ov, E.Goldenok

A way to give eventological distributions with the help of the new notion of set-distance *defined for any set of random events is considered. An* eventological distribution *is any probability distribution that defines a set of random events $\mathfrak{X}$ chosen from an algebra $\mathcal{F}$ of some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The same set of random events can be equivalently defined by giving anyone of the following probability distributions of intersections or of unions of random events [3]: $p(X)$, p_X, p^X, $u(X)$, u_X, u^X, $X \subseteq \mathfrak{X}$, where, for example, $p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right)$, $u_X = \mathbf{P}\left(\bigcup_{x \in X} x\right)$, $X \subseteq \mathfrak{X}$, are probability distributions of intersections and unions respectively. Since $p_X = \prod_{x \in X} \mathbf{P}(x) - \mathsf{Kov}_X$, $X \subseteq \mathfrak{X}$, the distribution of* probability of intersections *p_X can be defined by giving probabilities of events $\mathbf{P}(x)$, $x \in \mathfrak{X}$, and a distribution of* arity covariances *Kov_X, $X \subseteq \mathfrak{X}$. In some problems, for example in N random events problem [5], an using another characteristics of a set of random events with properties of a distance between them instead of arity covariance, a measure of statistical interdependencies of random events, is more available. But a distance is usually defined only for pairs of elements, not for any set of elements. In the paper the new notion of* set-distance $\Delta_X = |X| \cdot u_X - \sum_{x \in X} \mathbf{P}(x)$ of a set of random events *$X \subseteq \mathfrak{X}$ is introduced. So defined set-distance Δ_X is a function of a set $X \subseteq \mathfrak{x}$ which generalizes a distance between two events $x, y \in \mathfrak{X}$ (in new terms an usual distance "turns" into a* set-distance of duplet of events *$X = \{x, y\}$), usually defined as a probability of their symmetric difference: $\Delta_{xy} = \mathbf{P}(x \Delta y)$. In the paper properties of the new notion are considered and the theorem about metric giving eventological distributions is proved.*

Theorem (metric giving eventological distributions). *Any eventological distribution of a set of random events $\mathfrak{X}$ is defined by giving their probabilities $\mathbf{P}(x)$, $x \in \mathfrak{X}$, and set-distances Δ_X of all subsets of random events $X \subseteq \mathfrak{X}$.*