УДК 519

ЭВЕНТОЛОГИЧЕСКАЯ РЕГРЕССИЯ СЛУЧАЙНЫХ МНОЖЕСТВ СОБЫТИЙВ ВИДЕ СЕТ-СРЕДНИХ

А. И. Иванова
Научный руководитель: к.ф.-м.н., доцент Д.В. Семенова
Институт математики и фундаментальной информатики
Сибирский федеральный университет

Регрессия — это зависимость среднего значения какой-либо величины от некоторой другой величины или от нескольких величин. Если дано совместное распределение двух случайных множеств событий K и L, значения которых содержаться в конечных множествах $\mathfrak{X}$ и $\mathfrak{Y}$ соответственно, то регрессией K на L называется любой оператор $\Psi(K)$, приближенно представляющий статистическую зависимость L от K. Для определения зависимости между двумя случайными множествами событий в [1, 2] было предложено использовать эвентологическую регрессию, которая устанавливает вид средней сет-функциональной зависимости между этими двумя случайными множествами событий.

Случайное множество событий под $\mathfrak{X}$ — это измеримое отображение вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ в конечное измеримое пространство $\left(2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}\right)$, т.е. $K:(\Omega, \mathcal{F}, \mathbf{P}) \to \left(2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}\right)$, где $\mathfrak{X} \subset \mathcal{F}$ — конечное множество событий, избранных из алгебры $\mathcal{F}$ этого пространства. Вероятностное распределение *случайного множества событий* K под конченым множеством событий $\mathfrak{X}$ можно представить несколькими эквивалентными распределениями вероятностей, порождённых этим множеством событий [1]:

- эвентологическое распределение (Э-распределение)I-го рода случайного множества событий K под конченым множеством событий $\mathfrak{X}$ – это набор из $2^{|\mathfrak{X}|}$ вероятностей вида $\{p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}$, где $p(X//\mathfrak{X}) = P(K = X) = P\left(ter(X//\mathfrak{X})\right)$; $|\mathfrak{X}|$– мощность множества $\mathfrak{X}$; под $ter(X//\mathfrak{X})$ понимаемтеррасное событие I-го рода [1] $ter(X//\mathfrak{X}) = \left(\bigcap_{x \in X} x\right) \bigcap \left(\bigcap_{x \in X^c} x^c\right)$, которое наступает тогда, когда наступают события $x \in X$ и не наступают события $x \in X^c = \mathfrak{X} - X$.
- Э-распределение II-го рода случайного множества событий K под конченым множеством событий $\mathfrak{X}$ – это набор из $2^{|\mathfrak{X}|}$ вероятностей вида $\left\{p_{X//\mathfrak{X}}, X \subseteq \mathfrak{X}\right\}$, где $p_{X//\mathfrak{X}} = P(X \subseteq K) = P\left(ter_{X//\mathfrak{X}}\right)$, где $ter_{X//\mathfrak{X}}$ понимаем террасное событие II-го рода [1]: $ter_{X//\mathfrak{X}} = \bigcap_{x \in X} x$.

Задача эвентологической регрессии. *Пусть задано совместное распределение двух случайных множеств событий K и L, значения которых содержаться в конечных множествах $\mathfrak{X}$ и $\mathfrak{Y}$ соответственно. Пусть $\rho: 2^{\mathfrak{Y}} \to 2^{\mathfrak{Y}}$ – некоторая метрика в $\mathfrak{Y}$. Необходимо найти сет-функцию $\varphi: 2^{\mathfrak{X}} \to 2^{\mathfrak{Y}}$, которая доставляет минимум функционалу $\mathbf{E}\rho(L, \varphi(K))$, где $\mathbf{E}$ – оператор математического ожидания.*

В [1] доказано, что если в качестве метрики $\rho: 2^{\mathsf{Y}} \to 2^{\mathsf{Y}}$ взять α-симметричную метрику $\rho_\alpha(X,Y) = \alpha\,|X \setminus Y| + (1-\alpha)\,|Y \setminus X|$, функция эвентологической регрессии имеет вид

$$Y = \varphi(X) = Q_\alpha\left(Y\big|_{ter(X//\mathsf{X})}\right),\ X \subseteq \mathsf{X},\ Y \subseteq \mathsf{Y}$$

—условного сет-квантиля порядка α, где $Q_\alpha\left(Y\big|_{ter(X//\mathsf{X})}\right) = \left\{y : P\left(Y\big|_{ter(X//\mathsf{X})}\right) \geq \alpha\right\}$.

Совместное распределение двух случайных множеств событий K и L, значения которых содержаться в конечных множествах X и Y соответственноможно представить в виде следующей таблицы размерности $2^{|\mathsf{X}|} \times 2^{|\mathsf{Y}|}$.

Таблица 1

Совместное распределение двух случайных множеств событий K под X и L под Y

X \ Y	$ter(\varnothing//\mathsf{Y})$	$\dots$	$ter(Y//\mathsf{Y})$	$\dots$	$ter(\mathsf{Y}//\mathsf{Y})$	$P(K=X)$
$ter(\varnothing//\mathsf{X})$	$P\left(ter(\varnothing//\mathsf{X}\cup\mathsf{Y})\right)$	$\dots$	$P\left(ter(Y//\mathsf{X}\cup\mathsf{Y})\right)$	$\dots$	$P\left(ter(\mathsf{Y}//\mathsf{X}\cup\mathsf{Y})\right)$	$P(ter(\varnothing//\mathsf{X}))$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$ter(X//\mathsf{X})$	$P\left(ter(X//\mathsf{X}\cup\mathsf{Y})\right)$	$\dots$	$P\left(ter(X\cap Y//\mathsf{X}\cup\mathsf{Y})\right)$	$\dots$	$P\left(ter(X\cap\mathsf{Y}//\mathsf{X}\cup\mathsf{Y})\right)$	$P(ter(X//\mathsf{X}))$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$ter(\mathsf{X}//\mathsf{X})$	$P\left(\mathsf{X}//\mathsf{X}\cup\mathsf{Y})\right)$		$P\left(ter(\mathsf{X}\cap Y//\mathsf{X}\cup\mathsf{Y})\right)$		$P\left(ter(\mathsf{X}\cap\mathsf{Y}//\mathsf{X}\cup\mathsf{Y})\right)$	$P(ter(\mathsf{X}//\mathsf{X}))$
$P(L=Y)$	$P(ter(\varnothing//\mathsf{Y}))$	$\dots$	$P(ter(Y//\mathsf{Y}))$	$\dots$	$P(ter(\mathsf{Y}//\mathsf{Y}))$	

Пользуясь табл. 1 можно рассчитать условные вероятности по формуле

$$P\left(Y\big|_{ter(X)}\right) = \frac{P\left(ter(X\cap Y//\mathsf{X}\cup\mathsf{Y})\right)}{P(ter(X//\mathsf{X}))}$$

и найти:

- условный сет-квантиль $Q_\alpha\left(Y\big|_{ter(X)}\right) = \left\{y : P\left(Y\big|_{ter(X)}\right) \geq \alpha\right\}$
- условную сет-медиану $Med\left(Y\big|_{ter(X)}\right) = \left\{y : P\left(Y\big|_{ter(X)}\right) \geq \frac{1}{2}\right\}$
- условную сет-моду $P\left(Y\big|_{ter(X)} = Mod\,(Y\big|_{ter(X)})\right) = \max\limits_{X\subseteq\mathsf{X}} P(Y\big|_{ter(X)} = X)$

Рассмотрим пример «Выпить-Закусить». Имеется статистика чеков заказов за 9 дней. Разобьем исходный ассортимент на два множества. Пусть $\mathsf{X} = \{x, y, z, l\}$ - множество «Закусить», а $\mathsf{Y} = \{a, b, c, d, e\}$ -множество «Выпить». Пусть случайное множество событий K задано под X. K можно трактовать как набор заказанных блюд случайным покупателе. Пусть случайное множество событий L задано под Y и трактуется как набор заказанных напитков случайным покупателе. Из имеющихся заказов была получена оценка совместного распределения K и L. Построим сет-регрессии K на L в виде сет-квантиля, сет-моды, сет-медианы. На рис. 1 представлена диаграмма Эйлера-Венна для множества событий $\mathsf{X} \cup \mathsf{Y}$. По вертикали — события-терраски $ter(Y//\mathsf{Y})$, $Y \subseteq \mathsf{Y}$; по горизонтали — события-терраски $ter(X//\mathsf{X})$, $X \subseteq \mathsf{X}$; красными квадратами показан полученный график сет-функции эвентологической регрессии

$Y = \varphi(X) = Q_\alpha\left(Y\big|_{ter(X)}\right),\ X \subseteq \mathsf{X},\quad Y \subseteq \mathsf{Y}$, при α=4/7.

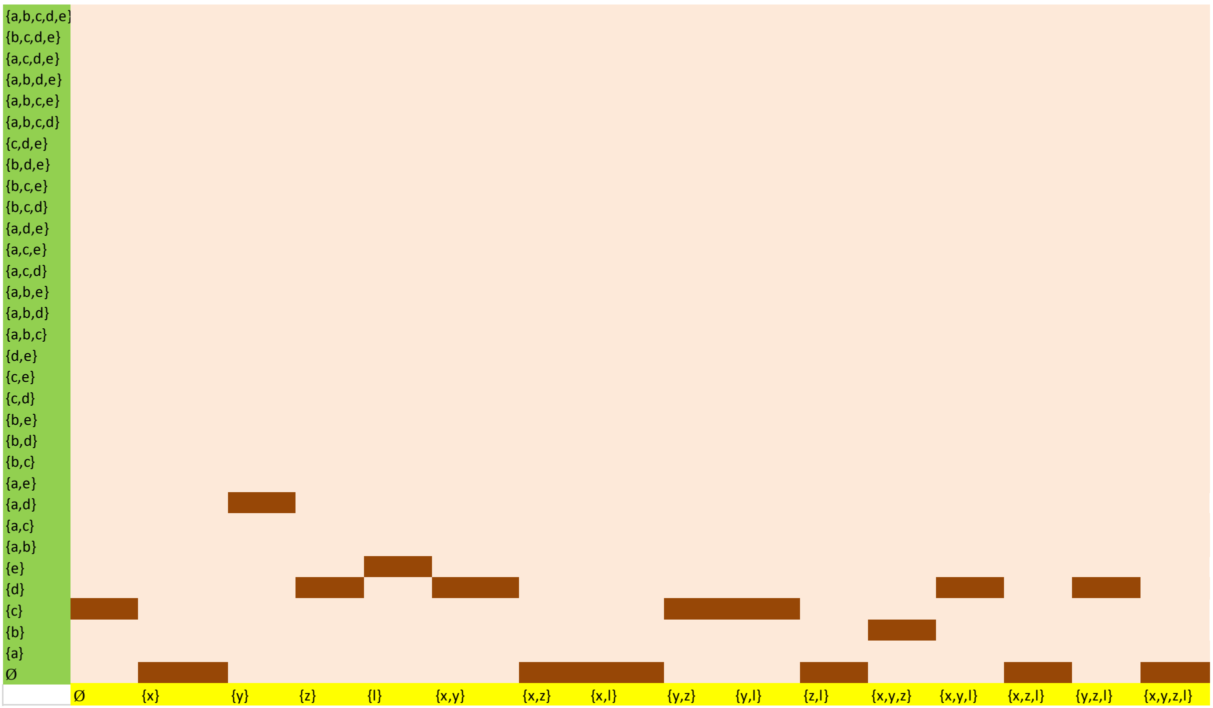

Рис. 1. Диаграмма Эйлера-Венна для множества событий $\mathsf{X} \cup \mathsf{Y}$. *По вертикали — события-терраски* $ter(Y\,//\,\mathsf{Y}) = \{u : u \in Y\}$, $Y \subseteq \mathsf{Y}$; *по горизонтали — события-терраски* $ter(X\,//\,\mathsf{X}) = \{x : x \in X\}$, $X \subseteq \mathsf{X}$; *красными квадратами показан полученный график сет-функции эвентологической регрессии* $Y = \varphi(X)$, $X \subseteq \mathsf{X}$, $Y \subseteq \mathsf{Y}$.

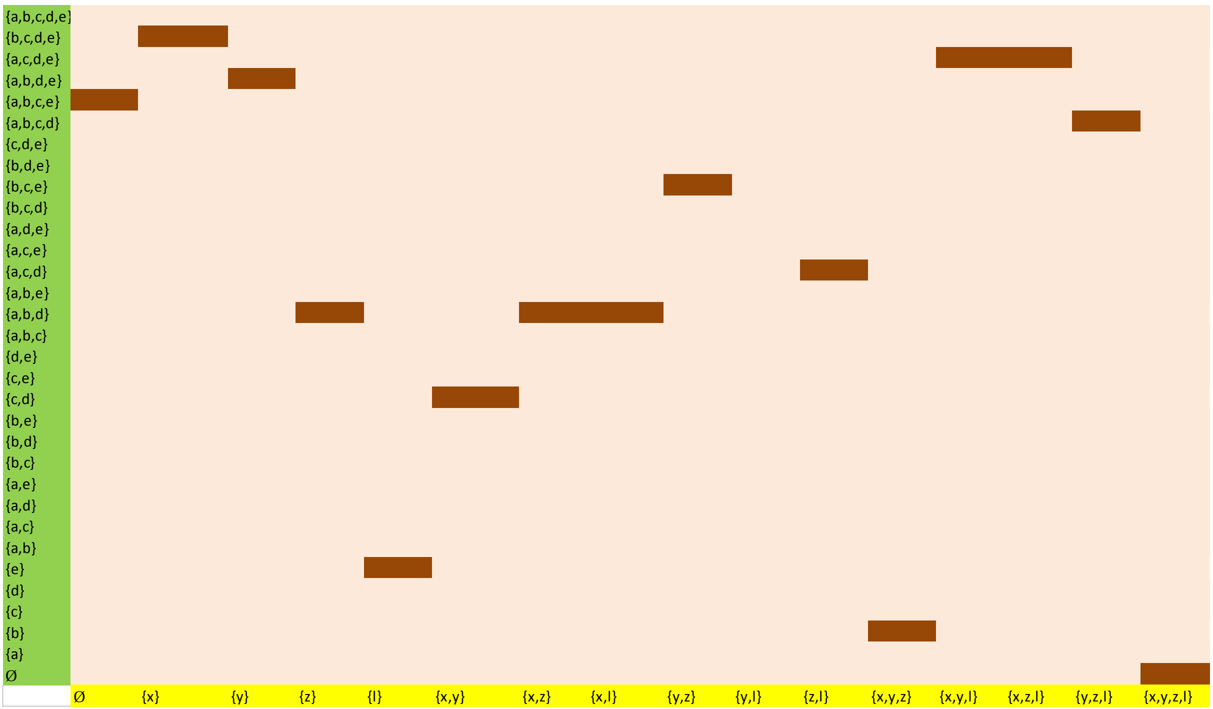

Рис. 2. Сет-медианадля примера «Выпить-закусить»

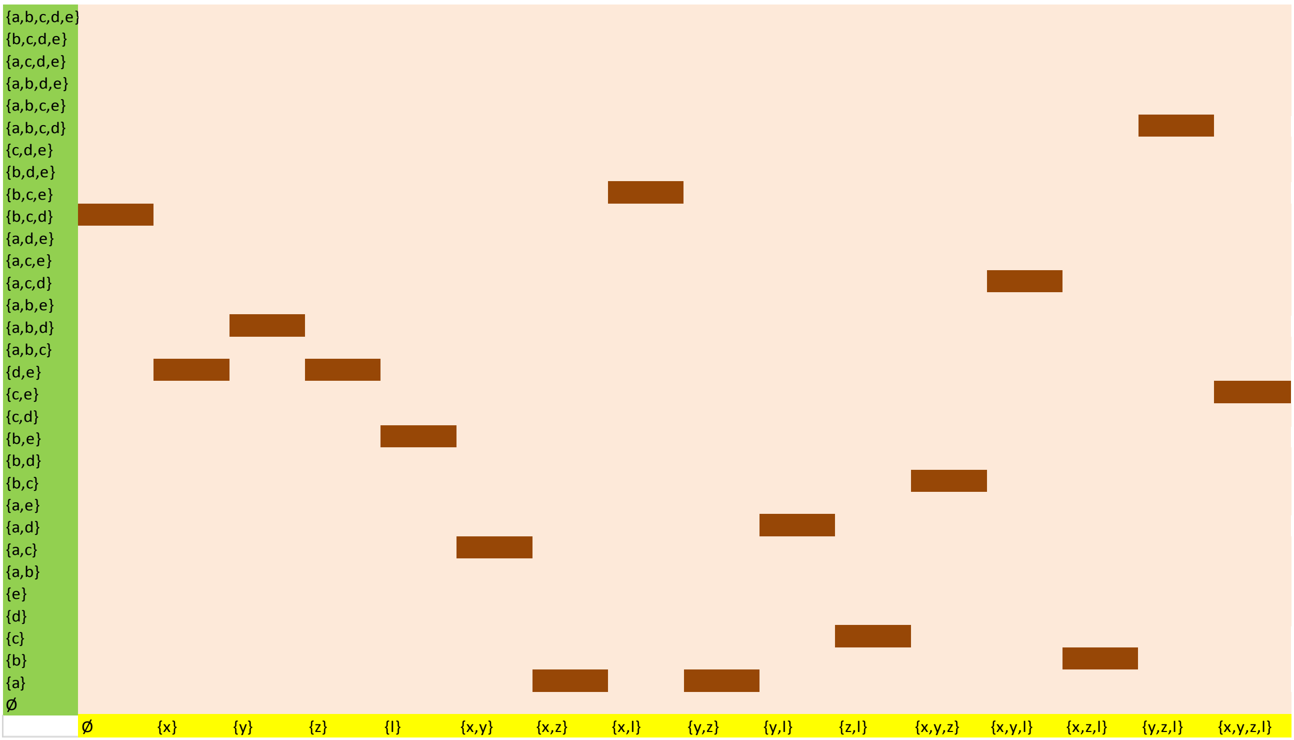

Рис.2. Сет-мода для примера «Выпить-закусить»

Эвентологическая регрессия может быть использована для решения задач моделирования, прогнозирования и контроля развития статистических систем природы и общества, события в которых, как правило, представлены неколичественной информацией [2, 3].