

Мультковариации и многоточечно зависимые распределения случайных множеств

А.О. Воробьёв

*Международное агентство социальных и маркетинговых исследований
125195, Москва, ул. Беломорская 6-28
e-mail: vorobov@mail.ru*

В работе рассматривается класс распределений случайных конечных абстрактных множеств (СКАМ), которые максимизируют энтропию при определенных ограничениях. Рассматривается связь этих распределений с мультковариациями случайного множества.

Под энтропией СКАМ K понимается следующее классическое выражение

$$\mathcal{S} = - \sum_{X \subseteq \mathfrak{X}} \mathbf{P}(K = X) \ln \mathbf{P}(K = X) = - \sum_{X \subseteq \mathfrak{X}} p(X) \ln p(X).$$

Известно, что абсолютный максимум энтропии достигается на равномерном распределении СКАМ:

$$p(X) = 1/2^{|\mathfrak{X}|}, \quad X \in 2^{\mathfrak{X}}.$$

Также известен факт, что максимум энтропии распределения СКАМ при фиксированных вероятностях покрытия случайного множества

$$p_x = \mathbf{P}(x \in K), \quad x \in \mathfrak{X}$$

достигается на классическом независимо-точечном распределении:

$$p(X) = \prod_{x \in X} p_x \prod_{x \in X^C} (1 - p_x), \quad X \in 2^{\mathfrak{X}}.$$

В статье решается общая задача поиска условного максимума энтропии СКАМ, при постоянных вероятностях покрытия множеств мощности меньшей или равной n , а также при ограничениях на область определения случайного множества.

МАКСИМИЗАЦИЯ ЭНТРОПИИ ПРИ ФИКСИРОВАННЫХ ВЕРОЯТНОСТЯХ ПОКРЫТИЯ

Зафиксируем вероятности покрытия СКАМ K для всех множеств мощности меньшей или равной n :

$$p_X = \mathbf{P}(X \subseteq K) = a_X, \quad |X| \leq n.$$

Найдем распределение СКАМ, которое максимизирует энтропию при таких ограничениях. Для этого запишем энтропию через вероятности покрытия случайного множества. Известно, что вероятности значений СКАМ выражаются через вероятности покрытия по формуле обращения Мебиуса:

$$p(X) = \sum_{Y \supseteq X} (-1)^{|Y \setminus X|} p_Y.$$

Отсюда получаем следующее выражение для энтропии:

$$\mathcal{S} = - \sum_{X \subseteq \mathfrak{X}} \left(\sum_{Y \supseteq X} (-1)^{|Y \setminus X|} p_Y \right) \ln \left(\sum_{Y \supseteq X} (-1)^{|Y \setminus X|} p_Y \right).$$

Для нахождения максимума приравняем к нулю частные производные \mathcal{S} по p_Z , $|Z| > n$.

$$\mathcal{S}'_{p_Z} = - \sum_{X \subseteq Z} (-1)^{|Z \setminus X|} \ln \left(\sum_{Y \supseteq X} (-1)^{|Y \setminus X|} p_Y \right) = 0, \quad |Z| > n.$$

Немного преобразуя выражение, получим следующую систему уравнений для вероятностей значения случайного множества:

$$\prod_{X \subseteq Z} p(X)^{(-1)^{|Z \setminus X|}} = \tau(Z) = 1, \quad |Z| > n,$$

где $\tau(Z)$ - мультиковариация случайного множества K .

Таким образом, доказана следующая теорема.

Теорема. *В классе распределений СКАМ, удовлетворяющих следующим условиям:*

$$p_X = \mathbf{P}(X \subseteq K) = a_X, \quad |X| \leq n,$$

максимум энтропии достигается на распределении, для которого:

$$\tau(Z) = 1, \quad |Z| > n.$$

Такие распределения будем называть n -точечно зависимые или много-точечно зависимые распределения.

Из известного факта, что любое распределение СКМ можно представить в виде произведения мультиковариаций:

$$\mathbf{P}(X) = \prod_{Y \subseteq X} \tau(Y),$$

сразу следует следующее следствие.

Следствие. n -точечно зависимое распределение СКМ представляется в виде:

$$\mathbf{P}(K = X) = p(X) = \prod_{Y \subseteq X, |Y| \leq n} \tau(Y).$$

Например, для случая $n = 2$ получаем следующее выражение для таких распределений:

$$\begin{aligned} p(X) &= \tau(\emptyset) \prod_{x \in X} \tau(\{x\}) \prod_{\{x,y\} \subset X} \tau(\{x,y\}) = \\ &= p(\emptyset) \prod_{x \in X} \frac{p(\{x\})}{p(\emptyset)} \prod_{\{x,y\} \subset X} \frac{p(\emptyset)p(\{x,y\})}{p(\{x\})p(\{y\})}. \end{aligned}$$

МАКСИМИЗАЦИЯ ЭНТРОПИИ ПРИ ФИКСАЦИИ ОБЛАСТИ ОПРЕДЕЛЕНИЯ СКМ

Допустим, что случайное множество K может принимать только ограниченное количество значений, т.е.

$$p(X) = 0, \quad X \notin \mathcal{A} \subseteq 2^{\mathfrak{X}}.$$

При таких условиях максимум энтропии, как легко понять, достигается на следующем распределении:

$$p(X) = 1/|\mathcal{A}|, \quad X \in \mathcal{A},$$

$$p(X) = 0, \quad X \notin \mathcal{A}.$$

Как и в предыдущей секции дополнительно зафиксируем вероятности покрытия для всех множеств мощности меньшей или равной n :

$$p_X = \mathbf{P}(X \subseteq K) = a_X, \quad |X| \leq n.$$

Аналогичными рассуждениями можно доказать следующую теорему:

Теорема. *В классе распределений СКМ, удовлетворяющих следующим условиям:*

$$p_X = \mathbf{P}(X \subseteq K) = a_X, \quad |X| \leq n$$

и

$$p(X) = 0, \quad X \notin \mathcal{A},$$

максимум энтропии достигается на распределении вида

$$p(X) = \prod_{Y \subseteq X, |Y| \leq n} \alpha(Y), \quad X \in \mathcal{A},$$

где $\alpha(Y)$ находятся из условий на вероятности покрытия и том факте, что

$$\sum_{X \in \mathcal{A}} p(X) = 1.$$

ЛИТЕРАТУРА

- [1] ВОРОБЬЕВ О.Ю. (1984) *Среднемерное моделирование*. Москва: Наука, 133 с.
- [2] ВОРОБЬЕВ О.Ю. (1993) *Сет-суммирование*. Новосибирск: Наука, 137 с.
- [3] ВОРОБЬЕВ О.Ю. (2002 будет опубликовано) *Теория случайных событий и её применение*. Красноярск: ИВМ СО РАН, 340 с.
- [4] ВОРОБЬЕВ А.О. (1998) *Прямые и обратные задачи для моделей распространения пространственных рисков*: Автореферат диссертации на соискание ученой степени канд. физ.-мат. наук. Красноярск: ИВМ СО РАН, 24 с.
- [5] VOROB'OV A.O., AND O.YU. VOROB'OV (1996) Inverse problems for generalized Richardson's model of spread *Computational Fluid Dynamics'96*. John Wiley & Sons Ltd, 104–110.