ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2020, том 493, с. 104–107

ПРОЦЕССЫ УПРАВЛЕНИЯ

УДК 51-7

АСИМПТОТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ ОЦЕНОК МАКСИМАЛЬНОЙ ЭНТРОПИИ

© 2020 г. Академик РАН Ю. С. Попков[1,2,*]

Поступило 06.04.2020 г.
После доработки 11.05.2020 г.
Принято к публикации 23.05.2020 г.

Сформулирована задача энтропийного оценивания функций плотности распределения вероятностей с учетом реальных данных (MEE). Получены глобальные условия существования неявной зависимости множителей Лагранжа от коллекции данных. Доказана асимптотическая эффективность MEE.

Ключевые слова: энтропийное оценивание, функции плотности, множители Лагранжа, вращение векторного поля, асимптотическая эффективность

DOI: 10.31857/S2686954320040165

Оценивание характеристик моделей с использованием данных является важной научной проблемой. В прикладных задачах, в которых она возникала, использовались модели с неизвестными параметрами, значения которых требовалось оценивать [1–4].

Решение этих задач осуществляется традиционными методами математической статистики – методом максимального правдоподобия и его производными, методом моментов, байесовыми методами и многочисленными их модификациями [5–7].

Задача оценивания характеристик моделей по реальным данным получила свое новое развитие в связи с появлением новых методов машинного обучения – рандомизированного машинного обучения (РМО) [8]. Они основаны на моделях со случайными параметрами, и оценивать необходимо функции плотности распределения вероятностей (ПРВ) этих параметров. Алгоритм оценивания (алгоритм РМО) формулируется в терминах функционального энтропийно-линейного программирования [9].

Появившись как задача оценивания функций ПРВ в процедурах РМО, она приобрела более общий контекст: метод максимизации энтропийных функционалов для построения оценок непрерывных функций с использованием реальных данных (Maximum Entropy Estimation – MEE).

Настоящее сообщение посвящено постановке общей MEE-задачи, исследованию ее решений и асимптотических свойств.

ПОСТАНОВКА MEE-ЗАДАЧИ

Рассмотрим скалярную непрерывную ограниченную функцию $\varphi(x,\theta) = \hat{y}$, которая характеризует модель с входом x и выходом y. Здесь x, $\hat{y}$ – скалярные переменные и $\theta \in \Theta \subset R^n$ – параметры. В результате r измерений имеем $\mathbf{x}^{(r)} = \{x[1],\ldots,x[r]\}$ и $\mathbf{y}^{(r)} = \{y[1],\ldots,y[r]\}$. Последние содержат случайные ошибки, различные для моментов измерений:

$$\xi^{(r)} = \{\xi[1],\ldots,\xi[r]\}.$$

Таким образом, после r измерений получаем следующие уравнения модели и наблюдений:

$$\begin{gathered}\mathbf{y}^{(r)} = \Gamma(\mathbf{x}^{(r)},\theta),\\ \Gamma(\mathbf{x}^{(r)},\theta) = \{\varphi(x[1],\theta),\ldots,\varphi(x[r],\theta)\}\\ \mathbf{v}^{(r)} = \hat{\mathbf{y}}^{(r)} + \xi^{(r)}.\end{gathered} \tag{1}$$

Предположения:

1) множество $\Theta = [\theta^-,\theta^+]$;

2) функция ПРВ $P(\theta)$ – непрерывно-дифференцируема на носителе Θ;

3) значения функции φ ограничены:

$$\varphi^- \le \varphi(x[t],\theta) \le \varphi^+, \quad t = 1,2,\ldots,r, \tag{2}$$

[1] *Федеральный исследовательский центр “Информатика и управление” Российской академии наук, Москва, Россия*
[2] *Институт проблем управления им. В.А. Трапезникова Российской академии наук, Москва, Россия*
**E-mail: popkov.yuri@gmail.com*

для всех $\theta \in \Theta$ и $x[t] \in R^1$;

4) случайный шум $\xi^{(r)} \in \Xi \subset R^r$ и множество

$$\Xi = \bigotimes_{t=1}^{r} \Xi_t, \quad \Xi_t = [\xi^-[t], \xi^+[t]]; \tag{3}$$

5) функция ПРВ $Q(\xi)$ – непрерывно-дифференцируема на носителе Ξ и имеет мультипликативную структуру:

$$Q(\xi) = \prod_{t=1}^{r} Q_t(\xi[t]). \tag{4}$$

Задача оценивания формулируется следующим образом: МЕЕ функций ПРВ $P^*(\theta), Q^*(\xi)$ максимизируют функционал обобщенной информационной энтропии

$$\begin{gathered}\mathcal{H}[P(\theta), Q(\xi)] = -\int_{\Theta} P(\theta) \ln P(\theta)\, d\theta - \\ -\sum_{t=1}^{r} \int_{\Xi_t} Q_t(\xi[t]) \ln Q_t(\xi[t]) \Rightarrow \max\end{gathered} \tag{5}$$

при ограничениях:

1) нормировка функций ПРВ:

$$\int_{\Theta} P(\theta)\, d\theta = 1; \quad \int_{\Xi_t} Q_t(\xi[t]) d\xi[t] = 1, \quad t = 1, 2, \ldots, r; \tag{6}$$

2) эмпирические балансы

$$\begin{gathered}\Phi[P(\theta), Q(\xi)] = \mathbf{y}^{(r)}, \\ \Phi[P(\theta), Q(\xi)] = \\ = \{\Phi_1[P(\theta), Q(\xi)], \ldots, \Phi_r[P(\theta), Q(\xi)]\}, \\ \Phi_t[P(\theta), Q(\xi)] = \\ = \int_{\Theta} \varphi(x[t], \theta) P(\theta)\, d\theta + \int_{\Xi_t} Q_t(\xi[t]) \xi[t]\, d\xi[t], \\ t = 1, 2, \ldots, r.\end{gathered} \tag{7}$$

Задача (5)–(7) относится к классу ляпуновских [10], которые характеризуются тем, что функционал и ограничения интегрального типа.

ЭНТРОПИЙНО-ОПТИМАЛЬНЫЕ ОЦЕНКИ ФУНКЦИЙ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Условия оптимальности в задачах оптимизации ляпуновского типа формулируются в терминах множителей Лагранжа. При этом используются производные Гато [10] входящих в задачу функционалов. Применяя указанную технику, получим энтропийно-оптимальные оценки функций ПРВ:

$$\begin{gathered}P^*(\theta|\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) = \frac{\exp\left(-\sum_{j=1}^{r} \lambda_j(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) \varphi(x[j], \theta)\right)}{\mathcal{P}(\lambda(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})}, \\ Q_t^*(\xi[t]|\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) = \frac{\exp(-\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})\xi[t])}{\mathcal{Q}_t(\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})}, \\ t = 1, 2, \ldots, r.\end{gathered} \tag{8}$$

В этих равенствах

$$\begin{gathered}\mathcal{P}(\lambda(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) = \\ = \int_{\Theta} \exp\left(-\sum_{j=1}^{r} (\lambda_j\, \mathbf{y}^{(r)}, \mathbf{x}^{(r)}) \varphi(x[j], \theta)\right) d\theta, \\ \mathcal{Q}_t(\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) = \int_{\Xi_t} \exp(-\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})\xi[t]) d\xi[t], \\ t = 1, 2, \ldots, r.\end{gathered} \tag{9}$$

Из равенств (8), (9) видно, что энтропийно-оптимальные ПРВ параметризованы множителями Лагранжа $\lambda_1, \ldots, \lambda_r$, которые определяются решением уравнений эмпирических балансов:

$$\begin{gathered}\frac{\mathcal{G}(\lambda(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}))}{\mathcal{P}(\lambda(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}))} + \frac{\mathcal{E}_t(\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}))}{\mathcal{Q}_t(\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}))} = y[t], \\ t = 1, 2, \ldots, r,\end{gathered} \tag{10}$$

где

$$\begin{gathered}\mathcal{G}(\lambda(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})) = \int_{\Theta} \varphi(x[t], \theta) \times \\ \times \exp\left(-\sum_{j=1}^{r} \lambda_j(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) \varphi(x[j], \theta)\right) d\theta, \\ \mathcal{E}_t(\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})) = \\ = \int_{\Xi_t} \exp(-\lambda_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})\xi[t]) d\xi[t], \\ t = 1, 2, \ldots, r.\end{gathered} \tag{11}$$

Решение этих уравнений – функция $\lambda^*(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})$, она зависит от выборки $(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})$, по которой строятся МЕЕ функций ПРВ.

СУЩЕСТВОВАНИЕ НЕЯВНОЙ ФУНКЦИИ $\lambda(\tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)})$

Преобразуем уравнения эмпирических балансов к виду

$$\mathbf{W}(\lambda \mid \tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)}) = \mathbf{0}, \tag{12}$$

где компоненты

$$W_t(\lambda \mid \tilde{y}[t], \mathbf{x}^{(r)}) = \int_\Theta (\varphi(x[t],\theta) - \tilde{y}[t]) \times$$
$$\times \exp\left(-\sum_{j=1}^{r} \lambda_j(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}) \varphi(x^{(r)}, \theta)\right) d\theta = 0, \tag{13}$$
$$t = 1, 2, \ldots, r,$$

и

$$\tilde{y}[t] = y[t] - \overline{\xi}[t], \quad y = 1, 2, \ldots, r. \tag{14}$$

Обозначим якобиан системы (12)

$$J_\lambda(\lambda \mid \tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)}) = \left[\frac{\partial W_t}{\partial \lambda_i},\ (t,i) = 1, 2, \ldots, r.\right] \tag{15}$$

Т е о р е м а 1. *Пусть*:

а) *функция* $\varphi(\mathbf{x}^{(r)}, \theta)$ *непрерывна по совокупности переменных*;

б) *для любых* $(\mathbf{x}^{(r)}, \tilde{\mathbf{y}}^{(r)}) \in R^r \times R^r$ *выполняются следующие условия*:

$$\det J_\lambda(\lambda \mid \tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)})) \neq 0, \tag{16}$$

$$\lim_{\|\lambda\| \to \infty} \mathbf{W}(\lambda \mid \tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)}) = \pm\infty. \tag{17}$$

Тогда существует единственная неявная функция $\lambda(\tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)})$, *определенная на* $R^r \times R^r$.

Т е о р е м а 2. *Пусть выполнены условия теоремы* 1. *Тогда функция* $\lambda(\tilde{\mathbf{y}}^{(r)}, \mathbf{x}^{(r)})$ – *аналитическая по совокупности переменных.*

АСИМПТОТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ МЕЕ

МЕЕ функций ПРВ представим через экспоненциальные множителями Лагранжа $z = \exp(-\lambda)$. Тогда равенства (8) примут следующий вид:

$$P^*(\theta, \mathbf{z}(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})) = \frac{\prod_{j=1}^{r} [z_j(\mathbf{y}^{(r)}, \mathbf{x}^r)]^{\varphi(x[j],\theta)}}{\int_\Theta \prod_{j=1}^{r} [z_j(\mathbf{y}^{(r)}, \mathbf{x}^r)]^{\varphi(x[j],\theta)} d\theta},$$

$$Q_t^*(\xi[t], z_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})) = \frac{[z_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})]^{\xi[t]}}{\int_{\Xi_t} [z_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})]^{\xi[t]} d\xi[t]}, \tag{18}$$
$$t = 1, 2, \ldots, r.$$

О п р е д е л е н и е 1. Будем называть оценки $P^*(\theta, \mathbf{z}^*)$ и $Q_t^*((\xi[t], z_t^*)$ а с и м п т о т и ч е с к и э ф ф е к т и в н ы м и , если

$$\lim_{r \to \infty} P^*(\theta, \mathbf{z}(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})) = P^*(\theta, \mathbf{z}^*),$$
$$\lim_{r \to \infty} Q_t^*(\xi[t], z_t(\mathbf{y}^{(r)}, \mathbf{x}^{(r)})) = Q_t^*(\xi[t], z_t^*), \tag{19}$$
$$t = 1, 2, \ldots, r,$$

где

$$\mathbf{z}^* = \lim_{r \to \infty} \mathbf{z}(\mathbf{y}^{(r)}, \mathbf{x}^{(r)}). \tag{20}$$

Рассмотрим уравнения эмпирических балансов (12), перейдя в них к экспоненциальным множителям Лагранжа:

$$\Phi_t(\mathbf{z}, \tilde{\mathbf{y}}^{(r)}, \mathbf{x}^r)) = \int_\Theta \prod_{j=1}^{r} [z_j(\tilde{\mathbf{y}}^{(r)}, \mathbf{x}^r)]^{\varphi(x[j],\theta)} \times$$
$$\times (\varphi(x[t], \theta) - \tilde{y}[t]) d\theta = 0,$$
$$t = 1, 2, \ldots, r. \tag{21}$$

Т е о р е м а 3 . *Пусть выполнены условия теорем* 1 *и* 2. *Тогда существует константа* $\alpha > 1$, *такая*, *что*

$$0 \le \left\|\frac{\partial \mathbf{z}}{\partial \tilde{\mathbf{y}}^{(r)}}\right\| \le \alpha \left[r \max_{t,i} \mid \frac{\partial \Phi_t}{\partial z_i}\right]^{-1} \left\|\frac{\partial \Phi}{\partial \tilde{\mathbf{y}}^{(r)}}\right\|, \tag{22}$$
$$0 \le \left\|\frac{\partial \mathbf{z}}{\partial \mathbf{x}^{(r)}}\right\| \le \alpha \left[r \max_{t,i} \mid \frac{\partial \Phi_t}{\partial z_i}\right]^{-1} \left\|\frac{\partial \Phi}{\partial \mathbf{x}^{(r)}}\right\|.$$

Л е м м а 1. *Пусть*

$$\left\|\frac{\partial \Phi}{\partial \mathbf{y}^{(r)}}\right\| \le \rho < \infty, \quad \left\|\frac{\partial \Phi}{\partial \mathbf{x}^{(r)}}\right\| \le \omega < \infty. \tag{23}$$

Тогда

$$\lim_{r \to \infty} \left\|\frac{\partial z}{\partial \mathbf{y}^{(r)}}\right\| = \lim_{r \to \infty} \left\|\frac{\partial z}{\partial \mathbf{x}^{(r)}}\right\| = 0, \tag{24}$$

и МЕЕ *функций* ПРВ (19) *асимптотически эффективна*.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 20–07–00470).

СПИСОК ЛИТЕРАТУРЫ

1. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Исследование зависимостей. М.: Финансы и статистика, 1985.
2. *Goldberger A.S.* A Course in Econometrics. Harvard University Press, 1991. 437 p.
3. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. Springer, 2001.
4. *Воронцов К.В.* Математические методы обучения по прецендентам. Курс лекций МФТИ, 2006.
5. *Боровков А.А.* Математическая статистика. М.: Наука, 1984.

6. *Лагутин М.Б.* Наглядная математическая статистика. М.: БИНОМ, Лаборатория знаний, 2013.

7. *Rousses G.* A Course of the Mathematical Statistics. Academic Press Inc. 2015. 600 p.

8. *Popkov Yu.S., Dubnov Yu.A., Popkov A.Yu.* Randomized Machine Learning: Statement, Solution, Applications // 2016 IEEE 8th International Conference on Intelligent Systems (IS). 2016. https://doi.org/10.1109/IS.2016.7737456.

9. *Popkov A.Yu., Popkov Yu.S.* New Methods of Entropy-Robust Estimation for Randomized Models under Limited Data // Entropy. 2014. №16. P. 675–698. https://doi.org/10.3390/e16020675

10. *Иоффе А.Д., Тихомиров В.М.* Теория экстремальных задач. М.: Наука, 1974. 481 с.

ASYMPTOTIC EFFICIENCY OF MAXIMUM ENTROPY ESTIMATES

Academician of the RAS Yu. S. Popkov[a,b]

[a]*Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russian Federation*

[b]*Trapeznikov Institute of Control Problems of the Russian Academy of Sciences, Moscow, Russian Federation*

The problem of entropy estimation of the probability density functions is formulated. The global existence conditions of implicit dependence of the Lagrange multipliers from data collection are obtained. The asymptotic efficiency of maximum entropy estimate is proved.

Keywords: entropy estimation, density functions, Lagrange multipliers, vector field rotation, asimptotic efficiency