Секция «Теория вероятностей и математическая статистика»

Анализ асимптотического поведения статистических оценок дивергенции Кульбака - Лейблера для дискретных распределений

Научный руководитель – Булинский Александр Вадимович

Мирмоминов Руслан Мэргыязович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия

E-mail: mirmominov98@gmail.com

Пусть P и Q – дискретные вероятностные меры, определенные на некотором конечном множестве $S = \{l_1, \ldots, l_K : K > 1, K \in \mathbb{N}\}$. Тогда *дивергенция Кульбака - Лейблера* между ними задается формулой

$$D_{KL} = D_{KL}(P, Q) = \sum_{k=1}^{K} p_k \ln p_k - \sum_{k=1}^{K} p_k \ln q_k,$$

где $p_k = P(l_k) > 0,\ q_k = Q(l_k) > 0,\ k = 1, \ldots, K$. Интерес представляют статистические оценки величины D_{KL}, которые строятся для неизвестных мер P и Q по независимым выборкам объема n из соответствующих распределений. Состоятельность и асимптотическая нормальность трех оценок D_{KL} рассматривалась в [1]. Для третьей из них в [2] установлено, что ее смещение сходится к нулю экспоненциально быстро. Нами найдена скорость сходимости к нулю дисперсии третьей из упомянутых оценок. Доказано, что эта дисперсия ограничена сверху величиной $\frac{C_1}{n} + \frac{C_2}{n^2} + \frac{C_3}{n^6} + C_4 e^{-\lambda n}$, где положительные C_i $(i = 1, \ldots)$ и λ явно указаны (определяются распределениями P и Q). Таким образом, оценка дисперсии имеет вид $O(\frac{1}{n})$ при $n \to \infty$. Такой порядок убывания дисперсии является оптимальным согласно неравенству Рао - Крамера.

Источники и литература

1) Zhang, Z. (2014). Nonparametric Estimation of Küllback-Leibler Divergence.
2) Zhang, Z. (2012). Entropy estimation in Turing's perspective. Neural Computation. 24(5), 1368–1389.