Секция «Теория вероятностей и математическая статистика»

Некоторые свойства статистических дивергенций

Хлюстов Дмитрий Кириллович
Студент (специалист)
Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия
E-mail: hlustov.d@gmail.com

Введенное Шенноном в [6] понятие энтропии в настоящее время находит широкое применение в теории вероятностей, теории информации, в задачах оптимального управления и в исследованиях нейронных сетей. Для случайной величины X, принимающей значения в пространстве $\mathbb{R}^n$ и имеющей плотность $p(x)$ по мере μ, ее дифференциальная энтропия по определению равна $H(X) = -\int_{\mathbb{R}^n} p(x)\log p(x)d\mu$, где $0\log 0 := 0$. Часто данную величину интерпретируют как меру неопределенности, содержащейся в случайной величине X.
В качестве обобщения понятия энтропии рассматривают так называемые f-дивергенции (см., например, [5]), используемые для характеризации близости вероятностных распределений. Для мер P, Q на пространстве $\mathbb{R}^n$ и выпуклой функции $f : \mathbb{R} \to \mathbb{R}$ полагают

$$D_f(P||Q) = \begin{cases} \int_{\mathbb{R}^n} f(\frac{dP}{dQ})dQ, \text{ если } P \ll Q, \\ +\infty, \text{ иначе.} \end{cases}$$

В зависимости от выбора f величина D_f может обладать различными свойствами, делающими ее удобнее для тех или иных приложений. В частности, при $f(t) = t\log t, t \geq 0$, имеем дивергенцию Кульбака-Лейблера $D_{KL}(P||Q)$, применяемую в статистике и теории информации (см. [2]). Важной для приложений задачей оказывается нахождение соотношений, связывающих различные дивергенции (см., например, [3]).
В данной работе развивается общий подход к получению соотношений разных типов, основанный на понятии совместной области, предложенном в [1]. Показывается, что теорема, сформулированная в статье [1] для двух дивергенций, может быть обобщена на произвольное конечное или некоторое бесконечное семейство соответствующих функционалов. На основании этих утверждений получены некоторые неравенства, связывающие несколько дивергенций. Например, доказано, что

$$D_{KL}(P||Q) \geq \frac{1}{2} D_{TV}(P||Q) \log\left(1 + \frac{\chi^2(P||Q)}{D_{TV}(P||Q)}\right),$$

где $D_{TV}(P||Q), \chi^2(P||Q)$ - соответственно расстояние полной вариации и дивергенция χ^2, возникающие при подстановке в определение f-дивергенции $f(t) = |t-1|$ и $f(t) = (t-1)^2$. Их свойства рассматриваются в [4].
Кроме того, основные результаты нашей работы позволяют установить новые соотношения, связывающие различные дивергенции и обобщающие результаты статьи [3].

Источники и литература

1) Harremoës P., Vajda I. On pairs of f-divergences and their joint range // IEEE Transactions on Information Theory. 57(6). 2011. p. 3230-3235.
2) Kullback S., Leibler R. A. On information and sufficiency // The Annals of Mathematical Statistics. 22(1). 1951. p. 79-86.

3) Nishiyama T., Sason I. On relations between the relative entropy and χ^{2}-divergence, generalizations and applications // Entropy. 22(5). 2020. p. 563-599.
4) Polyanskiy Y., Yihong W., Lecture notes on information theory. Lecture Notes for ECE563 (UIUC)., 2014
5) Rényi A. On measures of entropy and information // In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. 4. 1961. p. 547-562.
6) Shannon C. E. A mathematical theory of communication // The Bell System Technical Journal. 27(3). 1948. p. 379-423.