

М. П. К р и в е н к о (Москва, ИПИ РАН). **Свойства ядерной оценки многомерной плотности.**

Рассматривается ядерная оценка плотности распределения размерности $d > 1$ по наблюдаемым значениям $X_1, \dots, X_n$, имеющая вид: $\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K^H(x - X_i)$, где $K^H(x) = |H|^{-1/2} K(H^{-1/2}x)$, $\int K(x) dx = 1$, H — симметричная положительно определенная сглаживающая матрица.

Ядро $K(x)$ может формироваться из одномерного $k(x)$ так, что возникают два типа многомерных ядер: $K_p(x) = \prod_{i=1}^d k(x_i)$ — ядро мультипликативного типа, $K_S(x) \propto k\{(x^T x)^{1/2}\}$ — ядро сферически симметричного типа. Для каждого типа в качестве $k(x)$ далее рассматриваются равномерная, треугольная, нормальная плотности распределения, а также ядро Епанечникова, что находит отражение в обозначениях, например, $K_{SE}(x)$ — сферически симметричное ядро на основе одномерного ядра Епанечникова.

Воспользовавшись понятием канонического параметра сглаживания, становится возможным сравнивать асимптотическую эффективность применения некоторых ядер K_1 и K_2 при одной и той же плотности f с помощью величины

$$k(K_1, K_2) = \frac{C_d(K_2)}{C_d(K_1)},$$

где $C_d(K) = (R^4(K) \mu_2^{2d}(K))^{1/(d+4)}$, $R(K) = \int K^2(z) dz$, $\mu_2(K) = \int z_1^2 K(z) dz$.

Для сравнительного анализа различных ядер K_p и K_S находятся выражения для R и μ_2 , при этом уточняются и дополняются результаты [1] и [2], а затем строятся зависимости $k(K_p, K_S)$ от d , анализ которых позволяет обосновать следующее: сферически симметричные ядра имеют преимущества перед мультипликативными; при увеличении размерности преимущества растут, оставаясь ограниченными определенными уровнями.

Для каждого типа ядра (мультипликативного и сферически симметричного) можно найти оптимальное ядро — это ядро Епанечникова. Отдавая предпочтение сферически симметричному ядру, можно внутри соответствующего класса провести анализ потерь при использовании вместо оптимального K_{SE} другого ядра. Осуществляется это с помощью исследования зависимости $k(K_S, K_{SE})$ от d для различных ядер, что позволяет получить новые результаты: увеличение размерности может нивелировать потери от использования ядра, отличного от оптимального; для нормального ядра с увеличением размерности эффективность резко падает, правда, не ниже уровня приблизительно в 50%. Вывод представлений для характеристик ядер приведен в приложении 5.4 [3].

В завершение находятся выражения для наилучших значений параметров сглаживания для перечисленных ядер и нормальной плотности распределения (метод оценивания параметров сглаживания в духе "rule-of-thumb"), что крайне важно, в частности, по тому, что можно найти начальные приближения для параметра сглаживания при применении метода подстановки ("plug-in").

СПИСОК ЛИТЕРАТУРЫ

1. *Silverman B. W.* Density estimation for statistics and data analysis. – Chapman&Hall, 1986, 176 p., ISBN 0412246201.
2. *Wand M. P., Jones M. C.* Kernel smoothing. – Chapman&Hall, 1994, 224 p., ISBN 0412552701.
3. *Кривенко М. П.* Прикладные методы оценивания распределения многомерных данных малой выборки. М.: ИПИ РАН, 2011, 146 с., ISBN 9785919930013.