ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 24 МАТЕМАТИКИ Выпуск 3
2017

М. С. Тихов, М. Н. Ивкина (Нижний Новгород, ННГУ). **Метод Рида и Менча в зависимости «доза-эффект».**

Рассматривается зависимость «доза-эффект», которая описывается следующим образом [1]. Пусть случайная величина X имеет распределение, которая неизвестно, с функцией распределения $F(x)$ и плотностью $f(x) > 0$. Пусть имеется еще одна случайная величина U (доза), плотность распределения которой равна $g(x) > 0$ и которую мы также считаем неизвестной. Наблюдению доступны пары $\mathcal{U}^{(n)} = \{(u_i, w_i),\ 1 \leqslant i \leqslant n\}$, (повторная выборка), где w_i есть значение случайной величины W_i — индикатора события $(X_i < U_i)$ (эффект). Требуется по наблюдениям $\mathcal{U}^{(n)}$ оценить неизвестную функцию распределения $F(x)$. В случае, когда величины X и U — независимы, $\mathbf{E}(W|U = x) = F(x)$, т. е. является регрессией, и для оценки $F(x)$ можно использовать ядерную оценку $\widehat{F}_n(x) = \dfrac{S_{2n}(x)}{S_{1n}(x)}$, где

$$S_{1n}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{u_i - x}{h}\right), \quad S_{2n}(x) = \frac{1}{nh}\sum_{i=1}^{n} w_i K\left(\frac{u_i - x}{h}\right).$$

Здесь $K(x)$ есть ядерная функция (ядро) — четная плотность распределения с носителем распределения на отрезке $[-1, 1]$, а h есть параметр сглаживания, $h \to 0$ при $n \to \infty$, но $nh \to \infty$. В работе [2] показано, что разность $\xi_n = n^{2/5}\,(\widehat{F}_n(x) - F(x))$ при $n \to \infty$ и некоторых условиях регулярности состоятельна и асимптотически нормальна $N(a(x), b^2(x))$, где $a(x) = f'(x) + 2f(x)g'(x)/g(x)$, $b^2(x) = F(x)(1 - F(x)) \parallel K \parallel^2 / g(x)$.

Кроме оценки $\widehat{F}_n(x)$ в практическом исследовании для оценки используются и другие оценки, например, kNN-оценки [3]: здесь предельная дисперсия величины ξ_n есть $\widetilde{b}^2(x) = F(x)(1 - F(x)) \parallel K \parallel^2$ и не зависит от плотности $g(x)$. В [4] был предложен также другой метод оценивания $F(x)$ (метод Рида и Менча). «Этот метод исходит из того естественного допущения, что если некоторый тест-объект дал положительный ответ при какой-либо дозе, то он даст такой же ответ и при более высоких дозах; и наоборот, если некоторый тест-объект дал отрицательный ответ при определенной дозе, то он даст также отрицательный ответ и при всех меньших дозах»: на с. 237 [5] приведен пример применения метода. Однако в некоторых работах (см., например, [6]) эффективность метода Рида и Менча подвергалась сомнению.

В данном сообщении мы рассмотрим аналог метода Рида и Менча. Будем строить оценку для $F(x)$ по формуле: $\widetilde{F}_n(x) = \dfrac{S_{3n}(x)}{S_{4n}(x)}$, где

$$S_{3n}(x) = \frac{1}{nh}\sum_{i=1}^{n} w_i H\left(\frac{x - u_i}{h}\right), \quad S_{4n}(x) = S_{3n}(x) + \frac{1}{nh}\sum_{i=1}^{n}(1 - w_i)\left(1 - H\left(\frac{x - u_i}{h}\right)\right),$$

и исследуем ее асимптотическое поведение при $n \to \infty$. Функцию $H(x)$ выберем равной $\int_{-1}^{x} K(t)\,dt$, $-1 \leqslant x \leqslant 1$, и $H(x) = 0$ вне интервала $[-1, 1]$. Показано, что при $n \to \infty$,

$$\mathbf{E}(S_{3n}(x)) = F(x)g(x) + \frac{1}{2}(1-\mu_2)h\,(Fg)'(x) + \frac{1}{6}h^2(Fg)''(x) + o(h^2),$$
$$\mathbf{E}(S_{4n}(x)) = g(x) + \frac{1}{2}(1-\mu_2)g'(x)h + \frac{1}{6}g''(x)h^2 + o(h^2), \qquad (1)$$

где $\mu_2 = \int_{-1}^{1} t^2 K(t)\,dt$. Для дисперсий рассматриваемых статистик имеем:

$$\mathbf{D}(S_{3n}(x)) = \frac{1}{nh}F(x)g(x)\,\|H\|^2\,(1+o(1)),\ \mathbf{D}(S_{4n}(x)) = \frac{1}{nh}g(x)\,\|H\|^2\,(1+o(1)).$$

Кроме того, при некоторых условиях регулярности и при $h = n^{-1/3}$ имеет место соотношение $n^{1/3}(\widetilde{F}_n - F(x)) \xrightarrow[n\to\infty]{d} N((1-\alpha_2)f(x)/2, F(x)(1-F(x))\,\|H\|^2/g(x))$. Таким образом, оценка $\widetilde{F}_n(x)$ состоятельна и сходится со скорость $n^{-1/3}$, однако она не является $\sqrt[3]{n}$-состоятельной оценкой. Повысить скорость сходимости $\widetilde{F}_n$ можно с использованием двухшаговой процедуры, предложенной в работе [7]. При этой процедуре дисперсия новой оценки остается такой же, коэффициент при h в разложениях (1) аннулируется, а скорость сходимости оценок повышается до $n^{2/5}$. Заметим, что последняя процедура, примененная к оценке $\widehat{F}_n(x)$ изменяет ее скорость сходимости с $n^{2/5}$ до $n^{4/9}$.

СПИСОК ЛИТЕРАТУРЫ

1. *Криштопенко С. В., Тихов М. С., Попова Е. Б.* Доза-эффект. М.: Медицина, 2008, 285 с.
2. *Tikhov M. S.* Statistical Estimation on the Basis of Interval-Censored Data. — J. Math. Sci., 2004, v. 119, № 3, p. 321–335.
3. *Tikhov M., Ivkin M.* Multivariate k-Nearest Neighbors Distribution Function Estimates in Dose-effect Relationship. — Recent Adv. Math. Meth. in Appl. Sci.: Proc. Conf., 2014, p. 325–329.
4. *Reed L. J., Muench H.* A simple method of estimating fifty per cent endpoints. — Amer. J. Hygiene, 1938, v. 27, № 3, p. 493–497.
5. *Урбах В. Ю.* Статистический анализ в биологических и медицинских исследованиях. М.: Медицина, 1975, 297 с.
6. *Гуревич К. Г.* Оценка параметров кривой «доза-эффект» методом сплайн-интерполяции. — Вестник Московского ун-та. Сер. 2. Химия, 2000, т. 41, № 1, с. 69–70.
7. *Tikhov M. S., Dolgih I. S.* Asymptotically unbiased estimates of a distribution function in dose-effect relationships. — J. Mat. Sci., 2015, v. 205, № 1, p. 113–120.