М. С. Тихов (Нижний Новгород, ННГУ). **Устранение погрешности измерений в зависимости «доза–эффект».**

В данном сообщении рассматривается проблема «доза–эффект» (см. [1]), где по повторной выборке $\mathcal{U}^{(n)} = \{(U_i, W_i),\ 1 \leqslant i \leqslant n\}$ объема n (здесь $\{U_i\}$ — случайные или неслучайные величины, $W_i = I\{X_i < U_i\}$ есть индикатор события $\{X_i < U_i\}$) рассматривается задача оценивания неизвестной функции распределения (ф. р.) $F(x) = \mathbf{P}\{X_1 < x\}$ (см. [2]) или задача проверки гипотез согласия и однородности [3] в ситуации, когда вместо величин U_i наблюдаются величины $Y_i = U_i + \varepsilon_i$, а $W_i = I\{X_i < U_i\}$, ε_i есть неизвестная ошибка измерения, т. е. мы имеем выборку $\mathcal{Y}^{(n)} = \{(Y_i, W_i)\}$, по которой требуется оценить неизвестную ф. р. $F(x)$ величины X. Для оценивания $F(x)$ в [4] рассматривалась «наивная» оценка N-W вида $F_n(x) = S_{2n}(x)/S_{1n}(x)$, где $S_{jn}(x) = n^{-1}\sum_{i=1}^n W_i^{j-1} K_h(x-y_i)$, $j = 1, 2$, $K_h(x) = h^{-1}K(x/h)$, $K(x)$ — ядро (четная финитная плотность распределения); $h = h(n) \to 0$, но $nh \to \infty$ при $n \to \infty$. Предельное при $n \to \infty$ распределение статистики $F_n(x)$ указано в [5], где отмечено, что $F_n(x)$ сходится не к «истинной» функции распределения, а к функции $m(x) = \int F(x-u)\alpha(u)\,du$, $\alpha(u)$ — плотность распределения (п. р.) с. в. ε_i. Каким образом можно «снять» погрешность измерения в оценке функции распределения $F(x)$?

Один из методов устранения погрешности ε состоит в том, что сначала строится ядерная оценка по выборке $\mathcal{Y}^{(n)}$, а затем «устраняется» погрешность измерения ε в полученной оценке (см. [6, 7]). Но этот метод плохо подходит для задачи непараметрической оценки регрессии, поэтому мы предлагаем другой подход (см. [8]), который состоит в том, чтобы сначала «очистить» саму выборку, а затем по очищенной выборке оценить неизвестную функцию распределения (мы рассмотрим указанную проблему применительно к задаче оценивания).

Пусть для начала ε_i имеют нормальное распределение $\mathcal{N}(0, \sigma_0^2)$ с известной дисперсией σ_0^2 и с плотностью распределения $\varphi(x)$. Пусть $u(x) = \mathbf{E}(U|Y = x)$, где $(X_i, U_i, \varepsilon_i)$ независимы и одинаково распределены с (X, U, ε). Пусть $g(u)$ есть п.р. с.в. U, а $q(y)$ — п. р. с. в. Y. Тогда при наших предположениях $q(y) = \int g(u)\varphi(y-u)\,du$. Отсюда $u(x) = \rho(x)/q(x)$, где $\rho(x) = \int ug(u)\varphi(y-u)\,du$ и $q'(x) = -xq(x) + \rho(x)$. Если с.в. $U \in N(a, \sigma^2)$, то $u(x) = \sigma_0^2(\ln q(x))' + x = ((\sigma^2 - \sigma_0^2)x + a\sigma_0^2)/\sigma^2$ и в качестве оценки функции $u(x)$ можно использовать $\widetilde{u}_n(x) = ((s^2-\sigma_0^2)x + a\sigma_0^2)/s^2$, $\overline{y} = n^{-1}\sum_{j=1}^n y_j$, $s^2 = n^{-1}\sum_{j=1}^n (y_j - \overline{y})^2$.

Если с. в. U и ε имеют логистические распределения, то приведенную оценку $\widetilde{u}_n(x)$ можно использовать и в этом случае, поскольку функция $u(x)$ здесь близка к линейной для достаточно широких интервалов $\mathcal{I}$. Удовлетворительные оценки получаются и в случае, когда U имеет распределение экстремальных значений, а также распределение свертки двух логистических распределений.

В случае произвольного распределения и когда $\varepsilon \in \mathcal{N}(0, \sigma_0^2)$ мы предлагаем использовать оценку для функции $u(x)$ вида $u_n(x) = \sigma_0^2\rho_n(x)/q_n(x) + x$, где $q_n(x) = S_0(x)$, $\rho_n(x) = (q_n(x+h/2) - q_n(x-h/2))/h$.

Теорема. *Если вариация $\mu = V(K)$ ограниченного ядра K конечна, п. р. $q(x)$ имеет ограниченную непрерывную производную и $nh^2 \to \infty$ при $n \to \infty$, то $u_n(x) \xrightarrow{p} u(x)$ при $n \to \infty$ равномерно на конечных интервалах $\mathcal{I} = [a, b]$.*

Действительно, если $\lambda(t)$ — характеристическая функция ядра $K(x)$, то $\int |\lambda(t)|\,dt < \infty$ и если $\psi_n(x) = n^{-1}\sum_{j=1}^n \exp\{ity_i\}$, то при $n \to \infty$,

$$\sup_{x\in\mathcal{I}} |q_n(x) - q(x))| \leqslant (2\pi)^{-1}\int |\lambda(hu)|\,|\psi_n(u) - \mathbf{E}(\psi_n(u))|\,du \leqslant (n^{1/2}h)^{-1}\int |\lambda(t)|\,dt \to 0.$$

Кроме того, $V_{2n} = \sup_{x\in\mathcal{I}} |\rho_n(x) - \mathbf{E}(\rho_n(x))| \leqslant \sup_{x\in\mathcal{I}} |G_n(x) - G(x))|\,\mu/2h$, где $G(x)$ — ф. р., $G_n(x)$ — э. ф. р. с. в. Y, построенная по выборке $y_1, y_2, \ldots, y_n$, а $\sup_{x\in\mathcal{I}} |\mathbf{E}(\rho_n(x)) - q'(x))| \leqslant (\nu^2+2)L_3h^2$, $\nu^2 = \int x^2K(x)\,dx$, $|q'''(x)| \leqslant L_3$.

Отсюда следует состоятельность оценки $u_n(x)$.

Компьютерное моделирование показало, что данный способ хорошо «снимает» погрешность измерения для распределений, «близких» к нормальному.

СПИСОК ЛИТЕРАТУРЫ

1. *Тихов М. С.*, *Криштопенко С. В.*, *Попова Е. Б.* Доза–эффект. М.: Медицина, 2008, 228 с.
2. *Tikhov M. S.* Statistical estimation based on interval censored data. — In: Param. and semiparam. models with appl. to rel., surv. analysis, and qual. of life. Boston, MA: Birkhuser Boston, 2004, p. 211–218.
3. *Тихов М. С.*, *Криштопенко Д. С.* Асимптотические распределения суммируемых квадратичных уклонений оценок функции распределения в зависимости доза–эффект. — Обозрение прикл. и промышл. матем., 2009, т. 16 , в. 5, с. 772–786.
4. *Тихов М. С.*, *Ярощук М. В.* Статистическое оценивание распределений по интервально цензурированным выборкам в схеме непрямых наблюдений. — Нелинейный мир, 2007, т. 5, № 1–2, с. 4–8.
5. *Tikhov M. S.*, *Krishtopenko D. S.*, *Yarochuk M. V.* Asymptotic normality of the integrated square error at the fixed plan of experiment for indirect observations. — Comp. Model. and New Technologies, Riga, 2007, v. 11, № 1, p. 46–56.
6. *Fan J.* Asymptotic normality for deconvolution kernel density estimators. — Sankhyā Ser. A, 1991, v. 53, p. 97–110.
7. *Es B.*, *Gugushvili S.*, *Spreij P.* Deconvolution for an atomic distribution. — Electronic Journal of Statist., 2008, v. 2, p. 265–297.
8. *Надарая Э.* О непараметрических оценках плотности вероятности и регрессии. — Теория вероятн. и е° примен., 1965, т. 10, в. 1, с. 199–203.