В. М. Кочеганов, М. С. Тихов (Нижний Новгород, ННГУ)**. Оценивание эффективных доз в зависимости доза–эффект.**

Пусть $\{(X_i, U_i),\ 1 \leqslant i \leqslant n\}$ — потенциальная повторная выборка из распределения $F(x)G(x)$, $F(x) = \mathbf{P}\{X_i < x\}$, $G(x) = \mathbf{P}\{U_i < x\}$, вместо которой наблюдается выборка $\mathcal{U}^{(n)} = \{(U_i, W_i),\ 1 \leqslant i \leqslant n\}$, где $W_i = I\{X_i < U_i\}$ есть индикатор события $\{X_i < U_i\}$. Эта модель интерпретируется как зависимость доза–эффект: U_i рассматриваются как вводимые дозы, а W_i — как эффект от воздействия дозы U_i. Данная модель может использоваться и в иной интерпретации: так, в [2] представленная модель наблюдений применялась для оценки возраста менархе, где U_i есть возраст i-го обследуемого, а X_i — «возраст» интересующего нас события. Пусть $F(x) = \int_{-\infty}^{x} f(t)\,dt$, причем $f(x) > 0$.

Одной из основных задач зависимости доза–эффект является оценка среднеэффективных доз $\mathbf{E}D_{100\alpha} = F^{-1}(\alpha) = x_\alpha$, $0 < \alpha < 1$, при построении которой нам понадобится NW-оценка функции распределения $F(x)$:

$$F_{nh}(x) = \sum_{i=1}^{n} W_i K\biggl(\frac{x - U_i}{h}\biggr) \Bigg/ \sum_{i=1}^{n} K\biggl(\frac{x - U_i}{h}\biggr),$$

где $K(x) \geqslant 0$ — четная плотность распределения с носителем на $[-1, 1]$, а h есть ширина окна просмотра данных — неслучайная последовательность $h = h(n)$, сходящаяся к нулю при $n \to \infty$. Пусть $\|K\|^2 = \int_{-1}^{1} K^2(x)\,dx < \infty$.

В качестве оценки для x_α предложим следующие статистики:

а) «естественную» оценку $\widehat{x}_{1,\alpha}$ как решение уравнения $F_{nh_1}(\widehat{x}_{1,\alpha}) = \alpha$;

б) DNP-оценку (см. [3])

$$\widehat{x}_{2,\alpha} = n^{-1} \sum_{i=1}^{n} H((\alpha - F_{nh_1}(x))/h_2), \quad \text{где} \quad H(x) = \int_{-\infty}^{x} K(t)\,dt;$$

в) оценки

$$\widehat{x}_{3,\alpha} = \biggl[2 \sum_{i=1}^{n} i\, H((\alpha - F_{nh_1}(i/n))/h_2)\biggr] \Bigg/ \biggl[n \sum_{i=1}^{n} H((\alpha - F_{nh_1}(x))/h_2)\biggr]$$

$$\widehat{x}_{4,\alpha} = \biggl[\sum_{i=1}^{n} i\, K((\alpha - F_{nh_1}(i/n))/h_2)\biggr] \Bigg/ \biggl[n \sum_{i=1}^{n} K((\alpha - F_{nh_1}(x))/h_2)\biggr],$$

где $h_1 = cn^{-1/5}$, а $h_2/h_1 \to 0$, $nh_1^2 h_2 \to \infty$ при $n \to \infty$.

Все эти оценки (при некоторых условиях регулярности) являются состоятельными и асимптотически нормальными оценками: оценки $\widehat{x}_{1,\alpha}$, $\widehat{x}_{2,\alpha}$, $\widehat{x}_{3,\alpha}$ — с нормировкой $\sqrt{nh_1}$ и предельной дисперсией $\alpha(1-\alpha)\|K\|^2/f^2(x_\alpha)$, а оценка $\widehat{x}_{4,\alpha}$ тоже состоятельна, имеет нормировку $\sqrt{nh_1^3}$ и предельную дисперсию $\alpha(1-\alpha)\|K'\|^2/f^2(x_\alpha)$. Поэтому для значений α, близких к нулю или к 1, дисперсия перечисленных оценок может быть весьма большой, т. е. на краях распределения $F(x)$ оценки эффективных доз могут сильно отличаться от «истинных» значений доз. Для снижения влияния неизвестной плотности $f(x)$ мы предлагаем следующую процедуру оценивания: 1) построим оценку $F_{nh_1}(x)$ по выборке $\mathcal{U}^{(n)}$ при заданном h_1; 2) возьмем $\widetilde{h}_1 = (F_{nh_1}^{-1}(\alpha + k/(2n)) - F_{nh}^{-1}(\alpha - k/(2n)))^2$, $k = cn^{4/5}$ и построим оценку $F_{n\widetilde{h}_1}(x)$; 3) используя оценку $F_{n\widetilde{h}_1}(x)$, построим оценки $\widetilde{x}_{2,\alpha}$ или $\widetilde{x}_{3,\alpha}$.

Компьютерное моделирование показало, что статистики $\widetilde{x}_{2,\alpha}$ и $\widetilde{x}_{3,\alpha}$ точнее оценивают эффективные дозы $\mathbf{E}D_{100\alpha}$ на краях распределения, чем оценки $\widehat{x}_{2,\alpha}$ и $\widehat{x}_{3,\alpha}$.

СПИСОК ЛИТЕРАТУРЫ

1. *Криштопенко С. В.*, *Тихов М. С.*, *Попова Е. Б.* Доза–эффект. М.: Медицина, 2008, 288 с.
2. *Finney D. J.* Probit Analysis. Cambridge: University Press, 1980, 333 p.
3. *Dette H.*, *Neumeyer N.*, *Pilz K. F.* A note on nonparametric estimation of the effective dose in quantal bioassay. — Journal of the American Statist. Assoc., 2005, v. 100, № 470, p. 503–510.