М. В. Я р о щ у к (Нижний Новгород, ННГУ). **Имитационное моделирование зависимости доза–эффект и статистический анализ оценок функции эффективности.**

В работе, представленой данным сообщением, для описания зависимости доза–эффект будем рассматривать следующую математическую модель [1]. Пусть $\mathbf{X}^{(n)} = (X_1, X_2, \ldots, X_n)$ — независимые и одинаково распределенные случайные величины (н.о.р.с.в.) с неизвестной функцией распределения $F(x)$ и плотностью распределения $f(x) > 0$; $\mathbf{U}^{(n)} = (U_1, U_2, \ldots, U_n)$ — н. о. р. с. в., независимые от $\mathbf{X}$ с неизвестным распределением $G(x)$ и плотностью $g(x) > 0$. Наблюдается последовательность одинаково распределенных пар $\mathbf{Y}^{(n)} = \{(U_i, W_i),\ i = 1, 2, \ldots, n\}$, где $W_i = I\{U_i > X_i\}$ — индикатор события $\{U_i > X_i\}$. Рассматривается задача оценивания функции эффективности (ф.р.) $F(x)$ по выборке $\mathbf{Y}^{(n)}$. Здесь $\mathbf{X}$ интерпретируется как минимальный уровень дозы, с которой начинается реакция организма (эффект), а U — введенная в организм случайная доза.

В качестве оценок функции эффективности рассмотрим ядерные оценки Надарая–Ватсона (NW) $F_n(x) = S_{2n}(x)/S_{1n}(x)$, полагая $F_n(x) = 0$, если $S_{1n}(x) = 0$, где $S_{1n}(x) = (1/n)\sum_{i=1}^n K_h(U_i - x)$, $S_{2n}(x) = (1/n)\sum_{i=1}^n W_i K_h(U_i - x)$, $K_h(x) = (1/h)K(x/h)$. Здесь $K(x)$ — ядро, т. е. неотрицательная нормированная ограниченная четная функция с компактным носителем, $h = cn^{-1/5}$ — ширина окна просмотра данных. При некоторых условиях регулярности [1] оценка $F_n(x)$ асимптотически нормальна: $n^{2/5}(F_n(x) - F(x)) \xrightarrow{d} N(c^2 a_1(x), \sigma_1^2(x))$ при $n \to \infty$ с асимптотическим смещением $a_1(x) = (f'(x)g(x) + 2g'(x)f(x))\nu^2/g(x)$ и дисперсией $\sigma_1^2(x) = F(x)(1 - F(x))\|K\|^2/g(x) = \sigma_2^2(x)/g(x)$, где $\nu^2 = \int_{-\infty}^{\infty} x^2 K(x)\,dx$, $\|K\|^2 = \int_{-\infty}^{\infty} K^2(x)\,dx$.

Когда величины U_i являются неслучайными (фиксированные планы испытаний) с упорядоченными значениями и постоянным шагом деления, в качестве оценок можно использовать также $S_{2n}(x)$ [2]. Показано, что как $F_n(x)$, так и $S_{2n}(x)$ являются состоятельными и асимптотически нормальными $N(1/2)f'(x)\nu^2$, $\sigma_2^2(x))$.

Мы изучали также kNN-оценки, которые строятся, как и оценки Надарая–Ватсона, но в предположении, что ширина окна просмотра данных h случайна и ее значение выбирается таким образом, чтобы в интервал $(x - h, x + h)$ попадало k выборочных значений случайной величины U. Установлено, что дисперсия этих оценок равна σ_2^2, т. е. уже не зависит от значения плотности $g(x)$, однако смещение оценок увеличивается. К ним применялась двухшаговая процедура, устраняющая асимптотическое смещение, в результате получены состоятельные асимтотические нормальные оценки.

Важным моментом ядерного оценивания является выбор ширины окна h. Этот параметр определяет величину окрестности точки, в которой производится оценивание. При выборе достаточно больших значений h происходит «чрезмерное сглаживание» оценки, приводящее к смещению, а при малых — из оценки исключаются важные данные. Для выбора оптимального значения h мы использовали процедуру кросс-проверки (cross-validation) и бутстрап (bootstrap). Доказана состоятельность и асимптотическая нормальность оценки ширины окна, полученной этим методом.

Для исследования теоретических результатов, носящих асимптотический характер, на выборках конечного объема было проведено имитационное моделирование зависимости доза–эффект. На приведенных ниже графиках в качестве модельного распределения используется нормальное распределение $N(17,5; 2^2)$. Чтобы продемонстрировать асимптотическую нормальность полученных оценок, на всех рисунках изображена интегральная кривая нормального распределения. Наличие отрицательного смещения в правой части графика и положительного в левой у оценки NW на рис. 1 объясняется присутствием в выражении для $a_1(x)$ производной плотности, монотонность которой и влияет на знак смещения. Полученная с помощью двухшаговой процедуры асимптотически несмещенная оценка для ядра Епанечникова $K_0(x) = (3/4)(1 - x^2)I\{|x| \leqslant 1\}$ (рис. 2) достаточно хорошо оценивает распределение на участках, где много наблюдений над с. в. U, но плохо на краях, где наблюдения редки. Смещение и дисперсия оценок NW и kNN-оценок зависят от ширины окна. Смещение также возрастает по $f'(x)$, но при выполнении условий состоятельности $h \to 0$, $n \to \infty$ смещение стремится к нулю и $nh \to \infty, n \to \infty$ — дисперсия стремится к нулю. Таким образом, смещение, связанное с производными, уменьшается при увеличении объема выборки, а в пределе исчезает. Было также замечено, что применение ядер повышенного порядка, например, квартического, приводит к небольшому уменьшению смещения. На рис. 3 приводятся результаты численного моделирования kNN-оценки и двухшаговой kNN-оценки для квартического ядра $K_1(x) = (15/16)(1 - x^2)I\{|x| \leqslant 1\}$ при $n = 50$, $k = 24$.

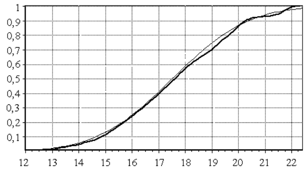

Рис. 1

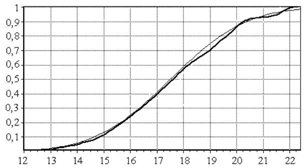

Рис. 2

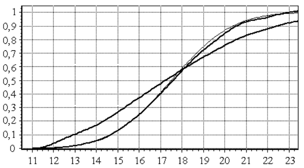

Рис. 3

Было произведено сравнение непараметрических оценок функции эффективности с оценками, построенными в помощью probit-анализа, и эти оценки мало отличаются, если они получены на основании нормально распределенных модельных данных. Более того, в ряде случаев доверительный интервал для медианы получается уже.

При фиксированном плане эксперимента для постоянного шага деления мы рассматривали оценки NW, а для непостоянного — оценки Пристли–Чао (PC) $F_{PC}(x) = \sum_{i=1}^{n} W_i(u_i - u_{i-1})K_h(x - u_i) = S_{2n}$ и нормированные оценки S_{2n}/S_{1n}, где $S_{1n} = \sum_{i=1}^{n}(u_i - u_{i-1})K_h(x - u_i)$. Установлено, что эти оценки являются состоятельными и асимптотически нормальными, но обладают асимптотическим смещением, которое мы устраняли с помощью двухшаговой процедуры. Однако у этой процедуры выявился свой недостаток: для значений на правом хвосте распределения иногда получались значения оценки, большие 1. Поэтому для больших x вместо значений W_i мы брали значения $(1 - W_i)$, и для корректности процедуры использовалась оценка $1 - S_{2n}/S_{1n}$ вместо S_{2n}/S_{1n}. Строилась комбинированная оценка, а затем асимптотически несмещенная оценка.

СПИСОК ЛИТЕРАТУРЫ

1. *Криштопенко С. В.*, *Тихов М. С.* Токсикометрия эффективных доз. Нижний Новгород: изд-во ННГУ, 1997, 156 с.
2. *Ярощук М. В.* Об оценивании распределений в зависимости доза–эффект. — Обозрение прикл. и промышл. матем., 2007, т. 14, в. 1, с. 178–180.