УДК 519.2

PC-ОЦЕНКИ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ В МОДЕЛИ ДОЗА-ЭФФЕКТ ПО СЛУЧАЙНЫМ ПЛАНАМ ЭКСПЕРИМЕНТА

М.С. Тихов

Нижегородский госуниверситет им. Н.И. Лобачевского

tikhovm@mail.ru

Поступила в редакцию 02.12.2011

Рассмотрены непараметрические PC-оценки функции распределения по случайным планам эксперимента в модели доза-эффект. Показано, что PC-оценки являются состоятельными и асимптотически. Предложена и исследована оценка предельной дисперсии.

Ключевые слова: доза-эффект-модель, непараметрические ядерные оценки, случайный план эксперимента.

Введение

В данной работе рассматривается проблема построения оценки функции распределений в модели доза-эффект, которая описывается следующим образом. Пусть последовательность $\mathcal{X}^{(n)} = \{(X_i, U_i), 1 \le i \le n\}$ независимых и одинаково распределенных с (X, U) пар случайных величин (с.в.) имеет совместную плотность распределения $f(x)g(u) > 0$ и функцию распределения $F(x)G(x)$, где

$$F(x) = \mathbf{P}(X < x), G(u) = \mathbf{P}(U < u),$$
$$F'(x) = f(x), G'(x) = g(x),$$

$W = I(X < U)$ есть индикатор события $(X < U)$. В зависимости доза–эффект вместо выборки $\mathcal{X}^{(n)}$ наблюдается выборка (см. [1])

$$\mathcal{U}^{(n)} = \{(W_i, U_i), 1 \le i \le n\},$$

где величина X интерпретируется как пороговая доза (латентная случайная величина), являющаяся самой большой дозой, при которой не наблюдается эффекта в эксперименте, а U есть введенная доза. Биологическая суть вышеизложенного становится понятной из следующих пояснений. Предположим, что речь идет о яде, который попадает в биообъект. Например, при биотестировании воды речь обычно идет о цериодафниях, которые чувствительны к загрязнению воды вредными веществами и для которых это загрязнение является ядом. Для каждого яда теоретически существует минимальная доза, вызывающая у тест-объекта его гибель. Оценить эту дозу для каждого биообъекта довольно трудно. Если биообъект в эксперименте выжил, то он получил дозу, заведомо меньшую минимальной летальной дозы. Для каждого биообъекта эта доза будет различной, что определяется индивидуальной чувствительностью особей биологического вида к тестируемому препарату. Однако будем считать, что в однородной массе величина X является случайной величиной с неизвестной функцией распределения $F(x)$, плотность распределения $g(U)$ случайной величины U также предполагается неизвестной. Требуется по выборке $U^{(n)}$ оценить неизвестную функцию распределения $F(x)$ и её квантили (называемые в токсикологии категориями эффективных доз) на всем интервале изменения переменной x. Медианная доза обозначается как ЕД50 и называется среднеэффективной дозой. Проблема осложняется тем, что сама величина X ненаблюдаема, а наблюдаются индикаторы w_i и уровни введенных доз u_i. Если вводимые дозы u_t неслучайны, то мы называем такую модель моделью с *фиксированным планом эксперимента*, если же величины u_t случайны, то такая модель называется моделью зависимости доза–эффект со *случайным планом эксперимента*.

В [1] в качестве оценки неизвестной функции распределения использовались оценки Надарая [2] и Ватсона [3] (NW-оценки) вида

$$\hat{F}_{NW}(x) = \frac{\sum_{i=1}^{n} W_i K_h(U_i - x)}{\sum_{i=1}^{n} K_h(U_i - x)}, \tag{1}$$

где $K_h(x) = (1/h)K(x/h)$ – ядерная функция (ядро), $h = h(n)$ есть ширина окна просмотра данных – неслучайная последовательность, сходящаяся к нулю при $n \to\infty$, но при этом $nh \to\infty$.

Обычно NW-оценки вида (1) используются в следующей схеме наблюдений:

$$y_t = m(u_t) + \varepsilon_t, t = 1, 2, ... \tag{2}$$

где $\varepsilon_1, \varepsilon_2,\dots$ – последовательность независимых и одинаково распределенных с.в., а u_t – случайные или фиксированные величины, $m(x)$ – неизвестная функция.

Для неслучайных u_t в схеме наблюдений (2) Priestley и Chao (см. [4]) предложили оценки для функции $m(x)$ по выборке $(u_t, y_t), 1 \le t \le n$, объема n следующего вида:

$$\hat{F}_{PC}(x) = \sum_{t=1}^{n-1} (u_{t+1} - u_t)\, y_t K_h(u_t - x). \qquad (3)$$

Мы будем называть их РС-оценками.

Если $u_t = t/n$, $t = 0,1,\ldots,n$, то РС-оценка принимает вид:

$$\hat{F}_{PC}(x) = n^{-1} \sum_{t=1}^{n-1} y_t K_h(t/n - x). \qquad (4)$$

Известно (см. [4, 5]), что как в схеме наблюдений (2), так и в зависимости доза–эффект для фиксированных планов наблюдений оценки (3) или (4) при некоторых условиях регулярности являются асимптотически нормальными. Однако в зависимости доза–эффект довольно часто планы эксперимента являются случайными, поэтому в данной работе мы изучим оценки вида (3) для случайных планов эксперимента, т.е. мы будем предполагать, что величины u_t являются случайными величинами, и докажем состоятельность и асимптотическую нормальность РС-оценок в этом случае, используя методы работы [6].

1. Основные результаты

Для дальнейшего изложения нам понадобится понятие порядковых и индуцированных порядковых статистик.

Пусть $U^{(1)} < U^{(2)} < \ldots < U^{(n)}$ – вариационный ряд, построенный по совокупности $(U_1, U_2, \ldots, U_n)$. Тогда $U^{(i)}$ называется i-й порядковой статистикой. Индуцированная порядковая статистика $W^{[i]}$ определяется следующим образом: если $U^{(i)} = U_j$, $W^{[i]} = W_j$.

Будем рассматривать статистики

$$\hat{\theta}_n(x) = \sum_{i=1}^{n-1} W^{[i]} (U^{(i+1)} - U^{(i)}) K_h(U_n^{(i)} - x)$$

и

$$F_n(x) = \frac{\sum_{i=1}^{n-1} W^{[i]} (U^{(i+1)} - U^{(i)}) K_h(U_n^{(i)} - x)}{\sum_{i=1}^{n-1} (U^{(i+1)} - U^{(i)}) K_h(U_n^{(i)} - x)} = \frac{\hat{\theta}_n(x)}{\sum_{i=1}^{n-1} (U^{(i+1)} - U^{(i)}) K_h(U_n^{(i)} - x)}.$$

Целью нашего исследования является доказательство асимптотической (при $n \to \infty$) нормальности оценок $\hat{\theta}_n(x)$ и $F_n(x)$. Свойство асимптотической нормальности оценок $\hat{\theta}_n(x)$ и $F_n(x)$ позволяет при достаточно больших n строить доверительные интервалы на функцию распределения $F(x)$, используя квантили нормального распределения.

Именно, мы покажем, что

$$\sqrt{nh}\,(\hat{\theta}(x) - K_h * F(x)) \xrightarrow[n\to\infty]{d} N(0, 2\sigma^2(x)),$$

$$\sqrt{nh}\,(F_n(x) - K_h * F(x)) \xrightarrow[n\to\infty]{d} N(0, 2\sigma^2(x)),$$

где $(K_h * F)(x)$ есть свертка функций $K_h(x)$ и $F(x)$ на интервале (0,1), т.е.

$$(F * K_h)(x) = \int_0^1 F(u) K_h(x-u)\, du, \; 0 < x < 1,$$

а $\sigma^2(x) = \dfrac{F(x)(1-F(x))\|K\|^2}{g(x)}$.

2. Условия

Условия (K)

(K1) $K(x) \ge 0$ для любого $x \in \mathbf{R}$.

(K2) $\int_{-\infty}^{\infty} K(x)\, dx = 1$.

(K3) $K(-x) = K(x)$ для любого $x \in \mathbf{R}$.

(K4) $K(x)$ – финитная функция, т.е. $K(x) = 0$ для $x \notin [-1,1]$.

(K5) $K(x)$ – ограниченная функция, т.е. $\sup_{x\in[-1,1]} |K(x)| = C < \infty$.

Для ядерной функции $K(x)$ определим следующие характеристики:

$$\nu^2 = \int_{-\infty}^{\infty} x^2 K(x)\, dx, \quad \|K\|^2 = \int_{-\infty}^{\infty} K^2(x)\, dx,$$

которые нам понадобятся в дальнейшем.

В силу условий (K1) – (K5) они существуют.

Примерами ядерных функций, удовлетворяющих условиям (K1) – (K5), являются:

- ядро Епанечникова

$$K_0(x) = \frac{3}{4}(1-x^2) I(|x| \le 1);$$

- квартическое ядро

$$K_0(x) = \frac{15}{16}(1-x^2)^2 I(|x| \le 1);$$

где $I(A)$ есть индикатор множества A.

Для ядра Епанечникова $\nu^2 = \frac{1}{5}$, $\|K\|^2 = \frac{3}{5}$, для квартического ядра $\nu^2 = \frac{1}{7}$, $\|K\|^2 = \frac{5}{7}$.

Условия (S)

(S1) Плотность $g(x) > 0$ есть непрерывная и ограниченная функция, которая имеет ограниченные производные до третьего порядка включительно.

(S2) Функция $F(x)$ имеет ограниченные непрерывные производные до третьего порядка включительно.

Примерами функций, удовлетворяющих условиям (S1), (S2), являются функции распределения и плотности нормального, логистического и логнормального распределений.

3. Результаты

Следующая теорема 1 устанавливает асимптотическую нормальность оценок $\hat{\theta}_n(x)$ и $F_n(x)$.

Пусть M – заданная константа.

Теорема 1. *Пусть* $h = Mn^{-1/5}$ *и выполнены условия* (K), (S).

Тогда

$$\sqrt{nh}\left(\hat{\theta}_n(x) - \left(K_h * F\right)(x)\right) \xrightarrow[n\to\infty]{d} N(0, 2\sigma^2(x)),$$

$$\sqrt{nh}\,(F_n(x) - K_h * F(x)) \xrightarrow[n\to\infty]{d} N(0, 2\sigma^2(x)).$$

Заметим, что дисперсия предельного распределения $\sigma_1^2(x) = 2\sigma^2(x)$ зависит от неизвестной функции распределения $F(x)$, и, следовательно, $\sigma^2(x)$ также неизвестна, поэтому в качестве оценки $\sigma^2(x)$ мы предлагаем использовать статистику

$$\hat{\sigma}_1^2(x) = \frac{n}{4}\sum_{i=1}^{n-1}(W^{[i+1]} - W^{[i]})^2 (U^{(i+1)} - U^{(i)})^2 \times \times K_h(U^{(i+1)} - x) K_h(U^{(i)} - x). \quad (4)$$

В следующей теореме утверждается, что оценка $\hat{\sigma}_1^2(x)$ является состоятельной оценкой дисперсии $\sigma_1^2(x)$.

Теорема 2. *Пусть выполнены условия* (K), (S). *Тогда при* $n \to \infty$ $\hat{\sigma}_1^2(x) \xrightarrow{p} \sigma_1^2(x)$.

4. Доказательства

Доказательство теоремы 1.

Пусть

$$\mu(u) = F(u)K_h(u - x),$$

$$d_{[i]} = \left(G(U^{(i+1)}) - G(U^{(i)})\right)K_h(U^{(i)} - x),$$

$$\Delta W_{[i]} = \left(W^{[i]} - F(U^{(i)})\right)K_h(U^{(i)} - x).$$

Представим разность $\hat{\theta}_n(x) - \left(F * K_h\right)(x)$ в следующем виде:

$$\hat{\theta}_n(x) - \left(F * K_h\right)(x) = N_1 + N_2 + N_3,$$

где

$$N_1 = \frac{1}{n}\sum_{i=1}^{n-1}\frac{\Delta W_{[i]}}{g(U^{(i)})} n d_{[i]},$$

$$N_2 = \sum_{i=1}^{n-1}\mu(U^{(i)})(U^{(i+1)} - U^{(i)}) - (F * K_h)(x),$$

$$N_3 = \sum_{i=1}^{n-1}\Delta W_{[i]}\left(U^{(i+1)} - U^{(i)}\right) - \sum_{i=1}^{n-1}\frac{\Delta W_{[i]}}{g(U^{(i)})} d_{[i]}.$$

Покажем, что при $n \to \infty$

$$N_2 = o_p\left((nh)^{-1/2}\right), \quad N_3 = o_p\left((nh)^{-1/2}\right).$$

Рассмотрим слагаемое N_2.

Из представления свертки

$$\left(F * K_h\right)(x) = \sum_{i=1}^{n-1}\int_{U^{(i)}}^{U^{(i+1)}} F(u)K_h(u - x)du$$

следует, что

$$N_2 = \sum_{i=1}^{n-1}\mu(U^{(i)})(U^{(i+1)} - U^{(i)}) - \int_{U^{(i)}}^{U^{(i+1)}} F(u)K_h(u - x)\,du = \sum_{i=1}^{n-1}\left(\mu(U^{(i)}) - \mu(\xi_i)\right)\left(U^{(i+1)} - U^{(i)}\right),$$

где $\xi_i \in [U^{(i)}, U^{(i+1)}]$.

Отсюда

$$\left|N_2\right| \le \sum_{i=1}^{n-1}\left|\mu(U^{(i)}) - \mu(\xi_i)\right|(U^{(i+1)} - U^{(i)}) \le$$

$$\le \sum_{i=1}^{n-1}\left(\left|K_h\left(U^{(i)} - x\right) - K_h\left(\xi_i - x\right)\right| + K_h\left(\xi_i - x\right)\left(F(U^{(i+1)}) - F(U^{(i)})\right)\right)\left(U^{(i+1)} - U^{(i)}\right) \le$$

$$\le \sum_{i=1}^{n-1}\left(\frac{L_1}{h^2} + \frac{L_2}{h}\right)\left(U^{(i+1)} - U^{(i)}\right)^2 = o_p\left((nh)^{-1/2}\right).$$

Для слагаемого N_3 имеем оценки:

$$\left|N_3\right| = \left|\sum_{i=1}^{n-1}\Delta W_{[i]}\left(U^{(i+1)} - U^{(i)}\right) - \sum_{i=1}^{n-1}\frac{\Delta W_{[i]}}{g(U^{(i)})} d_{[i]}\right| =$$

$$= \left|\sum_{i=1}^{n-1}\frac{\Delta W_{[i]}}{g(U^{(i)})}(U^{(i+1)} - U^{(i)}) \times \left(g(U^{(i)}) - g(\xi_i)\right)K_h(U^{(i)} - x)\right|,$$

где $\xi_i \in [U^{(i)}, U^{(i+1)}]$.

Значит, при $n \to \infty$

$$\left|N_3\right| \le \frac{L_3}{h}\left(\frac{1}{n}\sum_{i=1}^{n-1}\frac{\left(\Delta W_{[i]}\right)^2}{g^2\left(U^{(i)}\right)}\right)^{1/2} \times \left(\frac{1}{n}\sum_{i=1}^{n-1} n^2\left(U^{(i+1)} - U^{(i)}\right)^2\right)^{1/2} =$$

$$= O_p(1)\,O_p((nh)^{-1}) = o_p((nh)^{-1/2}).$$

Поэтому

$$\hat{\theta}_n(x) - \left(F * K_h\right)(x) = \frac{1}{n}\sum_{i=1}^{n-1}\frac{\Delta W_{[i]}}{g(U^{(i)})} K_h(U^{(i)} - x) \times n\,(G(U^{(i+1)}) - G(U^{(i)})) + o_p((nh)^{-1/2}).$$

В работе [7] показано, что $\frac{\Delta W_{[i]} K_h(U^{(i)}-x)}{g(U^{(i)})}$ асимптотически независимы от $(G(U^{(i+1)})-$ $-G(U^{(i)}))$. Кроме того, известно, что $G(U_j)$ имеет равномерное распределение на [0,1], поэтому $G(U^{(i)})$ есть i-я порядковая статистика из равномерного на [0,1] распределения.

Известно также (см., например, [8]), что если $\xi_1, \xi_2, \dots, \xi_n$ есть независимые случайные величины из равномерного на [0,1] распределения и $\{\xi_n^{(i)}, i=1,2,\dots,n\}$ – вариационный ряд, построенный по $\xi_1, \xi_2, \dots, \xi_n$, $D_i = \xi^{(i+1)} - \xi^{(i)}$, $\xi^{(0)}=0$, $\xi^{(n+1)}=1$, $i=1,2,\dots,n$, $\{\varsigma_j\}_{1\le j\le n}$ независимы и одинаково распределены с плотностью

$p(x)=e^{-x}$, $x>0$; $T=\sum_{j=1}^{n}\varsigma_j$,

то

$$(D_1, D_2, \dots, D_n) =^d (\varsigma_1/T, \varsigma_2/T, \dots, \varsigma_n/T).$$

Поэтому $nD_i =^d \frac{\varsigma_i}{T/n}$ и $\frac{T}{n} \xrightarrow[n\to\infty]{p} 1$.

В таком случае при $n\to\infty$

$$N_1 \sim \frac{1}{n}\sum_{i=1}^{n-1}\frac{\Delta W_{[i]}\varsigma^{(i)}}{g(U^{(i)})} \sim \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta W_{[i]}\varsigma^{(i)}}{g(U^{(i)})} = \frac{1}{n}\sum_{j=1}^{n}\frac{\Delta W_j \varsigma_j}{g(U_j)},$$

и асимптотическая нормальность N_1 получается теперь из центральной предельной теоремы Линдеберга-Леви (см. [9]).

Найдем дисперсию суммы N_1, считая, что величины $(X_j, U_j, \varsigma_j)_{1\le j\le n}$ независимы как тройками, так и между собой. Для этого воспользуемся формулой для дисперсии произведения независимых величин Z_1, Z_2:

$$\mathbf{D}(Z_1\cdot Z_2) = \mathbf{E}(Z_1^2)\mathbf{E}(Z_2^2) - \mathbf{E}^2(Z_1)\mathbf{E}^2(Z_2),$$

принимая во внимание, что

$$\mathbf{E}\left(\frac{(I(X_j<U_j)-F(U_j))^2}{g^2(U_j)}\right) = \int\frac{F(u)(1-F(u))}{g(u)}K_h^2(u-x)\,du = \sigma^2(x)(1+o(1)),$$

$$\mathbf{E}\left(\frac{I(X_j<U_j)-F(U_j)}{g(U_j)}\right)=0,\ \mathbf{E}(\varsigma_j)=2,$$

и учитывая возможность предельного перехода под знак математического ожидания, получим результат первой части теоремы 1. С небольшими изменениями показывается, что

$$\frac{1}{n}\sum_{i=1}^{n-1}(U^{(i+1)}-U^{(i)})K_h(U^{(i)}-x) \xrightarrow[n\to\infty]{p} 1,$$

поэтому

$$\sqrt{n}\,(F_n(x) - K_h * F(x)) \xrightarrow[n\to\infty]{d} N(0, 2\sigma^2(x)).$$

Оптимальное значение константы M, входящей в определение ширины окна h просмотра данных (см. формулировку теоремы 1), определяется из условия минимума среднеквадратичного отклонения, зависящего от интеграла (см. [10]) $b^2 = \int F(x)(1-F(x))dx$, который необходимо оценивать по выборке $U^{(n)}$. Для оценки величины b^2 можно воспользоваться следующей статистикой

$$\hat{b}^2 = \frac{1}{2n}\sum_{j=2}^{n}(W_j - W_{j-1})^2.$$

Статистика $\hat{b}^2$ асимптотически эквивалентна статистике

$$\tilde{b}^2 = n^{-1}\sum_{j=2}^{n}W_j(1-W_{j-1}),$$

поскольку $\hat{b}^2 = \tilde{b}^2 + O(n^{-1})$.

Покажем, что как $\tilde{b}^2$, так и $\hat{b}^2$ являются состоятельными оценками b^2.

Лемма 1. *Для фиксированных планов эксперимента при* $n\to\infty$ $\hat{b}^2 \xrightarrow{p} b^2$, $\tilde{b}^2 \xrightarrow{p} b^2$.

Доказательство. В силу независимости величин W_j и W_{j-1} получаем:

$$\mathbf{E}(\tilde{b}^2) = \mathbf{E}\left(\frac{1}{n}\sum_{j=2}^{n}W_j(1-W_{j-1})\right) = \mathbf{E}\left(\frac{1}{n}\sum_{j=2}^{n}W_j\right) - \mathbf{E}\left(\frac{1}{n}\sum_{j=2}^{n}W_jW_{j-1}\right) = \frac{1}{n}\sum_{j=2}^{n}F(u_j) - \frac{1}{n}\sum_{j=2}^{n}F(u_j)F(u_{j-1}) = b^2 + O\left(\frac{1}{n}\right).$$

Кроме того, имеем:

$$\mathbf{D}\left(\tilde{b}^2\right) = \frac{1}{n^2}\mathbf{D}\left(\sum_{j=2}^{n}W_j(1-W_{j-1})\right) = \frac{1}{n^2}\left(\mathbf{D}\sum_{j=1}^{n}W_j(1-W_{j-1})\right) + \frac{1}{n^2}\left(\sum_{i\ne j}\mathbf{Cov}\left(W_i(1-W_{i-1}), W_j(1-W_{j-1})\right)\right).$$

Вычислим сначала $\sum_{i\ne j}\mathbf{Cov}\left(W_i(1-W_{i-1}), W_j(1-W_{j-1})\right)$.

Не нарушая общности, рассмотрим случай $i<j$. Множество индексов разобьем на две части: 1) $i=$ $=j-1<j$ и 2) $i<j-1$. В первом случае

$$W_i(1-W_{i-1})W_j(1-W_{j-1}) = W_{j-1}(1-W_{j-1})(1-W_{j-2})W_j = 0,$$

поэтому здесь

$$\mathbf{Cov}\big(W_{j-1}(1-W_{j-2}), W_j(1-W_{j-1})\big)=$$
$$=-F(u_{j-1})(1-F(u_{j-2}))F(u_j)(1-F(u_{j-1})).$$

Во втором случае, т.е. когда $i<j-1$, величины $W_i(1-W_{i-1})$, $W_j(1-W_{j-1})$ независимы, откуда

$$\mathbf{Cov}\big(W_i(1-W_{i-1}), W_j(1-W_{j-1})\big)=0.$$

Так как

$$\mathbf{D}\big(W_j(1-W_{j-1})\big)=\mathbf{E}\big(W_j(1-W_{j-1})\big)-\mathbf{E}^2\big(W_j(1-W_{j-1})\big)=$$
$$=F(u_j)(1-F(u_j))(1-F(u_j)+F^2(u_j)),$$

то получаем, что при $n\to\infty$

$$\mathbf{D}\big(\tilde{b}^2\big)\sim n^{-1}\int F(x)(1-F(x))\times$$
$$\times\big(1-2F(x)+2F^2(x)\big)dx \underset{n\to\infty}{\to} 0.$$

Из неравенства Чебышева теперь следует состоятельность оценок $\tilde{b}^2$ и $\hat{b}^2$.

Поскольку величины, входящие в сумму (1) и (2), ограничены, то из [9, с. 291] следует также асимптотическая нормальность статистик $\hat{b}^2$ и $\tilde{b}^2$ с ожиданием b^2 и дисперсией $\frac{1}{n}\int F(x)(1-F(x))(1-2F(x)+2F^2(x))dx$.

Пусть

$$\hat{\sigma}^2(x)=\frac{1}{4n}\sum_{j=1}^{n-1}(W_{j+1}-W_j)^2K_h^2(U_j-x).$$

Аналогично предыдущему можно показать, что

$$\hat{\sigma}^2(x)\underset{n\to\infty}{\overset{p}{\to}}F(x)(1-F(x))\|K\|^2.$$

Доказательство завершено.

Рассмотрим теперь случайные планы эксперимента, где будем считать, что $\mathbf{P}(0\le U_i\le 1)=1$, и пусть

$$\hat{s}^2=\frac{1}{n}\sum_{i=1}^{n-1}(W^{[i+1]}-W^{[i]})^2.$$

Лемма 2. *Если выполнены условия* (K), (S), *то*

$$\hat{s}^2\underset{n\to\infty}{\overset{p}{\to}}s^2=\int F(u)(1-F(u))g(u)\,du.$$

Доказательство. Представим статистику $\hat{s}^2$ в следующем виде:

$$\hat{s}^2=\frac{1}{2n}\sum_{i=1}^{n}(\Delta w_{[i+1]}-\Delta w_{[i]}+r_{[i]})^2,$$

где

$$\Delta w_{[i]}=\big(W^{[i]}-F(U^{(i)})\big),$$

поскольку $F(x)$ удовлетворяет условию Липшица (так как плотность $f(x)$ ограничена), то

$$|r_{[i]}|=|W^{[i+1]}-W^{[i]}-(\Delta w_{[i+1]}-\Delta w_{[i]})|=$$
$$=|F(U^{(i+1)})-F(U^{(i)})|\le H_1(U^{(i+1)}-U^{(i)}).$$

Тогда

$$\hat{s}^2=S_1+R_1+R_2,$$

где

$$S_1=\frac{1}{2n}\sum_{i=1}^{n-1}(\Delta w_{[i+1]}-\Delta w_{[i]})^2,\quad R_1=\frac{1}{2n}\sum_{i=1}^{n-1}r_{[i]}^2,$$
$$R_2=\frac{1}{n}\sum_{i=1}^{n-1}(\Delta w_{[i+1]}-\Delta w_{[i]})r_{[i]}.$$

Покажем, что $R_1=o_p(n^{-1})$, $R_2=O_p(n^{-1})$. Действительно,

$$R_1=\frac{1}{2n}\sum_{i=1}^{n}r_{[i]}^2\le\frac{1}{2n}\sum_{i=1}^{n}H_1^2(U^{(i+1)}-U^{(i)})^2=$$
$$=\frac{H_1^2}{2n^3}\sum_{i=1}^{n}(n(U^{(i+1)}-U^{(i)}))^2=O_p(n^{-2})=o_p(n^{-1}),$$
$$|R_2|=\frac{1}{n}\sum_{i=1}^{n}|\Delta w_{[i+1]}-\Delta w_{[i]}|\,|r_{[i]}|\le$$
$$\le\frac{H_1}{n}\sum_{i=1}^{n-1}(U^{(i+1)}-U^{(i)})=\frac{H_1(U^{(n)}-U^{(1)})}{n}=O(n^{-1}).$$

Рассмотрим теперь слагаемое S_1. Имеем:

$$S_1=\frac{1}{2n}\sum_{i=1}^{n}(\Delta w_{[i+1]}-\Delta w_{[i]})^2=\frac{1}{2n}\sum_{i=2}^{n-1}(\Delta w_{[i]})^2+$$
$$+\frac{1}{2n}\sum_{i=1}^{n-1}(\Delta w_{[i]})^2-\frac{1}{n}\sum_{i=1}^{n-1}\Delta w_{[i+1]}\Delta w_{[i]}=S_{11}+S_{12}-S_{13}.$$

Очевидно,

$$S_{11}\sim S_{12}\sim\frac{1}{2n}\sum_{i=1}^{n}(\Delta w_i)^2,\quad S_{13}=\frac{1}{n}\sum_{i=1}^{n-1}\Delta w_{[i+1]}\Delta w_{[i]}.$$

Отсюда

$$\mathbf{E}(S_{11})\underset{n\to\infty}{\to}\frac{s^2}{2},\ \mathbf{E}(S_{12})\underset{n\to\infty}{\to}\frac{s^2}{2},$$
$$\mathbf{E}(S_{13})=\frac{1}{n}\sum_{i=1}^{n-1}\mathbf{E}(\Delta w_{[i+1]}\Delta w_{[i]}\mid U^{(i+1)},U^{(i)}).$$

В силу условной независимости индуцированных порядковых статистик (см. [13])

$$\mathbf{E}(S_{13})=\frac{1}{n}\sum_{i=1}^{n-1}\mathbf{E}(\Delta w_{[i+1]}\mid U^{(i+1)})\mathbf{E}(\Delta w_{[i]}\mid U^{(i)})=0.$$

Поскольку слагаемые $\Delta w_{[i+1]}$, $\Delta w_{[i]}$ ограничены, то дисперсии сумм S_{11}, S_{12} сходятся к нулю. Из неравенства Чебышева следует $S_{11}\underset{n\to\infty}{\overset{p}{\to}}\frac{s^2}{2}$, $S_{12}\underset{n\to\infty}{\overset{p}{\to}}\frac{s^2}{2}$, поэтому $\hat{s}^2\underset{n\to\infty}{\overset{p}{\to}}s^2$, что и требовалось доказать.

Доказательство теоремы 2. Рассмотрим оценку (4) или эквивалентную ей оценку

$$\hat{\sigma}_2^2(x)=\frac{n}{4}\sum_{i=1}^{n-1}(W^{[i+1]}-W^{[i]})^2(U^{(i+1)}-U^{(i)})^2\times$$
$$\times K_h^2(U^{(i)}-x).$$

Поскольку $\hat{\sigma}_2^2(x) \le \frac{C}{4n}\sum_{i=1}^{n-1}\left(n(U^{(i+1)}-U^{(i)})\right)^2$, то отсюда следует, что дисперсия статистики $n^{-1}\sum_{i=1}^{n-1}\left(n(U^{(i+1)}-U^{(i)})\right)^2$ сходится к нулю, поэтому $\mathbf{D}(\hat{\sigma}_2^2(x)) \underset{n\to\infty}{\to} 0$.

Если мы покажем, что

$$\mathbf{E}(\hat{\sigma}_2^2(x)) \underset{n\to\infty}{\to} F(x)(1-F(x))\|K\|^2,$$

то результат теоремы 2 будет следовать из неравенства Чебышева.

Заметим, что

$$\mathbf{E}(\hat{\sigma}_2^2(x)) = \mathbf{E}\left(\sum_{i=1}^{n-1}\left(\frac{n}{4}\sum_{i=1}^{n-1}\left(W^{[i+1]}-W^{[i]}\right)^2 \times\right.\right.$$

$$\left.\left.\times(U^{(i+1)}-U^{(i)})^2 K_h^2(U^{(i)}-x)|U^{(i+1)},U^{(i)}\right)\right),$$

а $n(U^{(i+1)}-U^{(i)}) - \frac{1}{n\,g(U^{(i)})} \overset{p}{\underset{n\to\infty}{\to}} 0$

для центральных членов вариационного ряда. Известно (см. [13]), что индуцированные порядковые статистики условно независимы, и поскольку величины $W_1, W_2, \ldots, W_n$ одинаково распределены, то из предыдущих рассуждений следует, что каждое слагаемое последней суммы сходится по вероятности к $\frac{F(x)(1-F(x))\|K\|^2}{g(x)}$. Тогда из теоремы Коши (см. [14, с.79]), примененной к последовательности рассматриваемых случайных величин, будет следовать, что

$$\mathbf{E}(\hat{\sigma}_2^2(x)) \underset{n\to\infty}{\to} \frac{F(x)(1-F(x))\|K\|^2}{g(x)}.$$

Результат теоремы 2 следует теперь из неравенства Чебышева.

Список литературы

1. Тихов М.С., Криштопенко С.В. Попова Е.С. Доза-эффект. М.: Медицина, 2008. 288 с.
2. Надарая Э.А. О непараметрических оценках плотности вероятности и регрессии // Теория вероятн. и ее примен. 1965. Т. 10, в.1. С.199–203.
3. Watson G.S. Smooth regression analysis // Sankhya. 1964. V.26. P. 359–372.
4. Priestley M.B., Chao M.T. Nonparametric function fitting // Journal of the Royal Statistical Society. 1972. Ser.B. V.34. P. 385–392.
5. Тихов М.С., Криштопенко Д.С., Ярощук М.В. Оценивание распределений в зависимости доза-эффект при фиксированном плане эксперимента // Статистические методы оценивания и проверки гипотез: Межвуз. сб. научных трудов. Пермь: Изд-во Пермского ун-та, 2006. С. 66–77.
6. Lewbel A., Schennach S. Estimator for inverse density weighted // Journal of Econometrics. 2007. V.136, №1. P.189–211.
7. Barbe P. Joint approximation of process based on spacing and order statistics // Stochastic Process and Their Applications. 1994. V.53. P. 339–349.
8. Гаек Я., Шидак З. Теория ранговых критериев. М.: Наука, 1971. 376 с.
9. Лоэв М. Теория вероятностей. М.: ИЛ, 1962. 719 с.
10. Тихов М.С. Построение и анализ статистических оценок для неполностью известных семейств распределений: Дис. … д-ра физ.-мат. наук. Нижний Новгород, 1993. 354 с.
11. Li J., Liu R.Y. Multivariate spacings based on data depth: I. Construction of nonparametric multivariate tolerance regions // Ann. Statist. 2008. V.36, No.3. P. 1299–1329.
12. Pino G.E. On the asymptotic distribution of k-spacings with applications to goodness-of-fit tests // Ann. Statist. 1979. V.7, No.5. P. 1058–1065.
13. David H.A., Nagaraja H.N. Order Statistics. John Wiley & Sons, 2003. 458 p.
14. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Т.1. М.: Физматлит, 2001. 680 с.

PC-ESTIMATORS OF DISTRIBUTION FUNCTION OVER RANDOM EXPERIMENT PLANS IN A DOSE-EFFECT MODEL

M.S. Tikhov

Nonparametric PC-estimators of a distribution function over random experiment plans in a dose-effect model are considered. PC-estimators have been shown to be consistent and asymptotically normal. We propose and investigate an estimator of limiting variance.

Keywords: dose-effect model, nonparametric kernel estimators, random experiment plan.