

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

МОДИФИЦИРОВАННЫЙ МЕТОД ОЦЕНИВАНИЯ РИДА И МЕНЧА В ЗАВИСИМОСТИ ДОЗА-ЭФФЕКТ

Тихов М.С., Шкилева К.Н.

Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского, г. Нижний Новгород

Поступила в редакцию 17.10.2019, после переработки 12.11.2019.

В этой статье пересматриваются оценки функции распределения в зависимости доза-эффект, введенные Reed'ом и Muench'ем (1938): предлагается новая статистика для оценки функции распределения, когда основная переменная не обязательно нормально распределена. Мы доказываем состоятельность и асимптотическую нормальность этих непараметрических оценок. Первоначальные статистики Рида и Менча, предназначенные для оценки медианной дозы ED_{50} , модифицированы таким образом, чтобы построить оценки функции распределения и эффективных доз $ED_{100\lambda}$ в широком диапазоне λ : от 0.05 до 0.95, которые достаточно устойчивы при отклонении от исходной модели. Разработан алгоритм стохастической аппроксимации для оценки функции распределения. Доказана сходимость этого алгоритма. Чтобы проиллюстрировать практическую полезность нашего подхода, разработанные в статье методики используются для вычисления среднего возраста прорезывания премоляров у мальчиков. Статистические данные взяты из работы Hayes и Mantel (1958). Проводится также исследование методом Монте-Карло на основе смоделированных данных. Результаты показывают, что непараметрические оценки функции распределения, рассмотренные в данной работе, хорошо работают на практике, в некоторых случаях даже для относительно небольших размеров выборки.

Ключевые слова: зависимость доза-эффект, метод Рида и Менча, стохастическая аппроксимация, асимптотическая нормальность, имитационное моделирование.

Вестник ТвГУ. Серия: Прикладная математика. 2019. № 4. С. 5–26.
<https://doi.org/10.26456/vtprm543>

Введение

В статье рассматриваются новые оценки функции распределения и ее квантилей в зависимости доза-эффект, которая описывается следующим образом.

Пусть случайная величина X имеет распределение с неизвестной функцией распределения $F(x)$ и плотностью $f(x)$. Пусть имеется еще одна случайная величина U , плотность распределения которой равна $g(x) > 0$ и которая может быть как известна, так и неизвестна. Наблюдению доступны пары (u_i, w_i) , $1 \leq i \leq n$, (повторная выборка), где w_i есть значение случайной величины W_i – индикатора события $(X_i < U_i)$. Требуется по наблюдениям $\mathcal{U}^{(n)} = \{(u_i, w_i)\}_{i=1}^n$ оценить неизвестную функцию распределения $F(x)$.

В работе [1] было замечено, что в случае, когда величины X и U – независимы, то условное ожидание $\mathbf{E}(W|U = x) = F(x)$, т.е. является регрессией, и поэтому для оценки $F(x)$ можно использовать ядерную оценку регрессии:

$$\hat{F}_n(x) = \frac{S_{2n}(x)}{S_{1n}(x)}, \quad (1)$$

где

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

Здесь $K(x)$ есть ядерная функция (ядро), т.е. четная плотность распределения, с носителем на отрезке $[-1, 1]$, а $h = h(n)$ есть параметр сглаживания. Из соображений минимизации среднеквадратичного уклонения в качестве h берут $h(n) = Cn^{-1/5}$. В [1], [2] показано, что при $n \rightarrow \infty$ разность $\sqrt{nh}(\hat{F}_n(x) - F(x))$ при фиксированном x и при некоторых условиях регулярности асимптотически нормальна $N(a(x), b^2(x))$, где $a(x) = (f'(x) + 2f(x)g'(x))/g(x)$, $b^2(x) = F(x)(1 - F(x)) \|K\|^2 / g(x)$, $\|K\|^2 = \int_{-\infty}^{+\infty} K^2(x) dx$.

Используются и другие оценки $F(x)$, например, kNN -оценки (см. [3]), предельная дисперсия нормированной разности которой $\tilde{b}^2(x) = F(x)(1 - F(x)) \|K\|^2$ не зависит от плотности $g(x)$. Многомерный аналог kNN -оценки предложен и исследован в работах [4], [5].

В приложениях обычно интересуются оценкой квантиля $x_\lambda = F^{-1}(\lambda) = \inf\{x \in R, F(x) \geq \lambda\}$, $0 < \lambda < 1$, распределения $F(x)$ вида $\hat{x}_{n\lambda} = \inf\{x \in R, \hat{F}_n(x) \geq \lambda\}$. Они изучались в работах [3], [6], [7]. В работах [8], [9] были предложены оценки квантиля, не требующие нахождения обратной функции.

Что касается асимптотического смещения $a(x)$, то устранению смещения ядерной оценки плотности посвящена работа [10], иной способ устранения смещения в регрессионной схеме $Y_i = m(X_i) + \varepsilon_i$, где ε_i – независимые случайные величины со средним 0 и дисперсией σ^2 , независимые от ковариат (X_1, X_2, \dots, X_n) , рассматривался в [11]. В работе [12] рассматривалась модификация этого метода устранения смещения (в зависимости доза-эффект) для оценок функции распределения путем устранения смещения как числителя, так и знаменателя.

В токсикологии медиана $x_{0.5}$ называется среднеэффективной дозой и обозначается ED_{50} . Монография D.Finney [13] является в этом вопросе классической. Одним из основных методов оценивания среднеэффективной дозы, который там предложен, является метод Спирмена-Кербера (см. также [14]). Однако этот метод предназначен, в основном, для оценки ED_{50} . Преимущество оценки (1) и ее квантиля перед оценками Спирмена-Кербера, состоит в том, что она позволяет

строить достаточно хорошие оценки в широком диапазоне значений λ , именно, для $0.05 \leq \lambda \leq 0.95$, т.е. от ED_5 до ED_{95} .

В книге [14] рассматривался еще один метод оценивания ED_{50} – это метод Риды и Менча (первоисточник – [15]). Однако качество оценок по методу Риды и Менча подвергалось сомнению (см., например, [16]). Заметим также, что рассматривавшиеся в [14] методы, основанные на выборках $\mathcal{U}^{(n)}$, базируются на распределениях типа нормального (или логнормального).

Поскольку в данной работе речь пойдет о модификации метода Риды и Менча, рассмотрим пример из книги [14] и поясним его применение.

Имеем следующие статистические данные.

Таблица 1: Оценка среднеэффективной дозы методом Риды и Менча

Логарифм дозы	Число тест-объектов в опыте	Частота эффекта		Положительные ответы	Накопленная частота эффекта				Пробит
		есть (3)	нет (4)		есть (6)	нет (7)	сумма (8)	процент (9)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2.4	6	0	6	0.0	0	17	17	0.0	3.16
2.8	7	1	6	14.3	1	11	12	8.2	3.93
3.2	7	3	4	42.9	4	5	9	44.5	4.82
3.6	6	5	1	83.3	9	1	10	90.0	5.97
4.0	6	6	0	100.0	15	0	15	100.0	6.84

Пояснения к Таблице 1: Например, в столбце (6) число $4 = 0+1+3$ – сумма значений, взятых из столбца (3), $9 = 4+5$, $15 = 9+6$, и т.д. В столбце (7) суммирование идет снизу вверх.

Приведем цитату из [14], с.237: Этот метод исходит из того естественного допущения, что если некоторый тест-объект дал положительный ответ при какой-либо дозе, то он дал такой же ответ и при более высоких дозах; и наоборот, если некоторый тест-объект дал отрицательный ответ при определенной дозе, то он дал также отрицательный ответ и при всех меньших дозах.

Расчет ED_{50} производится следующим образом. Пусть A – процент положительных ответов, ближайший к 50% снизу, а B – ближайший к 50% сверху (в примере $A = 44.5$, $B = 90.0$), а l_A и l_B – соответствующие дозы (в примере $l_A = 3.2$, $l_B = 3.6$). Тогда

$$ED_{50} = l_A + (l_B - l_A) \cdot \frac{50 - A}{B - A} = 3.2 + (3.6 - 3.2) \cdot \frac{50 - 44.5}{90 - 44.5} = 3.248.$$

Сравним это значение со значением, использующим оценку (1). Для оценки $\hat{F}_n(x)$ по данным таблицы 1 имеем ($n = 32$, $h = 0,9$): $ED_{50} = 3,255$, т.е. эти оценки близки.

Пусть

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

Определим $H(x) = \int_{-1}^x K(t) dt$, для $-1 \leq x \leq 1$, и $H(x) = 0$, если $x \notin [-1, 1]$.

Так, если $K_1(t) = 0.75(1 - x^2)$, $-1 \leq x \leq 1$, то $H(x) = 0.5 + 0.75x - 0.25x^3$.

В данной статье предлагается новая оценка функции распределения $F(x)$ – в некотором смысле модификация метода Рида и Менча. Именно, оценку функции распределения $F(x)$ в точке x будем находить по формуле (считая, что $\tilde{F}_n(x) = 0$, если $S_{5n}(x) = 0$):

$$\tilde{F}_n(x) = \frac{S_{3n}(x)}{S_{5n}(x)}, \quad S_{5n}(x) = S_{3n}(x) + S_{4n}(x). \quad (2)$$

Модифицированный метод Рида и Менча, т.е. оценка, найденная по формуле (2), дает $ED_{50} = 3,243$. В данной работе мы исследуем качество оценки (2), а также разбираем ситуацию, когда дозы не являются случайными величинами, а фиксированы заранее (фиксированный план эксперимента).

1. Предположения

Пусть $K(x)$ – ядерная функция, где $x \in \mathbf{R}$.

Предположения (К).

(К₁) $K(x) \geq 0$, причем $K(x) = 0$, $x \notin [-1, 1]$.

(К₂) $\int_{-\infty}^{\infty} K(x) dx = 1$.

(К₃) $K(x) = K(-x)$, $x \in \mathbf{R}$.

(К₄) $\|K\|_{\infty} = \sup_{x \in \mathbf{R}} |K| = \kappa < \infty$.

(К₅) Существуют непрерывные ограниченные производные функции $K(x)$ до третьего порядка включительно в интервале $(-1, 1)$.

Так как при приведенных условиях $\int_{-\infty}^{\infty} K^2(x) dx < \infty$, то положим

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

При условиях (К) преднорма $\|K\|^2$ ядра K конечна и существуют четвертые моменты для распределений с плотностями $K(x)$:

$$\nu^2 = \int_{-\infty}^{\infty} x^2 K(x) dx, \quad \mu^4 = \int_{-\infty}^{\infty} x^4 K(x) dx.$$

Примером функции, удовлетворяющей условиям (К), служит:

1. $K(x) = (3/4)(1 - x^2)I(|x| \leq 1)$ (ядро Епанечникова),

$$\|K\|^2 = 3/5 = 0.6, \quad \nu^2 = 1/5 = 0.2,$$

2. $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$ (квартическое ядро),

$$\|K\|^2 = 5/7 = 0.714, \quad \nu^2 = 1/7 = 0.143.$$

Пусть $h = h(n)$ – ширина окна просмотра данных.

Предположение (H).

(H₁) При $n \rightarrow \infty$, $h \rightarrow 0$, $nh \rightarrow \infty$.

Условие (H₁) означает, что по мере получения большего количества информации из выборки, т.е. при $n \rightarrow \infty$, усреднение данных происходит по более узкой области ($h \rightarrow 0$), но в то же время количество «локальной информации» (nh) должно увеличиваться.

Примером числовой последовательности, удовлетворяющей условию (H₁), является $h = cn^{-1/5}$, c – некоторая положительная константа.

Предположения (F).

(F₁) Случайные величины X и U – независимы.

(F₂) Плотности $f(x) > 0$ и $g(x) > 0$ на некотором интервале $(a, b) \subset \mathbf{R}$ ограничены и имеют ограниченные производные до третьего порядка включительно.

2. Результаты

Для того, чтобы оценить качество предложенной оценки рассмотрим ее математическое ожидание и дисперсию. Положим $S_j \equiv S_{jn}(x)$, $j = 3, 4, 5$. Имеем:

$$\begin{aligned} \mathbf{E}(S_3) &= \frac{1}{nh} \mathbf{E} \left(\sum_{i=1}^n W_i H \left(\frac{U_i - x}{h} \right) \right) = \frac{1}{h} \mathbf{E} \left(W_1 H \left(\frac{U_1 - x}{h} \right) \right) = \\ &= \frac{1}{h} \mathbf{E} \left(\mathbf{E} \left(W_1 \cdot H \left(\frac{U_1 - x}{h} \right) \right) | U_1 = u \right) = \frac{1}{h} \int_{-\infty}^{\infty} H \left(\frac{u - x}{h} \right) F(u) g(u) du = \\ &= \int_{-\infty}^{+\infty} H(t) F(x + ht) g(x + ht) dt = F(x)g(x) \int_{-1}^1 H(t) dt + \\ &+ (Fg)'(x)h \int_{-1}^1 tH(t) dt + \frac{1}{2}(Fg)''(x)h^2 \int_{-1}^1 t^2H(t) dt + o(h^2). \end{aligned} \quad (3)$$

Пусть $\mu_m(K) = \int_{-1}^1 t^m K(t) dt$. Так как $\int_{-1}^1 t^\beta H(t) dt = \frac{1}{\beta + 1} (1 - \mu_{\beta+1}(K))$, где $\mu_\beta(K) = \int_{-1}^1 t^\beta K(t) dt$ и $K(t) = K(-t)$, $t \in \mathbf{R}$, то

$$\int_{-1}^1 t^{2m-1} H(t) dt = \frac{1}{2m} (1 - \mu_{2m}(K)), \quad \text{и} \quad \int_{-1}^1 t^{2m} H(t) dt = \frac{1}{2m+1}, \quad m \in \mathbf{N}.$$

Заметим, что

$$\int_{-1}^1 H(t) dt = \int_{-1}^0 H(t) dt + \int_0^1 (H(t) - 1) dt + 1 = 1 - \mu_1(K) = 1,$$

поскольку $\mu_1(K) = 0$. Кроме того,

$$\int_{-1}^1 t H(t) dt = \frac{1}{2} (1 - \mu_2(K)), \quad \mu_3(K) = \int_{-1}^1 t^2 H(t) dt = \frac{1}{3},$$

где $\mu_2 = \int_{-1}^1 t^2 K(t) dt$, $K(t) = K(-t)$.

Значит, $\mathbf{E}(S_3) = F(x)g(x) + \frac{1}{2}(1 - \mu_2(K))h (Fg)'(x) + \frac{1}{6}h^2(Fg)''(x) + o(h^2)$ и если $\mu_2(K) = 1$, то $\mathbf{E}(S_3) = F(x)g(x) + \frac{1}{6}h^2(Fg)''(x) + o(h^2)$.

Пример 1. Пусть $K_0(t) = (1/\sqrt{5})K_1(x/\sqrt{5})$, $-\sqrt{5} < x < \sqrt{5}$, и ноль в остальных случаях, тогда $\mu_2(K_0) = 1$ и $1 - \mu_2(K_0) = 0$.

Далее мы будем предполагать, что $\mu_2(K_0) = 1$, тогда

$$\begin{aligned} \mathbf{E}(S_3 + S_4) &= \mathbf{E} \left(\frac{1}{nh} \sum_{i=1}^n W_i H \left(\frac{U_i - x}{h} \right) + \frac{1}{nh} \sum_{i=1}^n (1 - W)_i \left(1 - H \left(\frac{U_i - x}{h} \right) \right) \right) = \\ &= \int_{-1}^1 F(x + ht)g(x + ht)H(t) dt + \int_{-1}^1 (1 - F(x + ht))g(x + ht)(1 - H(t)) dt. \end{aligned}$$

Поскольку

$$\begin{aligned} \int_{-1}^1 (1 - F(x + ht))g(x + ht)(1 - H(t)) dt &= \int_{-1}^1 (1 - F(x - ht))g(x - ht)(1 - H(-t)) dt = \\ &= \int_{-1}^1 (1 - F(x - ht))g(x - ht)H(t) dt, \end{aligned}$$

то

$$\begin{aligned} \mathbf{E}(S_3 + S_4) &= \int_{-1}^1 F(x + ht)g(x + ht)H(t) dt + \int_{-1}^1 (1 - F(x - ht))g(x - ht)H(t) dt = \\ &= g(x) + \frac{1}{2}(1 - \mu_2(K))(2(Fg)'(x) - g'(x))h + \frac{1}{6}h^2 g''(x) + o(h^2) = g(x) + \frac{1}{6}h^2 g''(x) + o(h^2). \end{aligned}$$

Рассмотрим теперь дисперсии статистик S_3 и S_5 . Имеем:

$$\begin{aligned} \mathbf{D}(S_3) &= \mathbf{D} \left(\frac{1}{nh} \sum_{i=1}^n W_i H \left(\frac{U_i - x}{h} \right) \right) = \frac{1}{nh^2} \mathbf{D} \left(WH \left(\frac{U - x}{h} \right) \right) = \\ &= \frac{1}{nh^2} \left(\mathbf{E} \left(W^2 H^2 \left(\frac{U - x}{h} \right) \right) - \mathbf{E}^2 \left(WH \left(\frac{U - x}{h} \right) \right) \right) \sim \frac{1}{nh^2} \mathbf{E} \left(W H^2 \left(\frac{U - x}{h} \right) \right) - \\ &\quad - \frac{1}{n} F^2(x)g^2(x) \sim \frac{1}{nh} F(x)g(x) \int_{-1}^1 H^2(t) dt = \frac{1}{nh} F(x)g(x) \|H\|^2. \end{aligned}$$

Обозначим $G_i = W_i H_i + (1 - W_i)(1 - H_i)$, $H_i = H\left(\frac{U_i - x}{h}\right)$. Тогда

$$\mathbf{D}(S_3 + S_4) = \frac{1}{n^2 h^2} \mathbf{D}\left(\sum_{i=1}^n G_i\right) = \frac{1}{n h^2} \mathbf{D}(G_1), \mathbf{D}(G_1) = \mathbf{E}(G_1^2) - \mathbf{E}^2(G_1).$$

Значит,

$$\mathbf{D}(S_3 + S_4) \sim \frac{1}{n h^2} \mathbf{E}(G_1^2) - \frac{g(x)}{n} \sim \frac{1}{n h^2} \mathbf{E}(G_1^2).$$

Далее,

$$\begin{aligned} \frac{1}{n h^2} \mathbf{E}(G_1^2) &= \frac{1}{n h^2} \mathbf{E}\left(\left(W H\left(\frac{U-x}{h}\right) + (1-W)\left(1 - H\left(\frac{U-x}{h}\right)\right)\right)^2\right) = \\ &= \frac{1}{n h^2} \mathbf{E}\left(W H^2\left(\frac{U-x}{h}\right) + (1-W)\left(1 - H\left(\frac{U-x}{h}\right)\right)^2 + \right. \\ &\quad \left. + 2W(1-W)H\left(\frac{U-x}{h}\right)\left(1 - H\left(\frac{U-x}{h}\right)\right)\right) = \\ &= \frac{1}{n h^2} \mathbf{E}\left(W H^2\left(\frac{U-x}{h}\right) + (1-W)\left(1 - H\left(\frac{U-x}{h}\right)\right)^2\right) = \\ &= \frac{1}{n h^2} \int \left(F(x) H^2\left(\frac{y-x}{h}\right) + (1-F(x))\left(1 - H\left(\frac{y-x}{h}\right)\right)^2\right) g(y) dy = \\ &= \frac{1}{n h} \int_{-1}^1 (F(x+ht) H^2(t) + (1-F(x+ht))(1-H(t))^2) g(x+ht) dt \sim \\ &\sim \frac{1}{n h} \int_{-1}^1 (F(x) H^2(t) + (1-F(x))(1-H(t))^2) g(x) dt = \frac{1}{n h} g(x) \|H\|^2, \end{aligned}$$

поскольку

$$\int_{-1}^1 (1-H(t))^2 dt = \int_{-1}^1 (1-H(-t))^2 dt = \int_{-1}^1 H^2(t) dt = \|H\|^2.$$

Таким образом,

$$\mathbf{D}(S_5) \sim \frac{1}{n h} g(x) \|H\|^2.$$

Пусть $\|f\|_p = \left(\int_a^b f^p(x) dx\right)^{1/p}$, $\|f\|_\infty = \sup_{a \leq x \leq b} |f(x)| = M$, $p > 2$ (см. [19]).

Для $p > 1$ нам понадобится следующее неравенство

$$|x + y|^p \leq 2^{p-1} (|x|^p + |y|^p). \quad (4)$$

Рассмотрим $\mathbf{E}(|\xi_j - \mathbf{E}(\xi_j)|^{2+\delta})$, $\delta > 0$, где $\xi_j = \frac{1}{h} W_j H\left(\frac{U_j - x}{h}\right)$.

Тогда $S_{3n}(x) = \frac{1}{n} \sum_{j=1}^n \xi_j$. Используя неравенство (4), имеем:

$$|\xi_j - \mathbf{E}(\xi_j)|^{2+\delta} \leq 2^{1+\delta} (|\xi_j|^{2+\delta} + |\mathbf{E}(\xi_j)|^{2+\delta}).$$

Беря от обеих частей математическое ожидание, получаем

$$\mathbf{E}(|\xi_j - \mathbf{E}(\xi_j)|^{2+\delta}) \leq 2^{1+\delta} (\mathbf{E}(|\xi_j|^{2+\delta}) + |\mathbf{E}(\xi_j)|^{2+\delta}) \leq 2^{2+\delta} (\mathbf{E}(|\xi_j|^{2+\delta})).$$

Из неравенств Минковского, Гельдера и определения величины ξ_j следует:

$$\begin{aligned} & \mathbf{E}(|\xi_j - \mathbf{E}(\xi_j)|^{2+\delta}) \leq 2^{2+\delta} (\mathbf{E}(|\xi_j|^{2+\delta})) \leq \\ & \leq \left(\frac{2}{h}\right)^{2+\delta} \int_{-\infty}^{+\infty} |H((x-u)/h)|^{2+\delta} q(u) du = \frac{2^{2+\delta}}{h^{1+\delta}} \int_{-\infty}^{+\infty} |H((x-u)/h)|^{2+\delta} q(u) du = \\ & = \frac{2^{2+\delta}}{h^{1+\delta}} \int_{-\infty}^{+\infty} |H(y)|^{2+\delta} q(x+hy) dy = \frac{2^{2+\delta}}{h^{1+\delta}} q(x) \int_{-\infty}^{+\infty} |H(y)|^{2+\delta} dy (1+o(1)) \leq \\ & \leq \frac{2^{2+\delta}}{h^{1+\delta}} q(x) \|H\|_2^{2+\delta} \leq \frac{2^{2+\delta} C_1}{h^{1+\delta}}. \end{aligned}$$

Значит, для дроби Ляпунова,

$$L_n = \frac{\sum_{j=1}^n \mathbf{E}(|\xi_j - \mathbf{E}(\xi_j)|^{2+\delta})}{(\sum_{j=1}^n \mathbf{D}(\xi_j))^{1+\delta/2}} \leq \frac{2^{2+\delta} C_1 (nh)^{1+\delta/2}}{(nh)^{1+\delta} (F(x)g(x) \|H\|_2^2)} = \frac{C}{n^{2\delta/5}} \xrightarrow{n \rightarrow \infty} 0,$$

т.е. условие центральной предельной теоремы Ляпунова [17], с.241, выполнено и

$$\frac{S_{3n}(x) - \mathbf{E}(S_{3n}(x))}{\sqrt{\mathbf{D}(S_{3n}(x))}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Рассуждая аналогично можно показать, что,

$$\frac{S_{5n}(x) - \mathbf{E}(S_{5n}(x))}{\sqrt{\mathbf{D}(S_{5n}(x))}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$$

Положим $a_3 = \mathbf{E}(S_3)$, $a_5 = \mathbf{E}(S_5)$, $a_{3m} = \mathbf{E}(S_{3m})$ и рассмотрим отношение S_3/S_5 . Учитывая неравенство $\left| \frac{1}{1+x} - 1+x \right| \leq 2x^2$, $|x| \leq \frac{1}{2}$, имеем:

$$\begin{aligned} \frac{S_3}{S_5} &= \frac{S_3 - a_3 + a_3}{S_5 - a_5 + a_5} = \frac{S_3 - a_3}{a_3 (1 + (S_3 - a_3)/a_3)} + \frac{a_3}{a_3 (1 + (S_3 - a_3)/a_3)} = \\ &= \frac{S_3 - a_3}{a_5} \cdot \left(1 - \frac{S_5 - a_5}{a_5}\right) + \frac{a_3}{a_5} - \frac{a_3}{a_5} \cdot \frac{S_5 - a_5}{a_5} + Q_{1n}, \\ |Q_{1n}| &\leq \frac{2(S_5 - a_5)^2}{a_5^2} \cdot \frac{|S_3 - a_3|}{a_5} + \frac{2a_3(S_5 - a_5)^2}{a_5^2}. \end{aligned}$$

Тогда

$$\frac{S_3}{S_5} - \frac{a_3}{a_5} = \frac{S_3 - a_3}{a_5} - \frac{a_3(S_5 - a_5)}{a_5^2} + Q_{2n},$$

где $Q_{2n} = \frac{(S_3 - a_3)(S_5 - a_5)}{a_5^2} (1 + o(1))$, откуда

$$\frac{S_3}{S_5} \xrightarrow[n \rightarrow \infty]{p} F(x), \quad \text{т.е.} \quad \hat{F}_n(x) \xrightarrow[n \rightarrow \infty]{p} F(x). \quad (5)$$

Как и в [18], с.40 можно также показать, что

$$S_3 - a_3 \xrightarrow[n \rightarrow \infty]{a.s.} 0 \quad \text{и} \quad \sqrt{nh} \max_{1 \leq k \leq n} (S_{3k} - a_{3k})^2 \xrightarrow[n \rightarrow \infty]{p} 0.$$

Кроме того,

$$\mathbf{D}(\tilde{F}_n(x)) = \frac{\|H\|_2^2}{nh} \frac{F(x)(1-F(x))}{g(x)} (1 + o(1)).$$

Таким образом, мы доказали следующий результат.

Теорема 1. Пусть выполнены предположения **(K)**, **(H)**, **(F)**. Тогда

$$\sqrt{nh}(\tilde{F}_n(x) - \mathbf{E}(\tilde{F}_n(x))) \xrightarrow[n \rightarrow \infty]{d} N(0, \|H\|_2^2 F(x)(1-F(x))/g(x)).$$

3. Устранение смещения

Для устранения смещения мы воспользуемся подходом [11], модернизированным для нашей ситуации. Рассмотрим статистики $S_{3,n}(x)$ и $S_{5,n}(x)$. Введем диагональную матрицу

$$A = \begin{pmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & Z_n \end{pmatrix}, \quad \text{обозначая} \quad Z_i = \frac{1}{nh} H \left(\frac{x - U_i}{h} \right).$$

Определим «октаэдрическую» норму вектора $x = (x_1, x_2, \dots, x_n)^T$ как в [19], гл. 14, §2:

$$\|x\| = \sum_{j=1}^n |x_j|.$$

Тогда оценку $S_{3,n}(x)$ можно переписать в виде $S_{3,n}(x) = \|AW\|$, где $W = (W_1, \dots, W_n)^T$. Обозначим также $m = m(x) = \mathbf{E}(W|U = x)$. В этих обозначениях

$$S_{3,n}(x) = \|\hat{m}_1(x)\|, \quad \text{где} \quad \hat{m}_1(x) = AW := A_1 W.$$

Смещение этой оценки $\hat{m}_1(x)$ равно

$$\begin{aligned} b_1 &= \mathbf{E}(\hat{m}_1|U = x) - m(x) = \mathbf{E}(AW|U = x) - m(x) = A E(W|U = x) - m(x) = \\ &= (A - I)m(x) = -(I - A)m(x). \end{aligned}$$

Здесь I – единичная матрица соответствующей размерности.

Подставляя вместо $m(x)$ ее оценку $\hat{m}_1(x)$, получим оценку смещения:

$$\hat{b}_1 = -(I - A)\hat{m}_1(x) = -A(I - A_1)W.$$

После этого построим новую оценку

$$\hat{m}_2(x) = A_1 W - \hat{b}_1 = A_1 W + A(I - A_1)W = A_2 W,$$

где $A_2 = A_1 + A(I - A_1) = A + A(I - A)$.

Полагая $A_k = A_{k-1} + A(I - A_{k-1})$, получим

$$\hat{m}_k(x) = A_{k-1} W - \hat{b}_{k-1} = A_{k-1} W + A(I - A_{k-1})W = A_k W. \quad (6)$$

Предположим, что $A_k = I - (I - A)^k$, откуда $I - A_k = (I - A)^k$. Тогда

$$A_{k+1} = A_k + A(I - A_k) = I - (I - A)^k + A(I - A)^k = I - (I - A)^{k+1}.$$

Таким образом,

$$\hat{m}_k(x) = A_k W = (I - (I - A)^k) W, \quad \text{и} \quad \hat{b}_k = - (I - A)^{k-1} A W. \quad (7)$$

Кроме того, умножая $A = I - (I - A)$ на $I + (I - A) + (I - A)^2 + \dots + (I - A)^{k-1}$, получим, что

$$A(I + (I - A) + (I - A)^2 + \dots + (I - A)^{k-1}) = I - (I - A)^k,$$

поэтому

$$\begin{aligned} \hat{m}_k(x) &= A(I + (I - A) + (I - A)^2 + \dots + (I - A)^{k-1}) W = \\ &= \hat{m}_1(x) + A((I - A) + (I - A)^2 + \dots + (I - A)^{k-1}) W. \end{aligned} \quad (8)$$

Пусть $A = (a_{ij})_{n \times n}$, а $\|A\|$ — ее норма, т.е. $\|A\| = \sup_{\|x\|=1} \|Ax\|$.

Тогда $\|A\| = \max_{1 \leq k \leq n} \sum_{j=1}^n |a_{j,k}|$ (см. [19], гл. 14, §2).

Для двух $n \times n$ -матриц A и B при одном и том же определении векторных норм имеем:

$$\|A + B\| \leq \|A\| + \|B\|, \quad \|AB\| \leq \|A\| \cdot \|B\|.$$

Пусть $\|I - A\| = \lambda$. Для достаточно большого n , λ удовлетворяет условию $0 < \lambda < 1$. Тогда

$$\begin{aligned} \|\hat{b}_k\| &= \|(I - A)^{k-1} A W\| = \|(I - A)(I - A)^{k-2} A W\| \leq \\ &\leq \|(I - A)\| \cdot \|\hat{b}_{k-1}\| < \lambda \|\hat{b}_{k-1}\|. \end{aligned}$$

Следовательно, $\|\hat{b}_k\| < \lambda^{k-1} \|\hat{b}_1\|$, $k \geq 2$, откуда $\lim_{k \rightarrow \infty} \|\hat{b}_k\| = 0$.

Из (8) следует, что

$$\begin{aligned} \|\hat{m}_k(x)\| &\leq \|\hat{m}_1(x)\| + \|A\|(\lambda + \lambda^2 + \dots + \lambda^{k-1})\|W\| \leq \\ &\leq \|\hat{m}_1(x)\| + \|A\| \cdot \|W\| \cdot \frac{\lambda}{1 - \lambda} \leq \|\hat{m}_1(x)\| + \frac{\lambda(1 + \lambda)}{1 - \lambda} \|W\| \leq \\ &\leq \|\hat{m}_1(x)\| + \frac{\lambda(1 + \lambda)n}{1 - \lambda}. \end{aligned}$$

С другой стороны, из (8) следует

$$\begin{aligned} \|\hat{m}_k(x)\| &= \|A_{k-1}W - A \cdot A_{k-1}W + AW\| \geq \\ &\geq \|\hat{m}_1(x)\| - \|A - I\|(1 - \lambda^{k-1})\|W\| \geq \\ &\geq \|\hat{m}_1(x)\| - \lambda(1 - \lambda^{k-1})\|W\| \geq \|\hat{m}_1(x)\| - \lambda(1 - \lambda^{k-1})n. \end{aligned}$$

Так как

$$\mathbf{D}(\hat{m}_k|U = x) = (I - (I - A)^k)m(x) \cdot m^T(x)(I - (I - A)^k)^T,$$

то при фиксированном достаточно большом n и при $k \rightarrow \infty$,

$$\begin{aligned} \mathbf{D}(S_{3,n}(x)) &= \mathbf{E}((\hat{m}_k(x) - m(x))^2) = \mathbf{E}((\hat{m}_1(x) - m(x))^2)(1 + o(1)) = \\ &= \frac{\|H\|^2 F(x)g(x)}{nh} + o\left(\frac{1}{nh}\right). \end{aligned}$$

Аналогично, если через $\hat{q}_k(x)$ обозначим оценку, асимптотически устраняющую смещение оценки $S_{5,n}(x)$, то $\mathbf{D}(S_{5,n}(x)) = \frac{\|H\|^2 g(x)}{nh} (1 + o(1))$. Тогда оценка $\hat{F}_k(x)$ является асимптотически несмещенной (при $k \rightarrow \infty$) и при $n \rightarrow \infty$,

$$\mathbf{D}(\hat{F}_n(x)) = \frac{\|H\|^2 F(x)(1 - F(x))}{nh g(x)} (1 + o(1)).$$

4. Оценивание функции распределения методом стохастической аппроксимации

Для оценки функции распределения $F(u)$ мы будем использовать рекурсивный метод стохастической аппроксимации [20], [21]. Преимущество рекурсивных оценок перед их нерекурсивной версией заключается в следующем: чтобы обновить оценку от выборки объема n до оценки по выборке объема $n + 1$ требуется значительно меньше вычислений. Это свойство особенно важно в рамках оценивания функции распределения в зависимости доза-эффект, так как количество точек, в которых оценивается функция распределения, может оказаться большим. Вычисления производятся следующим образом:

Пусть

$$Z_n = H_h(x - U_n), \quad \tilde{F}_n(x) = \frac{S_{3,n}(x)}{S_{5,n}(x)}, \quad n = 1, 2, \dots$$

Положим $S_{3,1}(x) = W_1 Z_1$, $S_{5,1}(u) = W_1 Z_1 + (1 - W_1)(1 - Z_1)$, и для $n \geq 2$ будем последовательно вычислять

$$S_{3,n}(u) = (1 - \gamma_n) S_{3,n-1}(x) + \gamma_n W_n Z_n,$$

$$S_{5,n}(x) = (1 - \gamma_n) S_{5,n-1}(x) + \gamma_n (W_n Z_n + (1 - W_n)(1 - Z_n)), \quad 0 < \gamma_n < 1,$$

выбор γ_n уточним ниже.

Мы будем рассматривать *правильно (регулярно) меняющиеся* последовательности $(h_n)_{n \geq 1}$, $h_n = h(n)$.

Определение ([22], [23]). Пусть $\alpha \in \mathbf{R}$ и $(v_n)_{n \geq 1}$ есть неслучайная последовательность. Мы будем говорить, что последовательность $(v_n)_{n \geq 1}$ является *правильно меняющейся с показателем α* (и писать $(v_n)_{n \geq 1} \in RV_\alpha$), если

$$\lim_{n \rightarrow \infty} n \left(\frac{v_{n-1}}{v_n} - 1 \right) = \alpha. \quad (9)$$

Пример 2. Пусть $v_n = \frac{1}{n^\alpha (\ln n)^\beta}$, $\alpha > 0$, $\beta > 0$. Тогда

$$\begin{aligned} \frac{v_{n-1}}{v_n} - 1 &= \frac{n^\alpha \ln^\beta n}{(n-1)^\alpha \ln^\beta(n-1)} - 1 = \\ &= \left(1 + \frac{1}{n-1} \right)^\alpha \left(1 + \frac{1}{(n-1) \ln(n-1)} \right)^\beta (1 + o(1)) - 1 = \\ &= \left(1 + \frac{\alpha}{n-1} + o\left(\frac{1}{n}\right) \right) \left(1 + \frac{\beta}{(n-1) \ln(n-1)} + o\left(\frac{1}{n \ln n}\right) \right) - 1 = \frac{\alpha}{n} + o\left(\frac{1}{n}\right). \end{aligned}$$

Отсюда $\lim_{n \rightarrow \infty} n \left(\frac{v_{n-1}}{v_n} - 1 \right) = \alpha$.

Пусть

$$\beta_n = \prod_{j=2}^n (1 - \gamma_j), \quad n \geq 2.$$

Для доказательства теоремы 2 нам потребуется следующая лемма.

Лемма. Пусть $(v_n)_{n \geq 1} \in RV_\alpha$, $\lim_{n \rightarrow \infty} (n\gamma_n) = 1/\xi > 0$. Тогда

$$\lim_{n \rightarrow \infty} v_n \beta_n \sum_{k=2}^n \beta_k^{-1} \frac{\gamma_k}{\nu_k} = \frac{1}{1 + \alpha\xi}.$$

Доказательство. Обозначим

$$\beta_{mn} = \prod_{k=m+1}^n (1 - \gamma_k), \quad \text{если } 1 \leq m < n, \quad \text{и пусть } \gamma_n = \frac{1}{\xi n}.$$

Известно, что

$$(1 - \varepsilon_m) \exp\left(-\sum_{k=m+1}^n \gamma_k\right) \leq \beta_{mn} \leq \exp\left(-\sum_{k=m+1}^n \gamma_k\right)$$

для $n \geq m$, где $\varepsilon_m \rightarrow 0$ при $m \rightarrow \infty$. Так как (см. [24], с.16, 0.131)

$$\sum_{k=1}^n \frac{1}{k} = \gamma + \ln n + \frac{1}{2n} - \sum_{k=2}^{\infty} \frac{A_k}{n(n+1) \cdot \dots \cdot (n+k-1)},$$

где

$$A_k = \frac{1}{k} \int_0^1 x(1-x) \cdot \dots \cdot (k-1-x) dx,$$

(здесь γ – константа Эйлера), поэтому (см. [25])

$$\sum_{k=1}^n \frac{1}{k} - \gamma - \ln n = \int_n^\infty \frac{x - \lfloor x \rfloor}{x^2} dx \quad \text{и} \quad \frac{1}{2(n+1)} < \sum_{k=1}^n \frac{1}{k} - \gamma - \ln n < \frac{1}{2n},$$

$\lfloor x \rfloor$ – целая часть числа x . Отсюда получаем

$$(1 - \tilde{\varepsilon}_m) \left(\frac{m}{n}\right)^{1/\xi} \leq \beta_{mn} \leq (1 + \tilde{\varepsilon}_m) \left(\frac{m}{n}\right)^{1/\xi},$$

где $\tilde{\varepsilon}_m \rightarrow 0$ при $m \rightarrow \infty$.

Условие (9) эквивалентно тому, что $v_n = n^{-\alpha} L(n)$, где $L(n)$ есть медленно меняющаяся функция (см. [26]). Из определения медленно меняющейся функции следует, что $\frac{L(m)}{L(n)} \rightarrow 1$ при $m \rightarrow \infty$, поэтому $\frac{v_n}{v_m} = \left(\frac{m}{n}\right)^\alpha$ при $m \rightarrow \infty$. Отсюда следует, что

$$\lim_{n \rightarrow \infty} v_n \beta_n \sum_{k=m_0+1}^n \beta_k^{-1} \frac{\gamma_k}{\nu_k} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=m_0+1}^n \left(\frac{k}{n}\right)^{\alpha+1/\xi} \frac{n}{\xi k},$$

и поэтому

$$\begin{aligned} \lim_{n \rightarrow \infty} v_n \beta_n \sum_{k=m_0+1}^n \beta_k^{-1} \frac{\gamma_k}{\nu_k} &= \lim_{n \rightarrow \infty} \frac{1}{\xi n} \sum_{k=m_0+1}^n \left(\frac{k}{n}\right)^{\alpha+1/\xi-1} = \\ &= \frac{1}{\xi} \int_0^1 x^{\alpha+1/\xi-1} dx = \frac{1}{1+\alpha\xi}, \end{aligned}$$

что завершает доказательство леммы.

Пример 3. Пусть $\gamma_n = \frac{1}{n}$, $v_n = \frac{1}{n^\alpha}$, $\alpha > 0$, $n \geq 2$. Тогда $\beta_n = \frac{1}{n}$, $\beta_k^{-1} = k$ и

$$v_n \beta_n \sum_{k=2}^n \beta_k^{-1} \frac{\gamma_k}{\nu_k} = \frac{1}{n^{1+\alpha}} \sum_{k=2}^n k^\alpha \sim \int_0^1 x^\alpha dx = \frac{1}{1+\alpha}.$$

Теорема 2. Пусть выполнены предположения (К), (Н), (F) и

$\lim_{n \rightarrow \infty} (n\gamma_n) = 1/\xi$, $a = 1/5$. Тогда при $n \rightarrow \infty$,

$$\mathbf{E}(\tilde{F}_n(x)) - F(x) = \frac{h^2(f'(x)g(x) + 2f(x)g'(x))}{6(1-2a\xi)g(x)}(1+o(1)), \quad (10)$$

$$\mathbf{D}(\tilde{F}_n(x)) = \frac{\|H\|^2 F(x)(1-F(x))}{\xi(2-(1+ad)\xi)nhg(x)}(1+o(1)). \quad (11)$$

Доказательство. Рассмотрим $S_{3,n}(x)$. Мы имеем:

$$S_{3,n}(x) - F(x)g(x) = (1 - \gamma_n)(S_{3,n-1}(x) - F(x)g(x)) + \gamma_n(Z_n - F(x)g(x)) =$$

$$\begin{aligned}
&= \sum_{k=2}^{n-1} \left(\prod_{j=k+1}^n (1 - \gamma_j) \right) \gamma_k (Z_k - F(x)g(x)) + \\
&+ \gamma_n (Z_n - F(x)g(x)) + \left(\prod_{j=2}^n (1 - \gamma_j) \right) (Z_1 - F(x)g(x)) = \\
&= \beta_n \sum_{k=2}^n \beta_k^{-1} \gamma_k (Z_k - F(x)g(x)) + \beta_n (Z_1 - F(x)g(x)),
\end{aligned}$$

поэтому

$$\mathbf{E}(S_{3n}(x) - F(x)g(x)) = \beta_n \sum_{k=2}^n \beta_k^{-1} \gamma_k \mathbf{E}(Z_k - F(x)g(x)) + \beta_n \mathbf{E}(Z_1 - F(x)g(x)).$$

Повторяя теперь доказательство теоремы 1 и применяя лемму, получаем:

$$\mathbf{E}(S_{3,n}(x) - F(x)g(x)) = \frac{F(x)g''(x)h^2}{6(1 - 2a\xi)}(1 + o(1)).$$

Аналогично,

$$\mathbf{E}(S_{5,n}(x) - g(x)) = \frac{h^2 g''(x)}{6(1 - 2a\xi)}(1 + o(1)),$$

откуда следует соотношение (10) теоремы 2.

Что же касается, например, $\mathbf{D}(S_{3n}(x))$, то в случае $h_n = n^{-1/5}$, будем иметь:

$$\begin{aligned}
\mathbf{D}(S_{3n}(x)) &= \beta_n^2 \sum_{k=2}^n \beta_k^{-2} \gamma_k^2 h_k^{-d} F(x)g(x) \|H\|^2 (1 + o(1)) = \\
&= \frac{1}{\xi(2 - (1 + ad)\xi)nh} F(x)g(x) \|H\|^2 (1 + o(1)).
\end{aligned}$$

Теперь аналогичные соотношения можно получить и для $\mathbf{D}(S_{5n}(x))$, а далее, повторяя доказательство теоремы 1, выводим соотношение (11) теоремы 2.

5. Имитационное моделирование и применение

Приведем результаты численного моделирования и вычисления оценок эффективных доз. Моделировались случайные величины: X – нормальная случайная величина $N(17.5, 4)$; U – нормальная случайная величина $N(18, 4)$; ε – нормальная случайная величина $N(0, 0.09)$; $Y = U + \varepsilon$ – нормальная случайная величина $N(18, 4.09)$. Результаты моделирования сведены в таблицу 2.

Сравним оценки (1) и (2). Имеем:

1) Параметр сглаживания $h = 3.0$, значение функции распределения $\hat{F}_n(17.3455) = 0.5$; $\hat{F}_n(14.24) = 0.05$; $\hat{F}_n(20.923) = 0.95$.

Заметим, что теоретическая функция распределения это $F_X(x) = \Phi\left(\frac{x - 17.5}{2}\right)$, где $\Phi(x)$ – это функция распределения стандартной нормальной величины, а теоретические квантили равны: $x_{0.05} = 14.2103$, $x_{0.5} = 17.5$, $x_{0.95} = 20.7897$.

Таблица 2: Результаты моделирования нормальной случайной величины $Y \in N(18.0, 4.09)$ и индикатора $W = I(X < U)$

i	$Y_i = U_i + \varepsilon_i$	$W_i = I(X_i < U_i)$	i	$Y_i = U_i + \varepsilon_i$	$W_i = I(X_i < U_i)$
1	12.965	0	26	18.109	1
2	13.861	0	27	18.116	1
3	14.232	0	28	18.134	1
4	14.483	0	29	18.136	1
5	14.673	0	30	18.149	1
6	14.696	0	31	18.229	1
7	14.919	0	32	18.324	0
8	14.924	0	33	18.349	1
9	15.557	0	34	18.435	0
10	15.640	0	35	18.607	1
11	15.863	0	36	18.775	1
12	15.887	0	37	18.810	0
13	15.904	1	38	18.949	1
14	16.099	0	39	19.201	1
15	16.426	0	40	19.415	1
16	16.506	0	41	19.833	1
17	16.655	0	42	19.968	0
18	16.734	1	43	20.046	1
19	16.776	0	44	20.204	1
20	16.897	0	45	20.211	1
21	17.058	1	46	20.954	1
22	17.085	0	47	21.590	1
23	17.126	1	48	21.662	1
24	17.145	0	49	22.144	1
25	17.480	1	50	23.580	1

2) Параметр сглаживания $h = 6.0$, значение функции распределения $\tilde{F}_n(14.204) = 0.05$; $\tilde{F}_n(17.4335) = 0.5$; $\tilde{F}_n(20.725) = 0.95$.

Таким образом оценки квантилей, построенные по моделированной выборке, близки к теоретическим значениям, правда, сами параметры сглаживания отличаются: для модифицированной оценки Рида и Менча он больше. Его значение можно найти используя процедуру cross-validation.

Приведем также практический пример (таблица 3) применения предложенных оценок, взятый из работы [27], и сравним их значения со значениями, найденными в этой работе методом Кербера. Приведенный пример – это оценка возраста (год и месяц) выпадения премоляров у детей.

В работе [27] приведены оценки среднеэффективной дозы $ED_{50}^* = 9.547$ и сред-

Таблица 3: Время выпадения премоляров у детей (см. [27]: Table I. Kärber's Method as User for Estimation of Eruption Constants, Mandibular Canines, Females, Prince Georges and Montgomery Counties, Mariland, 1952, p.940)

i	U_i	n_i	m_i	i	U_i	n_i	m_i	i	U_i	n_i	m_i
1	6-9	12	0	25	8-9	18	3	49	10-9	32	23
2	6-10	14	2	26	8-10	36	5	50	10-10	24	23
3	6-11	30	0	27	8-11	26	7	51	10-11	28	25
4	7-0	34	0	28	9-0	20	6	52	11-0	24	23
5	7-1	34	0	29	9-1	40	14	53	11-1	20	18
6	7-2	20	0	30	9-2	42	17	54	11-2	30	29
7	7-3	40	1	31	9-3	52	12	55	11-3	32	32
8	7-4	26	0	32	9-4	22	7	56	11-4	24	24
9	7-5	30	1	33	9-5	34	16	57	11-5	22	20
10	7-6	32	0	34	9-6	28	11	58	11-6	30	30
11	7-7	26	3	35	9-7	26	18	59	11-7	20	19
12	7-8	24	0	36	9-8	34	17	60	11-8	36	35
13	7-9	14	0	37	9-9	34	19	61	11-9	22	21
14	7-10	24	1	38	9-10	18	14	62	11-10	22	22
15	7-11	32	1	39	9-11	24	18	63	11-11	40	38
16	8-0	34	4	40	10-0	44	33	64	12-0	30	30
17	8-1	24	0	41	10-1	24	16	65	12-1	34	34
18	8-2	40	1	42	10-2	36	27	66	12-2	44	42
19	8-3	42	3	43	10-3	38	32	67	12-3	32	31
20	8-4	32	3	44	10-4	46	38	68	12-4	40	40
21	8-5	46	6	45	10-5	36	31	69	12-5	42	42
22	8-6	30	5	46	10-6	22	16	70	12-6	28	28
23	8-7	28	5	47	10-7	26	22	71	12-7	36	35
24	8-8	26	4	48	10-8	40	33	72	12-8	44	44

неквадратичного уклонения $s = 0.0407$.

Приведем расчеты по формулам (1) и (2) (объем выборки $N = 2196$ равен сумме значений n_i):

1) Параметр сглаживания $h = 2.4$, значения функции распределения равны $\hat{F}_n(6.703) = 0.05$; $\hat{F}_n(9.58) = 0.5$; $\hat{F}_n(12.352) = 0.95$.

2) Параметр сглаживания $h = 4.5$, значения функции распределения равны $\tilde{F}_n(6.78) = 0.05$; $\tilde{F}_n(9.58) = 0.5$; $\tilde{F}_n(12.31) = 0.95$.

Таким образом, три оценки медианной среднеэффективной дозы близки, но две последние оценки позволяют строить хорошие оценки на краях функции распределения, именно, ED_5 и ED_{95} , а метод Кербера позволяет рассчитать только оценку математического ожидания (для нормального распределения значения медианы и математического ожидания совпадают, но для несимметричных распределений они отличаются). Отметим также, что, во-первых, в нашем случае оценка среднеквадратичного уклонения равна $\tilde{s} = 0.004543$ – это лучше, чем оценка Кербера среднеквадратичного уклонения, а, во-вторых, гладкость оценок функции распределения, как правило, выше, чем у оценок (2).

Заключение и выводы

В работе рассмотрена задача статистического оценивания неизвестной функции распределения модифицированным методом Рида и Менча (статистики (2)) по выборке фиксированного объема. Они состоятельны и асимптотически нормальны. Это позволяет при больших объемах выборки рассчитывать доверитель-

ные интервалы для значений функции распределения. Используя статистики (2) мы получаем также хорошие оценки среднеэффективных доз $ED_{100-\lambda}$ в пределах от 5% до 95%, причем гладкость оценок на хвостах распределений выше, чем у обычных оценок типа Надарая-Ватсона.

Список литературы

- [1] Тихов М.С. Статистическое оценивание на основе интервальных цензурированных данных // Статистические методы оценивания и проверки гипотез: межвуз. сб. Пермь: Пермский университет, 2000. С. 49–70.
- [2] Tikhov M.S. Statistical estimation based on interval censored data // Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life. Eds. by N. Balakrishnan, M.S. Nikulin, M. Mesbah, N. Limnios. Series: Statistics for Industry and Technology. Boston, MA: Birkhauser, 2004. Pp. 211–218. https://doi.org/10.1007/978-0-8176-8206-4_14
- [3] Тихов М.С., Ярошук М.В. Асимптотическая нормальность kNN-оценок в зависимости доза-эффект // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Математика. 2006. № 1(4). С. 129–137.
- [4] Tikhov M., Ivkin M. Multivariate k-Nearest Neighbors Distribution Function Estimates in Dose-effect Relationship // Proceedings of the 2014 International Conference on Mathematical Models and Methods in Applied Sciences, MMAS'2014. Pp. 325–329.
- [5] Tikhov M., Ivkin M. A new Yang-type estimator of Distribution Function at Quantal Response over Indirect Data // Wseas Transactions on Mathematics. 2014. Vol. 13. Pp. 684–693.
- [6] Бородина Т.С., Тихов М.С. Асимптотическая нормальность непараметрических оценок квантилей // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Математика. 2013. № 1. С. 196–207.
- [7] Tikhov M.S., Borodina T.S. Kernel Quantile Estimators in the Dose-Effect Dependence // Automatic Control and Computer Sciences. 2013. Vol. 47, № 2. Pp. 75–86. <http://doi.org/10.3103/S0146411613020089>
- [8] Dette H., Neumeier N., Pilz K. A Note on Nonparametric Estimation of the Effective Dose in Quantal Bioassay // Journal of the American Statistical Association. 2005. Vol. 100. Pp. 503–510.
- [9] Тихов М.С. Непараметрическое оценивание эффективных доз по данным бинарных откликов // Уфимский математический журнал. 2013. Т. 5, № 2. С. 94–108.
- [10] Hengartner N.W., Matzner-Lober E. Asymptotic unbiased density estimators // ESAIM: Probability and Statistics. 2009. Vol. 13. Pp. 1–14.

- [11] Cornillon P.A., Hengartner N., Matzner-Lober E. Recursive Bias Estimation for multivariate regression smoothers // *ESAIM: Probability and Statistics*. 2014. Vol. 18. Pp. 483–502.
- [12] Тихов М.С., Долгих И.С. Асимптотически несмещенные оценки функции распределения в зависимости доза-эффект // *Статистические методы оценивания и проверки гипотез*. 2005. С. 60–72.
- [13] Finney D.J. *Probit analysis*. Cambridge University Press, 2009. 256 p.
- [14] Урбах В.Ю. *Статистический анализ в биологических и медицинских исследованиях*. М.: Медицина, 1975. 297 с.
- [15] Reed L.J., Muench H. A simple method of estimating fifty per cent endpoints // *The American Journal of Hygiene*. 1938. Vol. 27, № 3. Pp. 493–497.
- [16] Гуревич К.Г. Оценка параметров кривой «доза-эффект» методом сплайн-интерполяции // *Вестник Московского университета. Серия 2: Химия*. 2000. Т. 41, № 1. С. 69–70.
- [17] Гнеденко Б.В. *Курс теории вероятностей*. М.: Эдиториал УРСС, 2005. 448 с.
- [18] Тихов М.С. Фурье-метод рекурсивного оценивания функции распределения в зависимости доза-эффект // *Вестник ТвГУ. Серия: Прикладная математика*. 2018. № 4. С. 31–49. <https://doi.org/10.26456/vtpmk516>
- [19] Гантмахер Ф.Р. *Теория матриц*. М.: Наука, 2010. 560 с.
- [20] Вазан М. *Стохастическая аппроксимация*. М.: Мир, 1972. 296 с.
- [21] Robbins H., Monro S. Stochastic Approximation Method // *The Annals of Mathematical Statistics*. 1951. Vol. 22, № 3. Pp. 400–407.
- [22] Galambos J., Seneta E. Regularly varying sequences // *Proceeding of the American Mathematical Society*. 1973. Vol. 41, № 1. Pp. 110–116.
- [23] Seneta E. Variants Karamata's Iteration Theorem // *Publications de l'Institut Mathématique*. 2006. Vol. 80(94). Pp. 241–251.
- [24] Градштейн И.С., Рыжик И.М. *Таблицы интегралов, сумм, рядов и произведений*. М.: ГИФМЛ, 1963. 1100 с.
- [25] Chen C.P. Inequalities for the Euler-Mascheroni constant // *Applied Mathematics Letters*. 2010. Vol. 23. Pp. 161–164.
- [26] Сенета Е. *Правильно меняющиеся функции*. М.: Наука, 1985. 142 с.
- [27] Hayes R.L., Mantel N. Procedures for computing the mean age of eruption of human teeth // *Journal of Dental Research*. 1958. Vol. 35, № 5. Pp. 938–947.

Образец цитирования

Тихов М.С., Шкилева К.Н. Модифицированный метод оценивания Риды и Менча в зависимости доза-эффект // Вестник ТвГУ. Серия: Прикладная математика. 2019. № 4. С. 5–26. <https://doi.org/10.26456/vtpmk543>

Сведения об авторах**1. Тихов Михаил Семенович**

профессор лаборатории прикладной теории вероятностей кафедры программной инженерии института информационных технологий, математики и механики Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского.

Россия, 603950, г. Нижний Новгород, пр. Гагарина, д. 23, ННГУ им. Н.И. Лобачевского. E-mail: tikhovm@mail.ru

2. Шкилева Ксения Николаевна

студентка института информационных технологий, математики и механики Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского.

Россия, 603950, г. Нижний Новгород, пр. Гагарина, д. 23, ННГУ им. Н.И. Лобачевского. E-mail: shkileva98@mail.ru

A MODIFIED REED-MUENCH METHOD OF ESTIMATION IN DOSE-EFFECT RELATIONSHIP

Tikhov Mikhail Semenovich

Professor of Applied Probability Theory laboratory at Software Engineering
department, Institute of Information Technology, Mathematics and Mechanics,
Lobachevsky State University of Nizhniy Novgorod
Russia, 603950, Nyzhniy Novgorod, 23 Gagarin av., UNN.
E-mail: tikhovm@mail.ru

Shkileva Ksenia Nikolaevna

Student of the Institute of Information Technology, Mathematics and Mechanics,
Lobachevsky State University of Nizhniy Novgorod
Russia, 603950, Nyzhniy Novgorod, 23 Gagarin av., UNN.
E-mail: shkileva98@mail.ru

Received 17.10.2019, revised 12.11.2019.

This paper revises the estimators for a distribution function in the dose-effect relationship as Reed and Muench (1938) by proposing a new statistics to estimate the distribution function when the main variable is not necessarily normally distributed. We prove the consistency and asymptotic normality of this estimates without assuming the form of a parametric family. The initial statistics of Reed and MENCH designed to estimate the median dose of ED_{50} are modified in such a way that it is possible to construct stable estimates of the distribution function and effective doses of $ED_{100\lambda}$ in a wide range of λ : from 0.05 to 0.95. A stochastic approximation algorithm for estimating of the distribution function is designed. The convergence theorem of this algorithm is proved. To illustrate the practical utility of our approach, the techniques developed in the paper are used for computing the mean age of eruption of the premolars in boys. The statistical data are from the Hayes&Mantel (1958) work. Finally, a Monte Carlo exercise is performed based on the on simulated data. The results show that the non-parametric estimates of the distribution function considered in this paper work well in practice, in some cases even for relatively small sample sizes.

Keywords: dose-effect relationship, method of Reed and Muench, stochastic approximation, asymptotic normality, simulation modeling.

Citation

Tikhov M.S., Shkileva K.N., "A modified Reed-Muench method of estimation in dose-effect relationship", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2019, № 4, 5–26 (in Russian).
<https://doi.org/10.26456/vtpmk543>

References

- [1] Tikhov M.S., “Statistical Estimation on the Basis of Interval-Censored Data”, *Statisticheskie Metody Ozenivaniya i Proverki Gipotez: Mezhvuz. Sb. [Interuniversity Transactions on Statistical Method of Estimation and Testing Hypotheses]*, Perm University, Perm, 2000, 49–70 (in Russian).
- [2] Tikhov M.S., “Statistical estimation based on interval censored data”, *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*, Statistics for Industry and Technology, eds. N. Balakrishnan, M.S. Nikulin, M. Mesbah, N. Limnios, Birkhauser, Boston, MA, 2004, 211–218, https://doi.org/10.1007/978-0-8176-8206-4_14.
- [3] Tikhov M.S., Yaroshchuk M.V., “Asymptotic normality in dose-effect relationship”, *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Matematika [Bulletin of Lobachevsky State University of Nizhny Novgorod. Series: Mathematics]*, 2006, № 1(4), 129–137 (in Russian).
- [4] Tikhov M., Ivkin M., “Multivariate k-Nearest Neighbors Distribution Function Estimates in Dose-effect Relationship”, *Proceedings of the 2014 International Conference on Mathematical Models and Methods in Applied Sciences, MMAS'2014*, 325–329.
- [5] Tikhov M., Ivkin M., “A new Yang-type estimator of Distribution Function at Quantal Response over Indirect Data”, *Wseas Transactions on Mathematics*, **13** (2014), 684–693.
- [6] Borodina T.S., Tikhov M.S., “Asymptotic normality of nonparametric estimates of quantiles”, *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Matematika [Bulletin of Lobachevsky State University of Nizhny Novgorod. Series: Mathematics]*, 2013, № 1, 196–207 (in Russian).
- [7] Tikhov M.S., Borodina T.S., “Kernel Quantile Estimators in the Dose-Effect Dependence”, *Automatic Control and Computer Sciences*, **47:2** (2013), 75–86, <http://doi.org/10.3103/S0146411613020089>.
- [8] Dette H., Neumeyer N., Pilz K., “A Note on Nonparametric Estimation of the Effective Dose in Quantal Bioassay”, *Journal of the American Statistical Association*, **100** (2005), 503–510 (in Russian).
- [9] Tikhov M.S., “Neparametricheskoe otsenivanie effektivnykh doz po dannym binarnykh otklikov”, *Ufimskij matematicheskij zhurnal [Ufa Mathematical Journal]*, **5:2** (2013), 94–108 (in Russian).
- [10] Hengartner N.W., Matzner-Lober E., “Asymptotic unbiased density estimators”, *ESAIM: Probability and Statistics*, **13** (2009), 1–14.
- [11] Cornillon P.A., Hengartner N., Matzner-Lober E., “Recursive Bias Estimation for multivariate regression smoothers”, *ESAIM: Probability and Statistics*, **18** (2014), 483–502.

- [12] Tikhov M.S., Dolgih I.S., “Asymptotically unbiased estimates of a distribution function in dose-effect relationships”, *Journal Mathematical Sciences*. V. 205, 2015, 113–120.
- [13] Finney D.J., *Probit analysis*, Cambridge University Press, 2009, 256 pp., <https://dspace.gipe.ac.in>.
- [14] Urbakh V.Yu., *Statisticheskij analiz v biologicheskikh i meditsinskikh issledovaniyakh [Statistical analysis in biological and medical research]*, Medicina Publ., Moscow, 1975 (in Russian), 297 pp.
- [15] Reed L.J., Muench H., “A simple method of estimating fifty per cent endpoints”, *The American Journal of Hygiene*, **27**:3 (1938), 493–497.
- [16] Gurevich K.G., “Evaluation of Parameters of the Dose-Effect Curve by Soline-Interpolation Method”, *Vestnik Moskovskogo universiteta. Seriya 2: Khimiya [Moscow University Chemistry Bulletin]*, **41**:1 (2000), 69–70 (in Russian).
- [17] Gnedenko B.V., *The Theory of Probability*, AMS Chelsea Publishing, Providence, Rhode Island, 2005, 529 pp.
- [18] Tikhov M.S., “Fourier methods for recursive estimating of distribution function in dose-effect relationship”, *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2018, № 4, 31–49 (in Russian), <https://doi.org/10.26456/vtppmk516>.
- [19] Gantmakher F.R., *The Theory of Matrices*, Nauka Publ., Moscow, 2010 (in Russian), 560 pp.
- [20] Vazan M., *Stokhasticheskaya approksimatsiya [Stochastic Approximation]*, Mir Publ., Moscow, 1972 (in Russian), 296 pp.
- [21] Robbins H., Monro S., “Stochastic Approximation Method”, *The Annals of Mathematical Statistics*, **22**:3 (1951), 400–407.
- [22] Galambos J., Seneta E., “Regularly varying sequences”, *Proceeding of the American Mathematical Society*, **41**:1 (1973), 110–116.
- [23] Seneta E., “Variants Karamata’s Iteration Theorem”, *Publications de l’Institute Mathématique*, **80(94)** (2006), 241–251.
- [24] Gradshtejn I.S., Ryzhik I.M., *Tablitsy integralov, summ, ryadov i proizvedenij [Table of Integrals, Series and Products]*, GIFML Publ., Moscow, 1963 (in Russian), 1100 pp.
- [25] Chen C.P., “Inequalities for the Euler-Mascheroni constant”, *Applied Mathematics Letters*, **23** (2010), 161–164.
- [26] Seneta E., *Pravilno menyayushchiesya funktsii [Regularly Varying Functions]*, Nauka Publ., Moscow, 1985 (in Russian), 142 pp.
- [27] Hayes R.L., Mantel N., “Procedures for computing the mean age of eruption of human teeth”, *Journal of Dental Research*, **35**:5 (1958), 938–947.