ISSN 2074-1871 Уфимский математический журнал. Том 14. № 4 (2022). С. 100-116.

УДК 519.2

ОТРИЦАТЕЛЬНАЯ БИНОМИАЛЬНАЯ РЕГРЕССИЯ В ЗАВИСИМОСТИ ДОЗА-ЭФФЕКТ

М.С. ТИХОВ

Аннотация. Эта статья посвящена проблеме оценки функции распределения и ее квантилей в зависимости доза-эффект с непараметрической отрицательной биномиальной регрессией. Большая часть математических исследований зависимости доза-эффект касалась моделей с биномиальной регрессией, в частности моделей с бинарными данными. Здесь предложены ядерные оценки функции распределения, ядро которых взвешивается отрицательной биномиальной случайной величиной при каждой ковариате. Эти ковариаты являются квазислучайными ван дер Корпута и Холтона последовательностями с медленным расхождением. Наши оценки состоятельны, т.е. сходятся к своим оптимальным значениям когда число наблюдений n возрастает до бесконечности. Предлагаемые оценки сравниваются с помощью их среднеквадратичных отклонений. Показано, что наши оценки имеют меньшую асимптотическую дисперсию по сравнению, в частности, с оценками типа Надарая-Ватсона и других оценок. Представлены непараметрические оценки квантилей, полученные путем инвертирования ядерной оценки функции распределения. Асимптотическая нормальность этих оценок с поправкой на смещение сохраняется при некоторых условиях регулярности. Мы даем также многомерное обобщение полученных результатов.

Ключевые слова: модель отрицательного биномиального отклика, эффективная доза, непараметрическая оценка.

Mathematics Subject Classification: 62G05, 62E20, 62P10

1. Введение

Задача оценивания неизвестного распределения является важнейшей задачей математической статистики как для полных, так и для неполных выборок. В настоящей статье рассматривается проблема построения эффективных оценок функции распределения (ф. р.) $F(x)$ и квантильной функции $F^{-1}(\lambda) = x_\lambda$, $0 < \lambda < 1$ в зависимости доза-эффект для модели *отрицательной биномиальной регрессии*, а также изучается асимптотическое поведение предложенных оценок. Цель нашего сообщения состоит в том, чтобы дать пригодные для использования на практике оптимальные оценки кривых доза-эффект. Такие задачи возникают в биологии [1], токсикологии [2], в оценке эффективных доз лекарственных препаратов [3]. Отметим также, что «зависимость доза-эффект» — это условное название. Рассматриваемая нами модель может быть применена, например, для оценки доверительных временных границ стадий развития ребенка в педиатрии (см. [1], [4]–[6]). Наиболее остро эта проблема возникает при оценивании квантилей либо малых, либо относительно высоких уровней.

Существует два основных подхода к оценке $F(x)$ и ее квантилей: параметрический подход с использованием известных распределений, в частности, пробит- и логит-модели, и непараметрический подход. Биологические механизмы действия и токсичности лекарств часто настолько сложны, что форма кривой $F(x)$ в значительной степени неизвестна и подгонка неправильной

M.S. Tikhov, Negative binomial regression in dose-effect relationships.

Поступила 18 ноября 2021 г.

модели может привести к большим и непредсказуемым отклонениям с недопустимыми доверительными границами. В таком случае для зависимости доза-эффект становится разумным использовать непараметрический подход, который состоит в следующем. Имеется модель бинарных откликов, которая носит условное название *зависимость доза-эффект* [3], [7]. Именно, пусть $\{(X_i, U_i), 1 \leqslant i \leqslant n\}$ — потенциальная повторная выборка из неизвестного распределения

$$F(x)G(x), \qquad F(x) = \mathbf{P}(X_i < x), \qquad G(x) = \mathbf{P}(U_i < x),$$

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 $(X_i < U_i)$. Задача — оценить неизвестную ф.р. $F(x)$ по выборке $\mathcal{U}^{(n)}$. Здесь U_i рассматриваются как дозы, а W_i — как эффект от воздействия дозы U_i. Пусть

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \qquad G(x) = \int_{-\infty}^{x} g(t)\,dt, \qquad \text{причем} \quad f(x) > 0,\ g(x) > 0.$$

Такую ситуацию будем называть *случайным* планом эксперимента. Тогда условное математическое ожидание будет равно

$$\mathbf{E}(W|U = x) = \mathbf{P}(X < U|U = x) = \mathbf{P}(X < x|U = x) = \mathbf{P}(X < x) = F(x),$$

т.е. неизвестная функция распределения $F(x)$ является регрессией и для оценки $F(x)$ по выборке $\mathcal{U}^{(n)} = \{(U_i, W_i),\ 1 \leqslant i \leqslant n\}$ можно использовать ядерные оценки регрессии.

Наряду со случайным планом будем рассматривать *фиксированные* планы эксперимента [8]. Именно, будем полагать вводимую дозу U неслучайной и положим $U_i = u_i$, $i = 0, 1, ..., n+1$, где $0 = u_0 < u_1 < ... < u_n < u_{n+1} = 1$. В статье мы будем изучать поведение ядерных оценок по фиксированным планам.

Для зависимости доза-эффект со случайными планами эксперимента и бинарными откликами [3], [7] в качестве оценки функции распределения $F(x)$ обычно берется статистика

$$F_n(x) = \frac{S_{2n}(x)}{S_{1n}(x)}, \qquad S_{1n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - U_i), \qquad S_{2n}(x) = \frac{1}{n}\sum_{i=1}^{n} W_i K_h(x - U_i),$$

если $S_{1n}(x) \neq 0$, где $K_h(x) = K(x/h)/h$, $K(x)$ — финитная симметричная плотность распределения (ядро), $h = h(n) \to 0$, $nh \to \infty$ при $n \to \infty$. Если $S_{1n}(x) = 0$, то $F_n(x)$ полагается равным нулю. Так, в качестве ядерной функции $K(x)$ часто используют ядро Епанечникова

$$K_1(x) = (3/4)(1 - x^2)I(|\,x\,| < 1),$$

а также квартическое ядро

$$K_2(x) = (15/16)(1 - x^2)^2 I(|\,x\,| < 1),$$

в качестве $h(n)$ берут $n^{-1/5}$.

При некоторых условиях регулярности (см. [7]), оказывается, что при $n \to \infty$ величина $n^{2/5}(F_n(x) - \mathbf{E}(F_n(x)))$ асимптотически нормальна $N(0, \sigma^2(x))$, где

$$\sigma^2(x) = F(x)(1 - F(x))\|\,K\,\|^2/g(x), \qquad \|\,K\,\|^2 = \int_{-\infty}^{\infty} K^2(x)\,dx.$$

Для фиксированных планов эксперимента предельная дисперсия оценки $F_n(x)$ равна

$$\sigma_1^2(x) = F(x)(1 - F(x))\|\,K\,\|^2.$$

Зависимость доза-эффект в модели с *биномиальной регрессией* (см., например, [9], [10]), можно описать следующим образом. Предположим, что ответ W_{ij} равен 1, если он дает интересующую реакцию, или $W_{ij} = 0$, если реакции нет, которая наблюдается на каждой фиксированной коварриате u_i.

Таким образом, W_{ij} есть j-й ответ m субъектов, когда ковариата равна u_i, $i = 1, 2, ..., n$, и ответы W_{ij} являются взаимно независимыми. Взаимосвязь определяется вероятностью того, что $W_{ij} = 1$ при условии u_i:

$$F(u_i) = \mathbf{P}(W_{ij} = 1) = \mathbf{P}(X_{ij} < u_i),$$

где $W_i = \sum_{j=1}^{m} W_{ij}$ имеет биномиальное распределение $B(m, p_i)$ с параметром $p_i = F(u_i)$, а хорошо известно, что максимальное правдоподобие для p_i дается отношением $w_i = W_i/m$ для каждого i. Данные (u_i, w_i), $i = 1, 2, ..., n$, позволяют построить оценку ф. р. вида

$$F_n(x) = \frac{\sum_{j=1}^{n} w_i \eta_i(x)}{\sum_{j=1}^{n} \eta_i(x)}, \qquad \text{где} \quad \eta_i(x) = K_h(x - u_i).$$

Для $m = 1$ мы имеем регрессионную модель Бернулли. В [11] показано, что при фиксированном x, разность $\sqrt{nh}(F_n(x) - \mathbf{E}(F_n(x)))$ асимптотически нормальна $N(0, \sigma_1^2(x)/m)$ при $n \to \infty$. При этом в каждом слое мы рассматриваем повторную выборку, в которой параметр $p_i = F(u_i)$ биномиального распределения, неслучаен. Для бесповторной выборки можно считать, что сам параметр биномиального распределения случаен, например, имеет бета распределение $\mathbf{B}(\alpha, \beta)$. В этом случае мы получим бета-биномиальное распределение, тогда вместо параметра p_i мы будем иметь параметр $\alpha_i/(\alpha_i + \beta_i)$ и в качестве его оценки будем брать $m_{1,i}/W_i$ как МП-оценку «средней» вероятности и «средней» функции распределения.

В представленном здесь сообщении, тезисы которого опубликованы в [12], рассматривается отрицательная биномиальная регрессионная модель (NBR-модель). Точнее, для заданного m рассматривается отрицательное биномиальное распределение величин Z_i при заданной ковариате u_i:

$$\mathbf{P}(Z_i = k) = \frac{\Gamma(k+m)}{\Gamma(k+1)\Gamma(m)} p_i^m (1 - p_i)^{k-m}, \quad k = m, m+1, ..., \ \Gamma(k+1) = k!\,.$$

Мы используем выборку $\mathcal{Z} = \{(z_i, u_i), i = 1, 2, ..., n\}$ для определения оценки $F(x)$ вида

$$T_n(x) = \frac{\sum_{j=1}^{n} m \eta_i(x)}{\sum_{j=1}^{n} z_i \eta_i(x)}. \tag{1.1}$$

Для так называемых квази-случайных последовательностей $\{u_i\}, i = 1, 2, ..., n$ с медленным расхождением (*low-discrepancy sequences*), где u_i — неслучайны, мы докажем состоятельность и асимптотическую нормальность построенных оценок при $n \to \infty$. Мы показываем, что предельная дисперсия оценок при нормировке $\sqrt{nh}$ равна

$$\sigma_2^2(x) = F^2(x)(1 - F(x)) \| K \|^2/m,$$

которая меньше предельной дисперсии

$$\sigma_1^2(x) = F(x)(1 - F(x)) \| K \|^2/m$$

оценок функции распределения $F(x)$ типа Надарая-Ватсона в биномиальной регрессии. На основе статистики (1.1) строятся оценки квантилей и доказывается их асимптотическая нормальность. На базе несмещенной оценки $\hat{p}$ параметра p, а также оценки максимального правдоподобия отрицательного биномиального распределения мы предлагаем оценку неизвестной функции распределения.

Отрицательное биномиальное распределение возникает естественным путем при малых значениях параметров биномиального распределения p_i и больших m, которое можно аппроксимировать распределением Пуассона. Известно, что смесь пуассоновского и Гамма-распределения (см. [13, с. 184]) приводит к отрицательому биномиальному распределению. Можно рассматривать также урновую схему Пойа и показать, что отрицательное биномиальное распределение можно получить предельным переходом из урновой схемы Пойа (см. [14, с. 495]).

В прикладных задачах, каковой является данная задача, приходится учитывать помимо теоретических аспектов вопросы привязки к конкретным ситуациям, в частности, выбор ковариат u_i. Их можно выбирать детерминированно с равномерным шагом, можно построить чисто случайные конструкции, можно, используя квази-метод Монте-Карло, отбирать их из заданного множества случайным образом. При удачном выборе множества удается получить почти оптимальные результаты. Здесь мы предлагаем рассматривать почти равномерные последовательности ковариат.

2. Основные условия

Пусть $\{X_i,\ i = 1, \dots, n\}$ — последовательность независимых одинаково распределенных с X на отрезке $[0, 1]$ случайных величин с функцией распределения $F(x)$, а

$$P = \{u_0, u_1,\ \dots,\ u_n, u_{n+1}\} \text{ — разбиение отрезка } [0,1], u_0 = 0 < u_1 < \ldots < u_n < 1 = u_{n+1}.$$

Будем предполагать, что выполнены следующие условия.

Условие 1. *При $n \to \infty$ ширина окна $h = n^{-1/5}$.*

Условие (1) будем записывать, как условие (**H**).

Условие 2. $K(x) \geqslant 0$, *причем* $K(x) = 0, x \notin [-1, 1]$.

Условие 3. $\int\limits_{-1}^{1} K(x)\,dx = 1$.

Условие 4. $K(x) = K(-x), x \in \boldsymbol{R}$.

Условие 5. *Существуют третьи непрерывные ограниченные производные функции $K(x)$ на отрезке* $[-1, 1]$.

Условие 6. $\| K \|_\infty = \sup\limits_{-1 \leqslant x \leqslant 1} |\, K(x) \,| = k_j < \infty$.

Положим

$$\| K \|^2 = \int_{-1}^{1} K^2(x)\,dx$$

и определим *вариацию* функции $f = f(x)$, $a \leqslant x \leqslant b$ (см. [15, с. 234]).

Пусть $g : [a, b] \to \mathbf{R}$. *Вариацией* функции $g = g(u)$ на отрезке $[a, b]$ называется следующая величина:

$$\bigvee(g) = \bigvee_a^b(g) = \sup_P \sum_{k=0}^{m} |\, g(u_{k+1}) - g(u_k) \,|,$$

т.е. точная верхняя грань по всем разбиениям P отрезка $[a, b]$.

Условие 7. *Вариация функции $K(x)$ ограничена, т.е.* $\bigvee(K) < \infty$.

Заметим, что если $K(x)$ — гладкая функция, то $\bigvee_0^1(K) = \int_0^1 |\, K'(x) \,|\, dx$.

В дальнейшем условия (2-7) будем записывать, как условие (**K**).

Условие 8. *Существует третья непрерывная ограниченная производная плотности распределения $f(x) = F'(x)$, причем $f(x) > 0$.*

Условие (8) будем записывать, как условие (**F**).

В работе будем предполагать, что выполнены условия (**H**), (**K**), (**F**), которые будем называть *условиями регулярности*.

3. Вспомогательные результаты

В этом параграфе представлены вспомогательные результаты, необходимые для изучения асимптотики введенных оценок.

Приведем неравенство Koksma-Hlawka (см. [16, с. 18]), которое позволяет оценить скорость сходимости интегральных сумм к соответствующему интегралу.

Пусть $\mathcal{B}$ — лебегова σ-алгебра на I^s, где $I = [0,1]$, и ρ_s — лебегова мера на $\mathcal{B}$, а P — множество точек $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N \in I^s$. Определим счетчик

$$A_n(B; P) = \sum_{i=1}^{n} I_B(u_i)$$

и отклонение

$$D_n(\mathcal{B}; P) = \sup_{B\in\mathcal{B}} \left| \frac{A_n(B; P)}{n} - \rho_s(B) \right|,$$

где $I_B(x)$ — индикатор множества B. Положим $D_n^*(P) = D_n(J_c^*; P)$, где J_c^* есть семейство подинтервалов на I^s вида $\prod\limits_{i=1}^{s}[0, u_i]$. Здесь $\rho_s\left(\prod\limits_{i=1}^{s}[0, u_i]\right) = u_1u_2...u_s$. Величину $D_n^*(P)$ называют *дискрепансом* (discrepancy) последовательности.

Определение 3.1. *Говорят, что последовательность $P = \{u_1, u_2, ...\}$ действительных чисел равномерно распределена (р.р.), если для любой пары действительных чисел $0 \leqslant a < b \leqslant 1$, имеем*

$$\lim_{n\to\infty} \frac{A_n([a, b), P)}{n} = b - a.$$

Мы будем иметь дело с р.р. последовательностями.

Теорема 3.1 ([16])**.** *(Неравенство Koksma-Hlawka). Если функция $f(u)$ $(0 \leqslant u \leqslant 1)$ есть непрерывная функция и имеет ограниченную вариацию $\bigvee(f)$ на $[0,1]$, то для любых $u_1, u_2, ..., u_n \in [0,1]$, мы имеем*

$$\left| \frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int f(u)\, du \right| \leqslant \bigvee(f) D_n^*(u_1, ..., u_n).$$

Для многомерного единичного куба $I^s = [0,1)^s, \bar{I}^s = [0,1]^s$ и вариации в смысле Харди и Краузе (см. [16, с. 19]) имеет место следующий результат.

Теорема 3.2. [16, с. 20] *Если функция $f(u)$ $(0 \leqslant u \leqslant 1)$ имеет ограниченную вариацию $\bigvee(f)$ на $\bar{I}^s$ в смысле Харди и Краузе, то для любых $\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_n \in I^s$, мы имеем*

$$\left| \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{u}_i) - \int f(\boldsymbol{u})\, \boldsymbol{du} \right| \leqslant \bigvee(f) D_n^*(\boldsymbol{u}_1, ..., \boldsymbol{u}_n).$$

В [16] показано, что $D_n^*(u_1, ..., u_n)$ есть непрерывная функция переменных $(u_1, ..., u_n)$ и что если $u_i = \dfrac{i}{n}$, то $D_n^*(u_1, ..., u_n) = \dfrac{1}{n}$. Аналогично можно показать, что для равномерного разбиения s-мерного единичного куба I^s дискрепанс конечного числа точек есть $D_n^* \asymp \dfrac{1}{n}$ (см. [17]). При конечных n его также можно вычислить используя алгоритм, данный в [18].

Замечание 3.1. *В одномерном случае если $u_0 = 0 < u_1 < u_2 < ... < u_n < u_{n+1} = 1$, то*

$$\begin{aligned} D_n^*(P) &= \max_{0\leqslant k\leqslant n} \sup_{u_k<u\leqslant u_{k+1}} \left| \frac{A_n([0,u); P)}{n} - u \right| = \max_{0\leqslant k\leqslant n} \sup_{u_k<u\leqslant u_{k+1}} \left| \frac{k}{n} - u \right| \\ &= \max_{0\leqslant k\leqslant n} \max\left(\left| \frac{k}{n} - x_k \right|, \left| \frac{k}{n} - x_{k+1} \right| \right), \end{aligned}$$

а это есть статистика Колмогорова.

Таким образом, если $u_1 < u_2 < ... < u_n$ рассматривать как вариационный ряд выборки и в качестве нулевой гипотезы взять равномерное распределение, то дискрепанс является максимальным уклонением выборочной функции распределения от равномерной. Большие значения D_n^* говорят о кучности последовательности P в некоторой области.

Этот результат обобщается и на многомерный случай [19]. В этом случае справедлив закон повторного логарифма, доказанный Кифером [20]: почти всюду

$$\overline{\lim_{n\to\infty}}\sqrt{2\pi}D_n^*(P)/\sqrt{2\ln\ln n} = 1\,.$$

Чтобы расширить возможные применения приводимых в п.4 результатов, рассмотрим *последовательности с низким расхождением* (*low-discrepancy sequences*) [16, Гл. 3] — *Ван дер Корпута и Холтона последовательности* [21].

Пусть

$$n = \sum_{j=0}^{\infty} a_j(n) b^j \tag{3.1}$$

— представление целого числа $n \geqslant 0$ по натуральному основанию $b \geqslant 2$, где $a_j(n) \in Z_b = \{0, 1, ..., b-1\}$ для каждого $j \geqslant 0$ и $a_j(n) = 0$ для всех достаточно больших j.

Определение 3.2. *Для $b \geqslant 2$, радикально-обратная функция ϕ_b по основанию b определяется следующим образом:*

$$\phi_b(n) = \sum_{j=0}^{\infty} a_j(n) b^{-j-1} \quad \text{для любого целого } n \geqslant 0, \tag{3.2}$$

где $a_j(n)$ берутся из представления (3.1) с тем же b.

Определение 3.3. *Для любого натурального $b \geqslant 2$, последовательность Ван дер Корпута по основанию b есть последовательность $\{u_0, u_1, ..., u_n, ...\}$ с $u_n = \phi_b(n)$ для любого $n \geqslant 0$.*

Пусть дана последовательность $S = \{u_0, u_1, ...\}$. Мы будем писать $D_n(S) = D_n(u_0, u_1, ..., u_{n+1})$ для расхождения первых n членов S и аналогично будем писать $D_n^*(S) = D_n^*(u_0, u_1, ..., u_n)$.

В [16] показано, что если S_b есть последовательность Ван дер Корпута по основанию b, то

$$D_N^*(S_b) \leqslant C_1 \frac{\ln N}{N}, \quad \text{для всех } N \geqslant 2,$$

где константа C_1 зависит только от b.

Определение 3.4. *Пусть s — произвольная размерность, а $b_1, b_2, ..., b_s \geqslant 2$, — взаимно простые натуральные числа. Определим последовательность Холтона, полагая*

$$\boldsymbol{u}(n) = (\phi_{b_1}(n), \phi_{b_2}(n), ..., \phi_{b_s}(n)) \in I^s \quad \text{для любого} \quad n \geqslant 0.$$

При $s = 1$ это определение сводится к определению последовательности Ван дер Корпута.

Теорема 3.3 ([16]). *Если S есть Холтона последовательность, то существуют константы C_2 и C_3, зависящие только от $b_1, b_2, ..., b_s$ такие, что для всех $N \geqslant 1$,*

$$C_2 \frac{(\ln N)^{s-1}}{N} \leqslant D_N^*(S) \leqslant C_2 \frac{(\ln N)^s}{N}.$$

В [17, с. 166] замечено, что при $s = 1$ наилучшими являются равномерные сетки, но с увеличением s они приближаются к наихудшим. Там показано, как надо изменить сетку, чтобы она стала лучше. В двумерном случае также известна явная квадратурная формула Фибоначчи [22, с. 92]:

$$\frac{1}{b_n}\sum_{k=1}^{b_n} f\left(\frac{k}{b_n}, \left\{\frac{b_{n-1}}{b_n}k\right\}\right), \quad b_1 = b_2 = 1, \quad b_n = b_{n-1} + b_{n-2},\ (n \geqslant 3),$$

где $\{a\}$ — дробная часть числа a. Кроме того, в [23, с. 92], приведена оценка сверху для следующих квадратурных формул:

$$\sup_{f\in H_2^r}\left|\int_{[0,1]^2} f(x_1,x_2)\,dx_1dx_2 = \frac{1}{A}\sum_{k=1}^{A} f\left(\frac{k}{A},\left\{\frac{b}{A}k\right\}\right)\right| \leqslant C\frac{1+\ln A}{A^r},$$

где $r>1$, A и b $(1<b<A)$ — взаимно простые целые, а функция $f(x_1,x_2)$ принадлежит классу H_2^r, если в единичном кубе $\bar{I}^s$ она имеет непрерывные производные вида

$$\frac{\partial^k}{\partial x_1^{k_1}...\partial x_s^{k_s}} \quad (0\leqslant k\leqslant rs,\ 0\leqslant k_\nu\leqslant r).$$

Нам нужна будет также теорема об асимптотическом поведении функций от оценок.

Теорема 3.4. [24, с. 86] *(Дельта-метод). Если $\varphi(n)\to\infty$ при $n\to\infty$, и*

$$\varphi(n)(T_n-\theta)\xrightarrow[n\to\infty]{d} N(0,\tau^2)$$

то

$$\varphi(n)(g(T_n)-g(\theta))\xrightarrow[n\to\infty]{d} N(0,\tau^2(g'(\theta))^2).$$

при условии, что существует непрерывная не равная нулю производная $g'(\theta)$, функции $g(\theta)$.

4. Основные результаты

4.1. NBR-оценки. Асимптотическое поведение. Пусть дана выборка $\mathcal{Z}^{(n)}=\{(z_i,u_i),\ i=1,2,...,n\}$, где z_i имеет отрицательное биномиальное распределение NB$(m,F(u_i))$, а последовательность u_i, $i=1,2,...,n$ есть последовательность Ван дер Корпута. Определим статистику

$$T_n(x)=\frac{\sum\limits_{i=1}^{n} m\eta_i(x)}{\sum\limits_{i=1}^{n} z_i\eta_i(x)},\quad \text{где}\quad \eta_i(x)=K_h(u_i-x).$$

Поскольку $\dfrac{1}{n}\sum\limits_{i=1}^{n}\eta_i(x)\to 1$ при $n\to\infty$, то мы рассмотрим оценку

$$\hat{F}_n(x)=\frac{m}{\frac{1}{n}\sum\limits_{i=1}^{n} z_i\eta_i(x)}. \tag{4.1}$$

Обозначим

$$S_1=\frac{1}{n}\sum_{i=1}^{n} z_i\eta_i(x),\qquad \nu_j(K)=\int\limits_{-1}^{1} t^jK(t)\,dt, j\in\mathbf{N}.$$

Теорема 4.1. *Пусть $\hat{F}_n(x)$ — оценка функции распределения $F(x)$, определенная формулой (4.1), $\{u_i, i=1,2,...,n\}$ — последовательность Ван дер Корпута, выполнены условия регулярности. Тогда*

$$\sqrt{nh}(F_n(x)-\boldsymbol{E}(F_n(x))\xrightarrow[n\to\infty]{d} N\left(0,\frac{(1-F(x))F^2(x)}{m}\|\,K\,\|^2\right).$$

Доказательство. Из [25, леммы 3.4] следует, что $\bigvee(K_h) = O(h^{-1})$, поэтому при $n \to \infty$ мы имеем:

$$\begin{aligned}
\mathbf{E}(S_1) = \mathbf{E}\left(\frac{1}{nh}\sum_{i=1}^{n} Z_i K\left(\frac{u_i - x}{h}\right)\right) &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{u_i - x}{h}\right)\mathbf{E}(Z_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{u_i - x}{h}\right)\frac{m}{F(u_i)} \\
&= m \cdot \frac{1}{h}\int_{-\infty}^{\infty} \frac{1}{F(u)} K\left(\frac{u-x}{h}\right) du + O\left(\frac{\ln n}{nh}\right) \\
&= m\int_{-x/h}^{(1-x)/h} \frac{K(t)}{F(x+ht)}\, dt + O\left(\frac{1}{n}\right) \\
&= m\int_{-\infty}^{\infty} \frac{K(t)}{F(x+ht)}\, dt + O\left(\frac{\ln n}{nh}\right) \\
&= m\int_{-\infty}^{\infty} \left(\frac{K(t)}{F(x)} + K(t)\left(\frac{1}{F(x+ht)} - \frac{1}{F(x)}\right)\right) dt + O\left(\frac{\ln n}{nh}\right) \\
&= \frac{m}{F(x)} + m\int_{-\infty}^{\infty} \frac{F(x) - F(x+ht)}{F(x)F(x+ht)} K(t)\, dt + O\left(\frac{\ln n}{nh}\right).
\end{aligned}$$

Теперь

$$\begin{aligned}
\gamma_n &= \int_{-\infty}^{\infty} \frac{F(x+ht) - F(x)}{F(x)F(x+ht)} K(t)\, dt \\
&= \int_{-\infty}^{\infty} \frac{f(x)ht + (1/2)f'(x)h^2t^2 + (1/6)f''(x)h^3t^3 + (1/24)f'''(\zeta)h^4t^4}{F(x)F(x+ht)} K(t)\, dt,
\end{aligned}$$

где ζ — некоторая «средняя» точка. Так как

$$\begin{aligned}
\left| h^4 \int\limits_{-1}^{1} \frac{f'''(\zeta)}{F(x)F(x+ht)} h^4 t^4 K(t)\, dt \right| &\leqslant h^4 \nu_4(K) \sup_{-1\leqslant t\leqslant 1} \frac{f'''(\zeta)}{F(x)F(x+ht)} \\
&\leqslant \frac{2C_3 h^4}{F(x)(F(x)-\varepsilon)} \quad \text{для} \quad n \geqslant n_1,
\end{aligned}$$

то

$$\begin{aligned}
\gamma_n &= \frac{1}{F^2(x)} \int_{-\infty}^{\infty} \frac{f(x)ht + (1/2)f'(x)h^2t^2}{(1 + (f(x)/F(x))ht) + O(h^2))} K(t)dt + O(h^4) \\
&= \frac{1}{F^2(x)} \int_{-\infty}^{\infty} \left(f(x)ht + \frac{1}{2}f'(x)h^2t^2 - \frac{f^2(x)}{F(x)}h^2t^2 - \frac{1}{2}\frac{f'(x)f(x)}{F(x)}h^3t^3 \right) K(t)dt + O(h^3) \\
&= \frac{1}{F^2(x)} \int_{-\infty}^{\infty} \left(f(x)ht + \frac{1}{2}f'(x)h^2t^2 - \frac{f^2(x)}{F(x)}h^2t^2 \right) K(t)dt + o(h^2) \\
&= \left(\frac{1}{2}\frac{f'(x)}{F^2(x)} - \frac{f^2(x)}{F^3(x)} \right) h^2 \nu_2(K) + o(h^2),
\end{aligned}$$

поэтому

$$\mathbf{E}(S_1) = \frac{m}{F(x)} - m\left(\frac{1}{2}\frac{f'(x)}{F^2(x)} - \frac{f^2(x)}{F^3(x)}\right) h^2\nu_2(K) + o(h^2), \qquad \nu_2(K) = \int_{-\infty}^{\infty} t^2 K(t)\, dt.$$

Рассмотрим теперь дисперсию статистики S_1. Имеем:

$$\begin{aligned}\mathbf{D}(S_1)=\mathbf{D}\left(\frac{1}{nh}\sum_{i=1}^{n}Z_iK\left(\frac{u_i-x}{h}\right)\right)&=\frac{1}{n^2h^2}\sum_{i=1}^{n}K^2\left(\frac{u_i-x}{h}\right)\mathbf{D}(Z_i)\\&=\frac{1}{n^2h^2}\sum_{i=1}^{n}K^2\left(\frac{u_i-x}{h}\right)\frac{m(1-F(u_i))}{F^2(u_i)}\\&\sim\frac{m}{nh^2}\int_{-\infty}^{\infty}\frac{1-F(u)}{F^2(u)}K^2\left(\frac{u-x}{h}\right)du\\&=\frac{m}{nh}\int_{-\infty}^{\infty}\frac{1-F(x+ht)}{F^2(x+ht)}K^2(t)\,dt\\&\sim\frac{m(1-F(x))}{nh\,F^2(x)}\|\,K\,\|^2.\end{aligned}$$

Чтобы доказать асимптотическую нормальность статистик S_1 проверим *условие Ляпунова*, для чего нам понадобится при $a>1$ следующее неравенство:

$$|\,x+y\,|^a\leqslant 2^{a-1}(\,|\,x\,|^a+|\,y\,|^a),\tag{4.2}$$

которое есть следствие того, что функция $|\,x\,|^a$ выпукла при $a>1$, значит

$$\left|\,\frac{x+y}{2}\,\right|^a\leqslant\frac{|\,x\,|^a+|\,y\,|^a}{2}\,.$$

Пусть

$$\xi_j=\frac{1}{nh}Z_jK\left(\frac{u_i-x}{h}\right).$$

Тогда $S_1=\sum\limits_{j=1}^{n}\xi_j$. Используя неравенство (4.2) при $a=4$, имеем

$$|\,\xi_j-\mathbf{E}(\xi_j)\,|^4\leqslant 8(|\,\xi_j\,|^4+|\,\mathbf{E}(\xi_j)\,|^4)\,.$$

Беря от обеих частей математическое ожидание, получаем

$$\mathbf{E}((\xi_j-\mathbf{E}(\xi_j))^4)\leqslant 8(\mathbf{E}(\xi_j^4)+(\mathbf{E}(\xi_j))^4\leqslant 16\mathbf{E}(\xi_j^4)\,.$$

Рассмотрим $A_n=\sum\limits_{j=1}^{n}\mathbf{E}(\xi_j^4)$. Имеем:

$$A_n=\frac{1}{n^4h^4}\sum_{j=1}^{n}\mathbf{E}(Z_j^4)K^4\left(\frac{u_j-x}{h}\right)\,.$$

Заметим, что если с.в. Z имеет отрицательное биномиальное распределение с параметрами m и $p=1-q$, то ее характеристическая функция равна

$$\varphi(t)=\left(\frac{p\cdot\exp(it)}{1-q\cdot\exp(it)}\right)^m,$$

дисперсия равна $\mathbf{D}(Z)=mq/p^2$, а четвертый начальный момент равен

$$\mathbf{E}(Z^4)=b_4\frac{q^4}{p^4}+b_3\frac{q^3}{p^3}+b_2\frac{q^2}{p^2}+b_1\frac{q}{p},$$

где

$$b_4=m(m^3+6m^2+11m+6),\quad b_3=6m(m^2+3m+2),\quad b_2=7m(m+1),\quad b_1=m.$$

С учетом этого замечания выводим, что

$$A_n \sim \frac{1}{n^4h^4}\sum_{j=1}^{n}\left(b_4\frac{(1-F(u_j))^4}{F^4(u_j)}+b_3\frac{(1-F(u_j))^3}{F^3(u_j)}+b_2\frac{(1-F(u_j))^2}{F^2(u_j)}+b_1\frac{1-F(u_j)}{F(u_j)}\right)K^4\left(\frac{u_j-x}{h}\right)$$
$$\sim \frac{1}{n^3h^3}\left(b_4\frac{(1-F(x))^4}{F^4(x)}+b_3\frac{(1-F(x))^3}{F^3(x)}+b_2\frac{(1-F(x))^2}{F^2(x)}+b_1\frac{1-F(x)}{F(x)}\right)\int_{-\infty}^{\infty}K^4(t)\,dt=\frac{C_1}{n^3h^3},$$

где C_1 — универсальная константа.

Значит, для дроби Ляпунова,

$$L_n=\frac{\sum_{j=1}^{n}\mathbf{E}((\xi_j-\mathbf{E}(\xi_j))^4)}{(\sum_{j=1}^{n}\mathbf{D}(\xi_j))^2}\leqslant\frac{C_1n^2h^2}{C_2n^3h^3}=\frac{C}{nh}\underset{n\to\infty}{\longrightarrow}0,$$

т.е. условие центральной предельной теоремы Ляпунова [26, с. 241], выполнено и

$$\frac{S_1-\mathbf{E}(S_1)}{\sqrt{\mathbf{D}(S_1)}}\xrightarrow[n\to\infty]{d}N(0,1).$$

Иными словами,

$$\sqrt{nh}\left(S_1-\frac{m}{F(x)}\right)\xrightarrow[n\to\infty]{d}N\left(m\left(\frac{f^2(x)}{F^3(x)}-\frac{1}{2}\frac{f'(x)}{F^2(x)}\right)\nu_2(K),\frac{m(1-F(x))}{F^2(x)}\|\,K\,\|^2\right).$$

Теперь рассмотрим асимптотическое поведение статистики $T=\dfrac{m}{S_1}$:

$$T=T_n(x)=\frac{\frac{1}{nh}\sum_{i=1}^{n}m\eta_i(x)}{\frac{1}{nh}\sum_{i=1}^{n}z_i\eta_i(x)}\sim\frac{m}{\frac{1}{nh}\sum_{i=1}^{n}z_i\eta_i(x)}=\frac{m}{S_{1n}(x)}=\frac{m}{S_1}.$$

Здесь величины z_i имеют отрицательное биномиальное распределение с соответствующими параметрами. Поэтому, используя дельта-метод, получаем

$$g(x)=\frac{m}{x},\quad g'(x)=-\frac{m}{x^2},\quad g'\left(\frac{m}{F(x)}\right)=-\frac{F^2(x)}{m},\quad\left(g'\left(\frac{m}{F(x)}\right)\right)^2=\frac{F^4(x)}{m^2}.$$

Для оценки $F_n(x)=\dfrac{m}{S_1}$, выполнено соотношение $F_n(n)\xrightarrow[n\to\infty]{p}F(x)$, поэтому

$$g(\theta_n)=g(\theta_0)+(\theta_n-\theta_0)g'(\theta_0)+O((\theta_n-\theta_0)^2)\Rightarrow g(\theta_n)-g(\theta_0)=(\theta_n-\theta_0)g'(\theta_0),$$
$$\sqrt{nh}(g(\theta_n)-g(\theta_0))\sim\sqrt{nh}(\theta_n-\theta_0)g'(\theta_0)\sim N\left(a,(g'(\theta_0))^2m\|\,K\,\|^2\frac{1-F(x)}{F^2(x)}\right).$$

Но

$$(g'(\theta_0))^2m\|\,K\,\|^2\frac{1-F(x)}{F^2(x)}=m\|\,K\,\|^2\frac{1-F(x)}{F^2(x)}\cdot\frac{F^4(x)}{m^2}=\frac{(1-F(x))F^2(x)}{m}\|\,K\,\|^2,$$

откуда

$$\sqrt{nh}(F_n(x)-\mathbf{E}(F_n(x))\xrightarrow[n\to\infty]{d}N\left(0,\frac{(1-F(x))F^2(x)}{m}\|\,K\,\|^2\right).$$

□

Замечание 4.1. *В оценке $\hat{F}_n(x)$ вместо статистики S_1 можно использовать статистику (см. [27])*

$$S_1^{PC}(x)=\sum_{i=1}^{n-1}(u_{i+1}-u_i)z_i\eta_i(x),$$

которая также будет асимптотически нормальна с теми же параметрами, что и у S_1.

Замечание 4.2. *Так как предельная дисперсия оценки $\hat{F}_n(x)$ зависит от неизвестной функции распределения $F(x)$ и, следовательно, неизвестна, то для ее оценки можно использовать статистику*

$$\hat{\sigma}^2(x) = \frac{m^2}{2n} \sum_{i=1}^{n-1} \frac{(z_{i+1} - z_i)^2}{S_1^4(x)} K_h(u_{i+1} - x) K_h(u_i - x),$$

которая является состоятельной оценкой функции

$$\frac{m(1 - F(x))}{F^2(x)} \| K \|^2.$$

Замечание 4.3. *На основе несмещенной оценки параметра p*

$$\hat{p} = \frac{m-1}{m+z-1}$$

отрицательного биномиального распределения $NB(m,p)$ (см. [28, с. 230]*) предложим еще одну оценку функции распределения $F(x)$ вида ($m \geqslant 2$)*

$$\hat{F}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \frac{m-1}{m+z_i-1} K\left(\frac{u_i - x}{h}\right).$$

Оценка $\hat{p}$ является несмещенной оценкой, нижняя граница Крамера-Рао для ее дисперсии имеет вид

$$\mathbf{D}(\hat{p}) \geqslant \frac{p^2 q}{m}.$$

Найдем сначала второй начальный момент, а потом дисперсию. По определению

$$\begin{aligned}
\mathbf{E}(\hat{p}^2) = \sum_{k=0}^{\infty} \frac{(m-1)^2}{(m+k-1)^2} \frac{\Gamma(m+k)}{k!\,\Gamma(m)} p^m q^k &= p^m \sum_{k=0}^{\infty} \frac{m-1}{m+k-1} \frac{\Gamma(m+k-1)}{k!\,\Gamma(m-1)} q^k \\
&= p^m {}_2F_1(m-1, m-1; m; q) \\
&= (m-1)p^m \int_0^1 \frac{t^{m-1}}{(1-tq)^{m-1}\, dt},
\end{aligned}$$

где ${}_2F_1(a,b;c;x)$ — гипергеометрическая функция Гаусса.

В таком случае,

$$\mathbf{D}(\hat{p}) = p^m\, {}_2F_1(m-1, m-1; m; q) - p^2.$$

Если $m = 2$, то

$$\mathbf{D}(\hat{p}) = -p^2 \left(1 + \frac{\ln p}{q}\right) \geqslant \frac{p^2 q}{2}$$

и при малых значениях q левая и правая части близки. Предельная дисперсия оценки $\hat{F}_n(x)$ будет равна

$$\sigma^2 = -F^2(x) \left(1 + \frac{\ln F(x)}{1 - F(x)}\right).$$

Если $m = 3$, то

$$\sigma_3^2 = \mathbf{D}(\hat{p}) = p^2 \left(\frac{2p \ln p}{q^2} + \frac{1+p}{q}\right) \geqslant \sigma_0^2 = \frac{p^2 q}{3}, \quad \text{т.к. } {}_2F_1(2,2;3;x) = \frac{2\ln(1-x)}{x^2} + \frac{2}{x(1-x)}$$

и при малых значениях q левая и правая части, т.е. σ_3^2 и σ_0^2, также будут близки. Отметим, что

$$\mathbf{E}(\hat{p}^2) = \frac{(m-1)p^m}{q^{m-1}} \left[(-1)^{m-1} \ln p + \sum_{k=1}^{m-2} \frac{(-1)m-k}{k} \left(\frac{q}{p}\right)\right].$$

Последнее соотношение впервые было получено в работе [29].

Можно также построить ядерную оценку функции распределения отталкиваясь от оценки максимального правдоподобия, которая равна

$$\tilde{p} = \frac{m}{m+z}.$$

В этом случае

$$\mathbf{E}(\tilde{p}) = \sum_{k=0}^{\infty} \frac{m}{m+k} \cdot \frac{\Gamma(m+k)}{k!\,\Gamma(m)} p^m q^k = p^m {}_2F_1(m, m; m+1; q),$$

а (см. [30, с. 565, 5.2.11 (**15**)])

$$\mathbf{E}(\tilde{p}^2) = \sum_{k=0}^{\infty} \frac{m^2}{(m+k)^2} \cdot \frac{\Gamma(m+k)}{k!\,\Gamma(m)} p^m q^k = m^2 p^m \int_0^{\infty} t e^{-mt} (1 - qe^{-t})^{-m}\, dt.$$

В частности, при $m = 1$ имеем:

$$\mathbf{E}(\tilde{p}) = -\frac{p \ln p}{q}, \qquad \mathbf{E}(\tilde{p}^2) = \frac{p}{q}\,\mathrm{dilog}(p).$$

При $m = 2$ имеем:

$$\mathbf{E}(\tilde{p}) = \frac{2p}{q}(q + p\ln p), \quad \mathbf{E}(\tilde{p}^2) = -\frac{4p^2}{q^2}(\ln p + \mathrm{dilog}(p)), \quad \text{где} \quad \mathrm{dilog}(x) = \int\limits_1^x \frac{\ln t}{1-t}\, dt.$$

Отсюда видно, что оценка максимального правдоподобия $\tilde{p}$ не является состоятельной оценкой параметра p, поэтому для $m = 2$ в качестве состоятельной оценки можно предложить статистику

$$\hat{\theta} = \frac{m}{2(m+z)} \frac{1-\hat{p}}{1+\hat{p}(\ln\hat{p} - 1)},$$

но эта оценка имеет риск больший, чем оценка (4.1).

4.2. Оценка квантиля. В данном разделе мы изучим асимптотическое поведение оценок квантилей в зависимости доза-эффект по фиксированным планам эксперимента в модели отрицательной биномиальной регрессии.

Определим оценку квантиля ξ_λ порядка $0 < \lambda < 1$ следующим образом:

$$\hat{\xi}_{n\lambda} = \inf\{x \in \mathbf{R} : \ \hat{F}_n(x) \geq \lambda\}\,. \tag{4.3}$$

Положим $a = \dfrac{(\lambda f'(\xi_\lambda) - 2f^2(\xi_\lambda))\nu_2(K)}{2\lambda\sigma}$.

В следующей теореме доказана асимптотическая нормальность оценок $\hat{\xi}_{n\lambda}$.

Теорема 4.2. *Пусть $\hat{\xi}_{n\lambda}$ — оценка квантиля порядка $0 < \lambda < 1$, определенная формулой (4.3), $\{u_i, i = 1, 2, ..., n\}$ — последовательность Ван дер Корпута, выполнены условия регулярности и $f(\xi_\lambda) > 0$. Тогда*

$$\sqrt{nh}(\hat{\xi}_{n\lambda} - \xi_\lambda - ah^2) \underset{n\to\infty}{\overset{d}{\to}} N\left(0, \frac{\lambda^2(1-\lambda)\,\|K\|^2}{mf^2(\xi_\lambda)}\right).$$

Доказательство. Пусть $\sigma^2 = \dfrac{(1-\lambda)\lambda^2}{m} \| K \|^2$, $\quad \delta = \delta(x) = \xi_\lambda + \dfrac{x\sigma}{\sqrt{nh} f(\xi_\lambda)}$.

Имеем:

$$\mathbf{P}\left(\frac{\sqrt{nh}f(\xi_\lambda))(\hat{\xi}_{n,\lambda}-\xi_\lambda)}{\sigma}\leqslant x\right)=\mathbf{P}(\hat{\xi}_{n,\lambda}\leqslant\delta)=\mathbf{P}(F_n(\delta)\geqslant\lambda)=\mathbf{P}\left(\frac{m}{\frac{1}{n}\sum_{i=1}^n z_i\eta_i(\delta)}\geqslant\lambda\right)$$

$$=\mathbf{P}\left(\frac{1}{n}\sum_{i=1}^n Z_iK_h(u_i-\delta)\leqslant\frac{m}{\lambda}\right)$$

$$=\mathbf{P}\left(\frac{1}{n}\sum_{i=1}^n (Z_iK_h(u_i-\delta)-\theta_i)\leqslant\frac{1}{n}\sum_{i=1}^n\left(\frac{m}{\lambda}-\theta_i\right)\right)$$

$$=\mathbf{P}\left(\frac{\sqrt{nh}\,\lambda^2}{m\sigma n}\sum_{i=1}^n (Z_iK_h(u_i-\delta)-\theta_i)\leqslant\frac{\sqrt{nh}\,\lambda^2}{m\sigma}\cdot\frac{1}{n}\sum_{i=1}^n\left(\frac{m}{\lambda}-\theta_i\right)\right),$$

где

$$\theta_i=\mathbf{E}\left(Z_iK_h(u_i-\delta)\right)=\frac{m}{F(u_i)}K_h(u_i-\delta).$$

Заметим, что $h^2=1/\sqrt{nh}$ и функция $K_h(u-\delta)$ равна нулю вне отрезка

$$\mathcal{J}_\lambda=[\xi_\lambda-h+xh^2\sigma/f(\xi_\lambda),\xi_\lambda+h+xh^2\sigma/f(\xi_\lambda)].$$

Кроме того, функция $1/F(u)>0$ монотонно убывает на $\mathcal{J}_\lambda$, $\{u_i,i=1,2,...,n\}$ есть последовательность Ван дер Корпута, поэтому $\bigvee_{\mathcal{J}_\lambda}(1/F(u))<\infty$ и из [16] следует, что

$$\alpha_n=\frac{1}{n}\sum_{i=1}^n\theta_i=\frac{1}{n}\sum_{i=1}^n\frac{m}{F(u_i)}K_h(u_i-\delta)=\int_{\mathcal{J}_\lambda}\frac{m}{F(u)}K_h(u-\delta)\,du+O\left(\frac{\ln n}{\sqrt{nh}}\right).$$

Делая замену

$$t=\frac{u-\delta}{h}$$

и учитывая, что $0\leqslant u\leqslant 1$, заключаем, что

$$\alpha_n=\int\limits_0^1\frac{m}{F(u)}K_h(u-\delta)\,du=\int\limits_{-\delta/h}^{(1-\delta)/h}\frac{m}{F(\xi_\lambda+\rho_1h)}\,K(t)\,dt\,,$$

где

$$\rho_1=t+\frac{x\sigma}{f(\xi_\lambda)}h$$

и для достаточно большого n $(n\geqslant n_1)$,

$$\alpha_n=\int\limits_{-1}^1\frac{m}{F(\xi_\lambda+\rho_1h)}\,K(t)\,dt\,.$$

Пусть $|\,x\,|\leqslant L$, где L — достаточно большое и $\omega_1=\omega/\lambda$. Тогда

$$F(\xi_\lambda+\rho_1h)=\lambda+f(\xi_\lambda)\rho_1h+\frac{f'(\xi_\lambda)}{2}\rho_1^2h^2+\omega h^3=\lambda(1+a_1h+b_1h^2+\omega_1h^3),$$

где

$$a_1=\frac{f(\xi_\lambda)}{\lambda}t,\qquad b_1=\frac{2x\sigma+t^2f'(\xi_\lambda)}{2\lambda}\,,$$

а из условий теоремы следует, что $|\,\omega_1\,|$ ограничена. Так как для $n\geqslant n_2$

$$\left|\frac{1}{1+a_1h+b_1h^2+\omega_1h^3}-1+a_1h-(a_1^2-b_1)h^2\right|$$

$$=\left|\frac{((b_1-a_1^2)\omega_1h^2+(a_1\omega_1+b_1^2-a_1^2b_1)h+2a_1b_1-\omega_1-a_1^3}{1+a_1h+b_1h^2+\omega_1h^3}\right|h^3\leqslant C_2h^3,$$

и $\int_{-1}^{1} tK(t)\,dt = 0$, то получаем, что

$$\alpha_n = \frac{m}{\lambda} - \frac{m\sigma}{\lambda^2}\left(x + \frac{(\lambda f'(\xi_\lambda) - 2f^2(\xi_\lambda))\nu_2(K)}{2\lambda\sigma}\right) h^2 + o(h^2).$$

Отсюда следует, что последовательность

$$\frac{\lambda^2}{m\sigma h^2}\cdot\left(\alpha_n - \frac{m}{\lambda}\right) + \left(x + \frac{(\lambda f'(\xi_\lambda) - 2f^2(\xi_\lambda))\nu_2(K)}{2\lambda\sigma}\right)$$

сходится к нулю равномерно по $|\,x\,| \leqslant L$ при $n\to\infty$, где $L > 0$ выбрано достаточно большим.

Пусть

$$\Sigma_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i K_h(u_i - \delta) - \theta_i\right).$$

Покажем, что

$$\frac{\sqrt{nh}\lambda^2}{m\sigma}\cdot\Sigma_n(x) \xrightarrow[n\to\infty]{d} N(0,1).$$

Для этого рассмотрим дисперсию величины $\Sigma_n(x)$:

$$\begin{aligned}\mathbf{D}(\Sigma_n(x)) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}\left(Z_i K_h(u_i-\delta)\right) =& \frac{1}{n^2}\sum_{i=1}^{n}K_h^2(u_i-\delta)\mathbf{D}(Z_i)\\ =& \frac{1}{n^2h^2}\sum_{i=1}^{n}K^2\left(\frac{u_i-\delta}{h}\right)\frac{m(1-F(u_i))}{F^2(u_i)}\\ =& \frac{1}{nh^2}\int_0^1 \frac{m(1-F(u))}{F^2(u)}K^2\left(\frac{u-\delta}{h}\right)du\,(1+o(1))\\ \sim& \frac{m(1-\lambda)}{nh\lambda^2}\|\,K\,\|^2\end{aligned}$$

равномерно по $|\,x\,| \leqslant L$ в последнем соотношении, откуда

$$\lim_{n\to\infty}\mathbf{D}\left(\frac{\lambda^2\sqrt{nh}}{m\sigma}\Sigma_n(x)\right) = 1\,.$$

Условия Ляпунова проверяются как при доказательстве теоремы 4.1.

Таким образом, выполнены условия центральной предельной теоремы Ляпунова [26, с. 241], поэтому для $|\,x\,| \leqslant L$

$$\frac{\lambda^2\sqrt{nh}}{m\sigma}\Sigma_n(x) \xrightarrow[n\to\infty]{d} N(0,1).$$

Осталось показать, что для любого $\varepsilon > 0$ можно выбрать $L > 0$ и $n \geqslant n_0$ так, что

$$\beta_n = \mathbf{P}\left(\frac{\sqrt{nh}f(\xi_\lambda))|\,\hat{\xi}_{n,\lambda} - \xi_\lambda\,|}{\sigma} > L\right) < \varepsilon.$$

Так как $\beta_n \leqslant \beta_{1n} + \beta_{2n}$, где

$$\beta_{1n} = \mathbf{P}\left(\frac{\sqrt{nh}f(\xi_\lambda))(\hat{\xi}_{n,\lambda} - \xi_\lambda)}{\sigma} > L\right),\quad \beta_{2n} = \mathbf{P}\left(\frac{\sqrt{nh}f(\xi_\lambda))(\hat{\xi}_{n,\lambda} - \xi_\lambda)}{\sigma} < -L\right),$$

то рассмотрим первое слагаемое. Рассуждая как выше, получим

$$\begin{aligned}\beta_{1n} = \mathbf{P}\Bigg(\frac{\sqrt{nh}\,\lambda^2}{m\sigma}\cdot\frac{1}{n}\sum_{i=1}^{n}\left(Z_iK_h(u_i-\delta(L)) - \theta_i\right) >& \frac{\sqrt{nh}\,\lambda^2}{m\sigma}\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\frac{m}{\lambda}-\theta_i\right)\Bigg)\\ =& \mathbf{P}\left(\frac{\lambda^2}{m\sigma h^2}\Sigma_n(L) > L + a\right) + o(1).\end{aligned}$$

Положим $x = L + a$ и пусть $\psi(x) = e^{tx}, t \geqslant 0$. Тогда

$$\mathbf{P}\left(\frac{\lambda^2}{m\sigma h^2}\Sigma_n(L) > x\right) \leqslant \mathbf{P}\left(\psi\left(\frac{\lambda^2}{m\sigma h^2}\Sigma_n(L)\right) > \psi(x)\right) \leqslant \frac{\mathbf{E}(\psi(\Sigma_n(L)))}{\psi(x)}.$$

Поэтому

$$\overline{\lim_{n\to\infty}} \ln \mathbf{P}\left(\frac{\lambda^2}{m\sigma h^2}\Sigma_n(L) > x\right) \leqslant -tx + \phi(t),$$

где

$$\phi(t) = \lim_{n\to\infty} \ln \mathbf{E}\left(\exp\left(\frac{t\lambda^2}{m\sigma h^2}\Sigma_n(L)\right)\right) = \frac{t^2}{2}.$$

Так как минимум функции $-tx + \phi(t)$ достигается при $t = x$ и равен $-x^2/2$, то из теоремы Gärtner-Ellis [31] следует, что

$$\lim_{n\to\infty} \beta_{1n} \leqslant \exp(-(L+a)^2/2).$$

Выберем L так, чтобы для заданного $\varepsilon > 0$ было $\exp(-(L+a)^2/2) < \varepsilon/2$. Аналогично разбирается второе слагаемое, поэтому для так выбранного L получим $\lim\limits_{n\to\infty} \beta_n < \varepsilon$. Отсюда следует результат теоремы 4.2. □

4.3. Многомерный случай. В данном разделе мы изучим асимптотическое поведение оценок двумерной функции распределения в зависимости доза-эффект по фиксированным планам эксперимента в модели отрицательной биномиальной регрессии ограничившись двумерным случаем.

Обозначим $F_{ij} = \dfrac{\partial^2}{\partial x_i \partial x_j} F(x_1, x_2),\ F_i = \dfrac{\partial}{\partial x_i} F(x_1, x_2),\ \nabla_F^T = (F_1, F_2)$,

$$\mathcal{H}_F = \begin{pmatrix} F_{11} & F_{12} \\ F_{12} & F_{22} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}, \quad \mathbf{J}^T = (1,1), \quad \mathbf{h} = \mathbf{HJ} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

Пусть $\mathcal{K}(\boldsymbol{x}) = \mathcal{K}(x_1, x_2)$ есть симметричная, финитная ограниченная, интегрируемая с квадратом плотность распределения, такая, что

$$\int \boldsymbol{x}\boldsymbol{x}^T \mathcal{K}(\boldsymbol{x})\, d\boldsymbol{x} = \nu_2(\mathcal{K})\mathbf{I}_s,$$

где $\nu_2(\mathcal{K})$ — действительное число, а $\mathbf{I}_2$ — единичная матрица порядка 2, $\mathcal{K}_{\mathbf{H}}(\boldsymbol{x}) = |\,\mathbf{H}\,|^{-1}\mathcal{K}(\mathbf{H}^{-1}\boldsymbol{x})$, $N = n_1 n_2$,

$$\mathbf{S}_1 = \mathbf{S}_1(\boldsymbol{x}) = \frac{1}{|\,\mathbf{H}\,|N} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Z_{ij}\mathcal{K}_{\mathbf{H}}(\boldsymbol{U}_{ij} - \boldsymbol{x}), \quad \hat{F}_N(\boldsymbol{x}) = \frac{m}{\mathbf{S}_1}. \tag{4.4}$$

Теорема 4.3. *Пусть $\hat{F}_N(\boldsymbol{x})$ — оценка функции распределения $F(\boldsymbol{x})$, определенная формулой (4.4), $\{\boldsymbol{u}_{ij}, i = 1, 2, ..., n_1; j = 1, 2, ..., n_2;\}$ — последовательность Холтона, выполнены условия регулярности. Тогда при $N \to \infty$,*

$$(i) \quad \boldsymbol{E}(\boldsymbol{S}_1(\boldsymbol{x})) = \frac{m}{F(\boldsymbol{x})} + \frac{m}{2F^3(\boldsymbol{x})}(2\nabla_F^T \boldsymbol{h}\boldsymbol{h}^T \nabla_F - \nu_2(\mathcal{K})\boldsymbol{h}\mathcal{H}_F \boldsymbol{h}^T)(1 + o(1));$$

$$(ii) \quad \boldsymbol{D}(\boldsymbol{S}_1(\boldsymbol{x})) = \frac{m(1 - F(\boldsymbol{x}))}{N\,|\,\boldsymbol{H}|F^2(\boldsymbol{x})} \,\|\,\mathcal{K}\,\|^2 (1 + o(1));$$

$$(iii) \quad \sqrt{N\,|\,\boldsymbol{H}|}(\hat{F}_N(\boldsymbol{x}) - \boldsymbol{E}(F_n(N)) \xrightarrow[n\to\infty]{d} N\left(0, \frac{(1 - F(\boldsymbol{x}))F^2(\boldsymbol{x})}{m}\,\|\,\mathcal{K}\,\|^2\right).$$

Доказательство. Ход доказательства аналогичен одномерному случаю, поэтому отметим отличия. Разложим функцию $F(\mathbf{x} + \mathbf{Ht})$, где $\mathbf{t}^T = (t_1, t_2)$ в ряд Тейлора:

$$\begin{aligned} F(\mathbf{x} + \mathbf{Ht}) = F(x_1 + t_1 h_1, x_2 + t_2 h_2) = &F(x_1, x_2) + \left[t_1 h_1 \frac{\partial}{\partial x_1} + t_1 h_1 \frac{\partial}{\partial x_2}\right] F(x_1, x_2) \\ &+ \frac{1}{2}\left[t_1 h_1 \frac{\partial}{\partial x_1} + t_1 h_1 \frac{\partial}{\partial x_2}\right]^2 F(x_1, x_2) + o(|\,\mathbf{H}\,|)\,. \end{aligned}$$

Тогда

$$\frac{1}{1+a_1h_1+a_2h_2+\frac{1}{2}(b_{11}h_1^2+2b_{12}h_1h_2+b_{22}h_2^2)} = 1-a_1h_1-a_2h_2+\frac{1}{2}\Big((2a_1^2-b_{11})h_1^2 + (4a_1a_2-b_{12})h_1h_2+(2a_2^2-b_{22})h_2^2\Big)+\ldots$$

$$=\frac{1}{F(x_1,x_2)}-\frac{\left[t_1h_1\frac{\partial}{\partial x_1}+t_1h_1\frac{\partial}{\partial x_2}\right]F^2(x_1,x_2)}{F(x_1,x_2)} + \frac{\left(\left[t_1h_1\frac{\partial}{\partial x_1}+t_1h_1\frac{\partial}{\partial x_2}\right]F(x_1,x_2)\right)^2}{F^3(x_1,x_2)} - \frac{1}{2}\frac{\left[t_1h_1\frac{\partial}{\partial x_1}+t_1h_1\frac{\partial}{\partial x_2}\right]^2F(x_1,x_2)}{F^2(x_1,x_2)}+\ldots$$

Отсюда

$$\mathbf{E}(S_1)=\frac{m}{F(x_1,x_2)}+m\left(\frac{\left(\left[t_1h_1\frac{\partial}{\partial x_1}+t_1h_1\frac{\partial}{\partial x_2}\right]F(x_1,x_2)\right)^2}{F^3(x_1,x_2)} - \frac{\nu_2(\mathcal{K})}{2}\frac{\left[t_1h_1\frac{\partial}{\partial x_1}+t_1h_1\frac{\partial}{\partial x_2}\right]^2F(x_1,x_2)}{F^2(x_1,x_2)}\right)$$

$$=\frac{m}{F(\boldsymbol{x})}+\frac{m}{2F^3(\boldsymbol{x})}\left(2\nabla_F^T\mathbf{h}\mathbf{h}^T\nabla_F-\nu_2(\mathcal{K})F(\boldsymbol{x})\mathbf{h}^T\mathcal{H}_F\mathbf{h}\right)(1+o(1)).$$

Точно так же,

$$\mathbf{D}(S_1)\sim\frac{m(1-F(\boldsymbol{x}))}{n|\,\mathbf{H}\,|F^2(\boldsymbol{x})}\|\,\mathcal{K}\,\|^2.$$

Условия Ляпунова проверяются как в одномерном случае, откуда мы получаем часть (iii) теоремы. □

СПИСОК ЛИТЕРАТУРЫ

1. D.J. Finney. *Probit Analysis.* Cambridge University Press, NY. 1971. 333 p.
2. M. Razzaghi. *Statistical Models in Toxicology.* Taylor & Fransis Group, NY. 2020. 270 p.
3. С.В. Криштопенко, М.С. Тихов, Е.Б. Попова. *Доза-эффект.* М.: Медицина. 2008. 288 с.
4. R.L. Hayes, N.Mantel. *Procedures for computing the mean age of eruption of human teeth* // J. Dental Research. **35**:5, 938–947 (1958).
5. M.C. Bisi, R. Stagni. *Evalution of toddler different strategies during the first six-months of independent walking: A longitudinal study* // Gait Posture. **41**:2, 574–579 (2015).
6. М.С. Тихов, К.Н. Шкилева. *Непараметрическое оценивание квантилей в модели бинарной регрессии* // Вестник ТвГУ. Серия: Прикладная математика. **1**, 5–19 (2020).
7. M.S. Tikhov. *Statistical Estimation Based on Interval Censored Data* // Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life (ed. N.Balakrishnan etc.). Springer, NY. 555 p., 211–218 (2004).
8. М.С. Тихов, Д.С. Криштопенко. *Оценивание распределений в зависимости доза-эффект при фиксированном плане эксперимента* // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр., Перм. ун-т, Пермь. 66–77 (2006). English translation: M.S. Tikhov, D.S. Krishtopenko, *Estimation of distribution under dose-effect dependence with fixed experiment plan* // J. Math. Scien. **220**:6, 753–762 (2017).

9. Е. Надарая, П. Бабилуа, Г. Сохадзе. *Об интегральной квадратической мере уклонения одной непараметрической оценки бернуллиевой регрессии* // Теория вероятн. и ее примен. **57**:2, 322–336 (2012).
10. H. Okumura, K. Naito. *Weighted kernel estimators in nonparametric binomial regression* // J. Nonparametr. Statist. **16**:1-2, 39–62 (2004).
11. M.S. Tikhov, T.S. Borodina. *Kernel estimators of quantiles in dose-effect relationships* // Automatic Control and Computer Sciences. **2**, 29–43 (2013).
12. M.S. Tikhov. *Negative binomial regression in dose-effect relationships* // The 5^{th} International Conference on Stochastic Methods (ICSM-5): Proc. of Int. Scien. Conf. Peoples Frindship University of Russia. M. 205–208 (2020).
13. М. Кендалл, А. Стьюарт. *Теория распределений.* М.: Наука. 1966. 588 с.
14. В. Феллер. *Введение в теорию вероятностей.* Т.1 . М.: Мир. 1984. 528 с.
15. И.П. Натансон. *Теория функций вещественной переменной.* М.: Лань. 2008. 560 с.
16. H. Niederreiter. *Random number generation and quasi-Monte Carlo method.* SIAM, Philadelphia. 1992. 241 p.
17. И.М. Соболь. *Многомерные квадратурные формулы и функции Хаара.* М.: Наука. 1969.
18. Т.М. Товстик. *Вычисление дискрепанса конечного числа точек в n-мерном единичном кубе* // Вестник СПбГУ, Сер.**1**, 118–121.
19. Л. Кейперс, Г. Нидеррейтер. *Равномерное распределение последовательностей.* М.: Наука. 1985. 408 с.
20. J. Kiefer. *On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm* // Pacific. J. Math. **11**:2, 649–660 (1961).
21. J.H. Halton. *Algorithm 247: Radical-inverse quasi-random point sequence* // CACM. **7**:12, 701–702 (1964).
22. L.K. Hua, Y. Wang. *Applications of number theory to numerical analysis.* Springer-Verlag, NY. 1981. 241 p.
23. Н.М. Колобов. *Теоретико-числовые методы в приближенном анализе.* М.: МЦНМО. 2014. 285 с.
24. E.L.Lehmann. *Elements of Large-Sample Theory.* Springer, NY. 1999. 632 p.
25. М.С.Тихов. *Непараметрическое оценивание эффективных доз по данным бинарных откликов* // Уфимск. матем. журн. **5**:2, 94–108 (2013).
26. Б.В. Гнеденко. *Курс теории вероятностей.* М.: УРСС. 2005. 488 с.
27. M.B. Priestly, M.T. Chao. *Nonparametric function fitting* // J. Royal Statist. Soc., Ser. B. **34**, 385–392 (1972).
28. Н.Л. Джонсон, С. Коц, А. Кемп. *Одномерные дискретные распределения.* М.: БИНОМ. Лаборатория знаний. 2010. 559 с.
29. M. DeGroot. *Unbiased sequential estimation for binomial populations* // Ann. Math. Stat., **30**:1, 80–101 (1959).
30. Ф.П. Прудников, Ю.А. Брычков, О.И. Маричев. *Интегралы и ряды.* Т.1. Элементарные функции. М.: ФИЗМАТЛИТ. 2002. 632 с.
31. R.S. Ellis. *Large deviations for a general class of random vectors* // Ann. Statist. **12**:1, 1–12 (1984).

Михаил Семенович Тихов,
Нижегородский государственный университет им. Н.И.Лобачевского,
пр. Гагарина, 23,
603950, г. Нижний Новгород, Россия
E-mail: `tikhovm@mail.ru`