ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

# ОТРИЦАТЕЛЬНАЯ $\lambda$-БИНОМИАЛЬНАЯ РЕГРЕССИЯ В ЗАВИСИМОСТИ ДОЗА-ЭФФЕКТ

**Тихов М.С.**
Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, г. Нижний Новгород

*Поступила в редакцию 14.09.2022, после переработки 12.12.2022.*

Эта статья посвящена проблеме оценки функции распределения и ее квантилей в зависимости доза-эффект с непараметрической отрицательной $\lambda$-биномиальной регрессией. Здесь предложены ядерные оценки функции распределения, ядро которых взвешивается отрицательной $\lambda$-биномиальной случайной величиной при каждой ковариате. Наши оценки состоятельны, т.е. сходятся к своим оптимальным значениям когда число наблюдений $n$ возрастает до бесконечности. Показано, что эти 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. В первой части анализируются соотношения между моментами отрицательного $\lambda$-биномиального распределения. Получена новая характеризация распределения Пуассона.

**Ключевые слова:** модель отрицательного $\lambda$-биномиального отклика, эффективная доза, непараметрическая оценка.

*Вестник ТвГУ. Серия: Прикладная математика. 2022. № 4. С. 53–75.*
https://doi.org/10.26456/vtpmk649

## Введение

Одной из важнейших задач математической статистики является оценка по исходной выборке неизвестной функции распределения. Данная задача, а также построение эффективной оценки квантильной функции рассматриваются в настоящей статье в зависимости «доза-эффект» для модели $\lambda-$отрицательной биномиальной регрессии. Изучено асимптотическое поведение оценок функции распределения и её квантилей.

Зависимость «доза-эффект» находит свое отражение в задачах медицины, а именно в токсикологии [1] и фармакологии [2], однако спектр применимости данного подхода гораздо шире: в биологических и экологических моделях, таких как очистка воды и влияние загрязняющих факторов на экосистемы [3], ихтиологии [4] и других. Само название «доза-эффект» является условным по области возникновения первых задач, к которым был применен указанный подход.

Перейдем к математической постановке. Рассматривается задача оценивания неизвестной функции распределения (ф.р.) $F(x) = \mathbf{P}(X < x)$ случайной величины (с.в.) $X$, когда сама величина ненаблюдаема, а наблюдается повторная выборка $\mathcal{U}^{(n)} = \{(u_i, w_i), i = 1, 2, ..., n\}$, где $W_i = I(U_i > X_i)$ и по выборке $\mathcal{U}^{(n)}$ требуется оценить $F(x)$. Здесь $X$ интерпретируется как порог чувствительности субъекта, граница, с которой начинается его реакция. Плотности распределения $f(x)$ и $g(x)$ величин $X$ и $U$ неизвестны. Если с.в. $X$ и $U$ независимы, то (см.[5], [6])

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

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

Это замечание показывает, что $F(x)$ является регрессией и для оценки функции распределения можно использовать ядерные оценки регрессии.

Для оценки ф.р. $F(x)$ в зависимости «доза-эффект» будем использовать статистику

$$F_n(x) = \frac{S_{2n}(x)}{S_{1n}(x)}, \text{ если } S_{1n}(x) \neq 0, \quad \text{и} \quad F_n(x) = 0, \text{ если } S_{1n}(x) = 0,$$

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

Здесь $K_h(x) = (1/h)K(x/h)$ – финитная симметричная плотность распределения, называемая также ядром или ядерной функцией. Имеется много различных ядерных функций, но чаще всего используется ядро Епанечникова $K(x) = \frac{3}{4}\,(1 - x^2)\, I\,(\,|x|\, < 1)$, или квартическое ядро $K(x) = \frac{15}{16}\,(1 - x^2)^2\, I\,(\,|x|\, < 1)$ из-за их определенных оптимальных свойств. Выбор ядерной функции не оказывает сильного влияния на качество оценки (см. [7]). Более важным является выбор ширины окна просмотра $h$, поскольку взяв её, относительно имеющихся величин маленькой, получим поточечное восстановление исходных частот для каждой из доз, так как в точке будет учитываться узкий локальный набор из исходной выборки. В противоположном случае, когда значение $h$ слишком велико, будем наблюдать чрезмерное сглаживание и сильное усреднение получаемой оценки за счет влияния большей части исходной выборки в каждой точке. Из условия оптимальности среднеквадратичного риска ширина окна просмотра выбирается равной $h = h(n) = cn^{-1/5}$, или с помощью процедуры кросс-проверки [8].

Что касается качества оценок $F_n(x)$, то при некоторых условиях регулярности (см. [5],[6]) имеет место их асимптотическая нормальность:

$$\sqrt{nh}\,(F_n(x) - \mathbf{E}(F_n(x)) \xrightarrow[n\to\infty]{d} N(0, \sigma^2(x)), \text{ где } \sigma^2(x) = \frac{F(x)\,(1 - F(x))}{g(x)}\,||K||^2,$$

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

Такая дисперсия получается при случайном плане эксперимента, т.е. когда величина $U$ случайна, если же вводимая доза $U$ неслучайна, т.е. $U_i = u_i$, где $0 = u_0 < u_1 < ... < u_n < u_{n+1} = 1$, то такой план эксперимента будем называть ***фиксированным*** [9] и в модели биномиальной регрессии [10] предельная дисперсия равна $\sigma^2(x) = F(x)\,(1 - F(x))\,||\,K\,||^2$.

## 1. Необходимые понятия и соотношения

*1.1 Отрицательное $\lambda$ -биномиальное распределение* ($\lambda$-NBD)

В статье будет рассматриваться отрицательная биномиальная регрессионная модель ($\lambda$-NBD), т.е. план испытаний, когда количество испытаний при фиксированном значении $u_i$ имеет отрицательное $\lambda$-биномиальное распределение, поэтому мы сначала изучим свойства этого распределения.

Разложим функцию $g(x) = (1 - x)^{-a}$ в ряд Тейлора по переменной $x$. Имеем:

$$(1 - x)^{-\,a} = 1 + ax + \frac{a(a+1)}{2!}x^2 + \frac{a(a+1)(a+2)}{3!}q^3 + ... \,.$$

Если $x = \lambda q,\ a = \dfrac{r}{\lambda}$, то получим

$$(1 - \lambda q)^{-\,\frac{r}{\lambda}} = 1 + rq + \frac{r(r+\lambda)}{2!}q^2 + \frac{r(r+\lambda)(r+2\lambda)}{3!}q^3 + ... = \tag{1}$$

$$= 1 + \sum_{k=1}^{\infty} \frac{r(r+\lambda)\cdot ...\cdot(r+(k-1)\lambda)}{k!}\, q^k, \tag{2}$$

откуда

$$1 = (1 - \lambda q)^{\,\frac{r}{\lambda}} + \sum_{k\,=\,1}^{\infty} \frac{r(r+\lambda)\cdot ...\cdot(r+(k-1)\lambda)}{k!}(1 - \lambda q)^{\,\frac{r}{\lambda}}q^k.$$

Рассматривая $0 < \lambda < 1, 0 < q < 1$, определим

$$\begin{aligned} &p_\lambda(0) = (1 - \lambda q)^{\,\frac{r}{\lambda}}, \\ &p_\lambda(k) = \sum_{k\,=\,1}^{\infty} \frac{r(r+\lambda)\cdot ...\cdot(r+(k-1)\lambda)}{k!}\,(1 - \lambda q)^{\frac{r}{\lambda}}q^k,\ k = 1, 2, ..., \end{aligned} \tag{3}$$

которое есть распределение вероятностей $(\lambda - NBD)$.
При $\lambda \to 1$ получаем отрицательное биномиальное распределение $NB(p, r)$, а при $\lambda \to 0$ получаем распределение Пуассона с параметром $qr$.

Обозначим исходную величину, имеющую $\lambda$-NBD распределение, через $Y_\lambda$ и найдем её математическое ожидание. Имеем:

$$\mathbf{E}(Y_\lambda) = (1 - \lambda q)^{\frac{r}{\lambda}} \sum_{k\,=\,1}^{\infty} \frac{r(r+\lambda)\cdot ...\cdot(r+(k-1)\lambda)}{(k-1)!}\, q^k = \frac{rq}{1 - \lambda q}.$$

Для величины $Z_\lambda = Y_\lambda + r/\lambda$ математическое ожидание будет равно

$$\mathbf{E}(Z_\lambda) = \mathbf{E}\left(Y_\lambda + \frac{r}{\lambda}\right) = \frac{rq}{1 - \lambda q} + \frac{r}{\lambda} = \frac{r}{\lambda\,(1 - \lambda q)}.$$

Воспользовавшись равенством $k^2 = k((k-1)+1) = k(k-1)+k$, получим

$$\mathbf{E}(Y_\lambda^2) = (1-\lambda q)^{\frac{r}{\lambda}} \sum_{k=1}^{\infty} k^2 \frac{r(r+\lambda)\cdot \ldots \cdot (r+(k-1)\lambda)}{k!}\, q^k = \frac{rq(rq+1)}{(1-\lambda q)^2}\,.$$

и дисперсия будет равна

$$\mathbf{D}(Y_\lambda) = \frac{rq}{(1-\lambda q)^2}\,.$$

Заметим, что

$$\lim_{\lambda\to 1} p_\lambda(k) = \frac{r(r+1)\ldots(r+k-1)}{k!}(1-q)^r q^k$$

есть вероятностная функция отрицательной биномиальной случайной величины с параметрами $(r, 1-q)$. В этом случае мы имеем ожидание и дисперсию, равные соответственно

$$\lim_{\lambda\to 1} \mathbf{E}(Y_\lambda) = \frac{rq}{1-q}, \qquad \lim_{\lambda\to 1} \mathbf{D}(Y_\lambda) = \frac{rq}{(1-q)^2}.$$

Отметим также, что

$$\lim_{\lambda\to 0} p_\lambda(k) = \frac{(rq)^k}{k!}\, e^{-rq}$$

есть вероятностная функция Пуассоновской случайной величины с параметрами $rq$, а ее математическое ожидание и дисперсия будут равны

$$\lim_{\lambda\to 0} \mathbf{E}(Y_\lambda) = rq, \qquad \lim_{\lambda\to 0} D(Y_\lambda) = rq\,.$$

Кроме математического ожидания и дисперсии нам понадобятся также моменты более высокого порядка: $\mathbf{E}(Y_\lambda^m), m \geqslant 3$. Из соотношения $k^n = \sum_{j=0}^{n} S_2(n,j)(k)_j$, $(k)_j = k(k-1)\cdot\ldots\cdot(k-j+1)$, определим числа Стирлинга 2-го рода $S_2(n,j)$ (см. [12]), для которых

$$S_2(n,k) = S_2(n-1,k-1) + k\cdot S_2(n-1,k).$$

Явная формула для вычисления этих чисел есть $S_2(n,k) = \dfrac{1}{k!}\sum\limits_{j=0}^{k}(-1)^{k+j}\dbinom{k}{j} j^n\,.$

Тогда

$$\mathbf{E}(Y_\lambda^m) = \sum_{k=1}^{\infty} S_2(m,k)\frac{r(r+\lambda)\cdot\ldots\cdot(r+(k-1)\lambda)}{(1-\lambda q)^k}\, q^k,\ m \geqslant 3,$$

в частности,

$$\mathbf{E}(Y_\lambda^3) = \frac{rq(r^2q^2+1+(3r+\lambda)q)}{(1-\lambda q)^3}\,.$$

Мы можем вычислить характеристическую функцию (х.ф.) величины $Y_\lambda$, которая равна

$$\varphi = \varphi(t) = 1 + \sum_{k=1}^{\infty} e^{\theta k}\cdot p_\lambda(k) = \frac{(1-\lambda q)^{\frac{r}{\lambda}}}{(1-\lambda q e^{\theta})^{\frac{r}{\lambda}}}\,, \quad \theta = it\,. \tag{4}$$

Теперь выведем соотношения для начальных моментов распределения (3). Для центральных моментов биномиального распределения и распределения Пуассона аналогичные соотношения между моментами рассмотрены в [13], с.174(5.9), с.179(5.21) и приведены также в книге [14], с.173(4.16).

Обозначим $\mu_k = \mathbf{E}(X^k)$, $k = 1, 2, ...$, начальный момент порядка $k$ и разложим $\varphi$ в ряд по $\theta$. Получаем

$$\varphi = 1 + \sum_{j=1}^{\infty} \frac{\mu_j \theta^j}{j!} = \left( \frac{1-\lambda q}{1-\lambda q e^{\theta}} \right)^{\frac{r}{\lambda}} . \tag{5}$$

Дифференцируя обе части равенства (5) по $\theta$, приходим к равенству

$$\sum_{j=1}^{\infty} \frac{\mu_{j-1}\theta^{j-1}}{(j-1)!} = \frac{rqe^{\theta}}{1-\lambda q e^{\theta}} \cdot \frac{(1-\lambda q)^{\frac{r}{\lambda}}}{(1-\lambda q \cdot e^{\theta})^{\frac{r}{\lambda}}} .$$

Умножая обе части на $(1-\lambda q e^{\theta})$, мы будем иметь:

$$(1-\lambda q e^{\theta}) \sum_{j=1}^{\infty} \frac{\mu_{j-1}\theta^{j-1}}{(j-1)!} = rqe^{\theta} \sum_{j=1}^{\infty} \frac{\mu_j \theta^j}{j!} .$$

Разлагая $e^{\theta}$, используя правило умножения рядов и приравнивая коэффициенты при $\theta^k$, находим следующие выражения для начальных моментов $(k+1)$-го порядка через такие же моменты более низкого порядка:

$$\mu_{k+1} = \frac{rq}{1-\lambda q} \sum_{j=0}^{k} \binom{k}{j} \mu_j + \frac{\lambda q}{1-\lambda q} \sum_{j=0}^{k-1} \binom{k}{j} \mu_{j+1} . \tag{6}$$

В частности, из (6) выводим:

$$\mathbf{E}(Y_{\lambda}^4) = \frac{rq(1 + r^3q^3 + (\lambda^2 + 4\lambda r + 6r^2)q^2 + (4\lambda + 7r)q)}{(1-\lambda q)^4} .$$

Для отрицательного биномиального распределения соотношение (6) преобразуется в равенство

$$\mu_{k+1} = \frac{rq}{1-q} \sum_{j=0}^{k} \binom{k}{j} \mu_j + \frac{q}{1-q} \sum_{j=0}^{k-1} \binom{k}{j} \mu_{j+1} , \tag{7}$$

а для распределения Пуассона $\mathcal{P}(a)$ оно примет вид:

$$\mu_{k+1} = a \sum_{j=0}^{k} \binom{k}{j} \mu_j = \sum_{l=1}^{k+1} S_2(k+1, l) a^l . \tag{8}$$

Теперь заметим, что

$$\mu_{k+1} = \mathbf{E}(X^{k+1}) = \mathbf{E}(X \cdot X^k) = a \sum_{j=0}^{k} \binom{k}{j} \mu_j = a\mathbf{E}((X+1)^k) ,$$

откуда легко следует характеризация распределения Пуассона [15]:
*Случайная величина* $Z \in \mathcal{P}(a)$ *(т.е. имеет распределение Пуассона), тогда и только тогда, когда для любой ограниченной функции*

$$f : \mathbf{N}\bigcup\{0\} \to \mathbf{R}, \quad \mathbf{E}(af(Z+1)) - \mathbf{E}(Zf(Z)) = 0,\ a > 0,$$

(здесь можно ограничиться степенными функциями $f(x) = x^k$, из чего получаем такие же моменты, как у распределения Пуассона, а далее подставляем в разложение $\varphi$ в ряд по степеням $\theta$ и ввиду выполнения условия Карлемана получаем распределение Пуассона).

Из (6) также выводим, что

$$\mathbf{E}((Y_\lambda - \mathbf{E}(Y_\lambda))^3) = \frac{rq(1+\lambda q)}{(1-\lambda q)^3}\,, \quad \text{значит, асимметрия равна} \quad \beta_1 = \frac{1+\lambda q}{\sqrt{rq}}\,,$$

а эксцесс равен

$$\beta_2 = \frac{\lambda^2 q^2 + 4\lambda q + 1}{rq}\,.$$

Далее, согласно определению

$$\mu_k = \sum_{j=1}^{\infty} j^k \frac{r(r+\lambda)\cdot\ldots\cdot(r+(j-1)\lambda)}{j!}(1-\lambda q)^{\frac{r}{\lambda}} q^j\,.$$

Дифференцируя по $q$, выводим

$$\mu_{k+1} = q\frac{d\mu_k}{dq} + \frac{rq}{1-\lambda q}\mu_k\,. \tag{9}$$

Обозначая через $\psi(t)$ характеристическую функцию величины $V_\lambda = Y_\lambda - rq/(1-\lambda q)$ и действуя как в случае начальных моментов, мы получим аналогичные соотношения для центральных моментов величины $Y_\lambda$:

$$\nu_{k+1} = \frac{rq}{(1-\lambda q)^2}\sum_{j=0}^{k-1}\binom{k}{j}\nu_j + \frac{\lambda q}{1-\lambda q}\sum_{j=1}^{k-1}\binom{k}{j}\nu_{j+1}\,, \tag{10}$$

где $\nu_j$ – центральный момент, $\nu_0 = 1$, $\nu_1 = 0$, что дает возможность легко их вычислить.

Из формулы (10) выводим: $\nu_4 = \dfrac{qr(1+\lambda^2q^2+(4\lambda+3r)q)}{(1-\lambda q)^4}$.

Для отрицательного биномиального распределения формула (10) преобразуется к виду

$$\nu_{k+1} = \frac{rq}{(1-q)^2}\sum_{j=0}^{k-1}\binom{k}{j}\nu_j + \frac{q}{1-q}\sum_{j=1}^{k-1}\binom{k}{j}\nu_{j+1}\,,$$

а для распределения Пуассона (см. [13], с.179; [14], с.173)

$$\nu_{k+1} = a\sum_{j=0}^{k-1}\binom{k}{j}\nu_j\,.$$

Рассмотрим снова характеристическую функцию отрицательного $\lambda$-биномиального распределения (4).

У этого распределения существуют начальные моменты $\mu_j$ любого порядка $j$, поэтому

$$\varphi = 1 + \sum_{j=1}^{\infty} \frac{\mu_j \theta^j}{j!} \,. \tag{11}$$

Дважды дифференцируя по $\theta$ соотношения (5) и (11), выводим:

$$\sum_{j=0}^{\infty} \frac{\mu_{j+2}}{j!}\theta^j = \frac{q^2 r(r+\lambda)2^{2\theta}}{(1-\lambda q e^{\theta})^2} \frac{(1-\lambda q)^{r/\lambda}}{(1-\lambda q e^{\theta})^{r/\lambda}} + \frac{q r e^{\theta}((1-\lambda q e^{\theta}))}{(1-\lambda q e^{\theta})^2} \frac{(1-\lambda q)^{r/\lambda}}{(1-\lambda q e^{\theta})^{r/\lambda}} \,. \tag{12}$$

Умножим обе части равенства (12) на $(1-\lambda q e^{\theta})^2$, получим

$$((q^2 r(r+\lambda) - \lambda r q^2)e^{2\theta} + q r e^{\theta}) \left( \sum_{k=0}^{\infty} \frac{\mu_k}{k!}\theta^k \right) = \sum_{j=0}^{\infty} \frac{\mu_{j+2}}{j!}\theta^j \,.$$

Далее, разложим $e^{2\theta}$ и $e^{\theta}$ в ряд, перемножим ряды и приравняем члены при одинаковых степенях $\theta^m$. В результате мы получим равенство

$$\mu_{m+2} = ((q^2 r(r+\lambda) - \lambda r q^2) \sum_{j=0}^{m} \binom{m}{j} 2^{m-j}\mu_j + q r \sum_{j=0}^{m} \binom{m}{j}\mu_j \,.$$

Пусть $\lambda \to 0$, а $qr = a$. Тогда

$$\mu_{m+2} = a^2 \sum_{j=0}^{m} \binom{m}{j} 2^{m-j}\mu_j + a \sum_{j=0}^{m} \binom{m}{j}\mu_j \,.$$

Но

$$a^2 \mathbf{E}((X+2)^m) = \mathbf{E}(X(X-1)X^m)) = \mu_{m+2} - \mu_{m+1} \,,$$

откуда уже нетрудно вывести характеризацию распределения Пуассона:
*Случайная величина $Z \in \mathcal{P}(a)$ тогда и только тогда, когда для любой ограниченной функции*

$$f : \mathbf{N}\bigcup\{0\} \to \mathbf{R}, \quad \mathbf{E}(af(Z+2)) = \mathbf{E}(Z(Z-1)f(Z)), \text{ и } \mathbf{E}(Z) = a,\ a > 0 \,. \tag{13}$$

Соотношение (13) можно обобщить, потребовав

$$\mathbf{E}(a^r f(Z+r)) = \mathbf{E}((Z)_r f(Z)),\ a > 0, \text{ где } (Z)_r = Z(Z-1)\cdot \ldots \cdot (Z-r+1), \tag{14}$$

с начальными условиями

$$\mu_{k+1} = \sum_{j=1}^{k+1} S_2(k+1, j)a^j, \quad k = 0, 1, ..., r-2 \,.$$

Эту же характеризацию можно получить исходя из биномиального распределения $B(n,p)$ ) когда $np \to a$ при $n \to \infty$, для которого

$$\mu_{m+2} = ((n^2p^2 + np^2)\sum_{j=0}^{m}\binom{m}{j}2^{m-j}\mu_j + np(1-p)\sum_{j=0}^{m}\binom{m}{j}\mu_j\,.$$

*1.2 Метод Стейна-Тихомирова*

Для доказательства центральной предельной теоремы мы воспользуемся модифицированным вариантом метода Стейна-Тихомирова (см. [16], [17]). Именно, введем оператор

$$\Delta(\varphi_n(t)) = \varphi_n'(t) + t\varphi_n(t). \tag{15}$$

Если докажем (см. [18]), что для $t \in \mathbf{R}$ $\lim\limits_{n\to\infty} \Delta(\mathbf{E}(e^{itS_n}) = 0$, то распределение случайной величины $S_n$ будет асимптотически нормально. Но это следует из (15) и того, что $\varphi_n(t) \underset{n\to\infty}{\to} 1$ равномерно на множестве $\{t:\ |t| \le T\}$.

Пусть $S_n = \sum_{i=1}^{n}\xi_i,\ \mathbf{E}(\xi_i) = 0,\ \sigma_i^2 = \mathbf{D}(\xi_i) = \mathbf{E}(\xi_i^2)\,,\ \sum_{i=1}^{n}\sigma_i^2 = 1$ и $S_n^{(i)} = \sum\limits_{j\ne i}\xi_j$ . Тогда $S_n = S_n^{(i)} + \xi_i$ .

Для некоторой функции $\psi$ мы имеем

$$\mathbf{E}(S_n\psi(S_n)) = \mathbf{E}\left(\sum_{i=1}^{n}\xi_i\psi(S_n)\right) = \mathbf{E}\left(\sum_{i=1}^{n}\xi_i\psi(S_n^{(i)}+\xi_i)\right).$$

Предположим, что функция $\psi(x)$ имеет вторую непрерывную производную и $||\,\psi''||\ = \sup\limits_x|\psi''(x)| < \infty$. Тогда

$$\mathbf{E}(S_n\psi(S_n)) = \mathbf{E}\left(\sum_{i=1}^{n}\xi_i\psi(S_n)\right) = \mathbf{E}\left(\sum_{i=1}^{n}\xi_i\psi(S_n^{(i)}+\xi_i)\right),$$

$$\xi_i\psi(S_n^{(i)}+\xi_i) = \xi_i\psi(S_n^{(i)}) + \xi_i^2\psi'(S_n^{(i)}) + \frac{1}{2}\xi_i^3\psi''(S_n^{(i)}+\theta_i\xi_i),$$

$(\xi_i, S_n^{(i)})$ – независимы, и поэтому для некоторых $\gamma_i$,

$$\mathbf{E}(\psi'(S_n)) = \mathbf{E}\left(\sum_{i=1}^{n}\sigma_i^2\psi'(S_n^{(i)})\right) + \mathbf{E}\left(\sum_{i=1}^{n}\sigma_i^2(\psi'(S_n) - \psi'(S_n^{(i)}))\right) =$$

$$= \mathbf{E}\left(\sum_{i=1}^{n}\sigma_i^2\psi'(S_n^{(i)})\right) + \mathbf{E}\left(\sum_{i=1}^{n}\sigma_i^2\xi_i\psi''(S_n^{(i)}+\gamma_i\xi_i))\right)$$

и

$$|\,\mathbf{E}(\psi'(S_n^{(i)}) - S_n\psi(S_n))\,| \ \le\ ||\,\psi''\,||\ \sum_{i=1}^{n}(\mathbf{E}(\,|\,\xi_i\,|^3) + \sigma_i^2\mathbf{E}(\,|\,\xi_i\,|))\ \le 2||\,\psi''\,||\ \sum_{i=1}^{n}\mathbf{E}(\,|\,\xi_i\,|^3),$$

поскольку предполагая, что $p, q > 0,\ a, b > 0,\ \dfrac{1}{p} + \dfrac{1}{q} = 1$ из вогнутости функции $\ln(x)$ следует

$$\ln\left(\frac{a}{p} + \frac{b}{q}\right) \geq \frac{\ln a}{p} + \frac{\ln b}{q} = \ln\left(a^{\frac{1}{p}} b^{\frac{1}{q}}\right), \text{ значит, } a^{\frac{1}{p}} b^{\frac{1}{q}} \leq \frac{a}{p} + \frac{b}{q},$$

а если $a = |\,\xi_i|^3 = b,\ p = 3,\ q = \dfrac{3}{2}$, то

$$a^{\frac{1}{p}} b^{\frac{1}{q}} = |\,\xi_i| \cdot |\,\xi_i|^2 \leq \frac{|\,\xi_i|^3}{3} + \frac{2|\,\xi_i|^3}{3} = |\,\xi_i|^3.$$

Если окажется, что $||\,\psi''\,|| < \infty$, а

$$\sum_{i=1}^{n} \mathbf{E}(\,|\,\xi_i\,|^3) \underset{n\to\infty}{\longrightarrow} 0, \tag{16}$$

то, тем самым, будет доказана центральная предельная теорема.

В качестве функции $\psi(x)$ выберем функцию $\psi(x) = \exp(ix)$, для которой $|\,\psi''(x)\,| = 1$ и, значит, $||\,\psi''\,|| = 1$, а условие (16) есть условие Ляпунова, поскольку $\sum_{i=1}^{n} \sigma_i^2 = 1$.

Преимущество метода Стейна состоит в том, что он позволяет оценивать также границы уклонения в нормальной аппроксимации распределения сумм случайных величин.

Далее, рассмотрим случайную величину $Z_i = Y_\lambda + r/\lambda$ и будем использовать выборку $\mathcal{Z} = \{(z_i, u_i), i = 1, 2, ..., n\}$ чтобы построить оценку функции распределения $F(x)$ вида:

$$T_n(x) = 1 - \frac{1}{l} + \frac{1}{l^2} \cdot \frac{\sum\limits_{i=1}^{n} r \cdot K_h(x - u_i)}{\sum\limits_{i=1}^{n} z_i \cdot K_h(x - u_i)}.$$

Так как $\dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} K_h(x - u_i) \to 1$ при $n \to \infty$ , то рассмотрим оценку:

$$\hat{F}_n(x) = 1 - \frac{1}{l} + \frac{1}{l^2} \cdot \frac{r}{\sum\limits_{i=1}^{n} z_i \cdot K_h(x - u_i)}. \tag{17}$$

В работе [11] рассматривалась модель отрицательной биномиальной регрессии в зависимости «доза-эффект» и было показано, что предельная дисперсия построенной там оценки функции распределения меньше, чем у оценки в модели биномиальной регрессии. Именно, в [11] рассматривался фиксированный план испытаний и на уровне $u_i$ проводились испытания до появления $r$ «успехов» ($W_i = 1$), где в качестве вероятности «успеха» рассматривается $p_i = F(u_i)$. Было показано, что построенные оценки асимптотически нормальны. В настоящей статье мы рассматриваем обобщение этой ситуации, которое включает как модель отрицательной биномиальной регрессии, так и модель Пуассоновской регрессии.

В данной статье будем также рассматривать фиксированный план эксперимента, где $u_i = i/n$, но результаты будут верны, когда $U_i$ заданы с помощью квази-случайных последовательностей с низким расхождением (low-discrepancy sequences), модель $\lambda$-отрицательной биномиальной регрессии ($\lambda - NBD$) и случайные величины с плотностью распределения $f(x)$, заданной на интервале $(0,1)$, для которой $f(x) > 0, x \in (0,1)$. В модели биномиальной регрессии [10] при ядерном сглаживании края распределения $F(x)$ оцениваются не очень хорошо. В нашей модели на краях распределения наблюдений получается в среднем больше, чем в середине и такое планирование позволяет лучше оценивать $F(x)$.

## 2. Вспомогательные результаты

Для исследования оценок по фиксированным планам эксперимента нам понадобится понятие *вариации* функции ([19], с.234).

Пусть на отрезке $[\,a,b\,]$ задана функция $f(x)$. Разобьем $[\,a,b\,]$ на части точками $x_0 = a < x_1 < x_2 < \ldots < x_n = b$ и составим сумму

$$\boldsymbol{VS}(f) = \sum_{k=0}^{n-1} |\, f(x_{k+1}) - f(x_k)\,|.$$

**Определение 1.** *Точная верхняя грань множества всевозможных сумм* $\boldsymbol{VS}(f)$ *называется вариацией функции* $f(x)$ *на отрезке* $[\,a,b\,]$ *и обозначается* $\boldsymbol{V}_a^b(f)$, *т.е.*

$$\boldsymbol{V}_a^b(f) = \sup_{\mathcal{P}_n} \boldsymbol{VS}(f) = \sup_{\mathcal{P}_n} \sum_{k=0}^{n-1} |\, f(x_{k+1}) - f(x_k)\,| \quad \text{где} \quad \mathcal{P}_n = \{x_0, x_1, \ldots, x_n, x_{n+1}\}.$$

Если $\boldsymbol{V}_a^b(f) < \infty$, то говорят, что $f(x)$ есть *функция конечной* (*ограниченной*) *вариации* на отрезке $[\,a,b\,]$. Если функция $f : [\,a,b\,] \to \mathbf{R}$ принадлежит классу $C^1[\,a,b\,]$, т. е. имеет непрерывную производную первого порядка, модуль которой интегрируем, то $f$ есть функция ограниченной вариации на этом отрезке, а ее вариацию можно вычислить как

$$\boldsymbol{V}_a^b(f) = \int_a^b |\, f'(x)\,|\, dx.$$

Далее, пусть $P$ есть множество точек $x_1, x_2, \ldots, x_N$ из $[\,0,1\,]$ и $\mathcal{B}-$ лебегова $\sigma-$алгебра на $[\,0,1\,]$, а $\mu$ – лебегова мера на $\mathcal{B}$. Определим

$$A(B,P) = \sum_{i=1}^{N} I_B(x_i), \quad D_N(B,P) = \left|\, \frac{A(B,P)}{N} - \mu(B) \,\right|,$$

где $I_B(x)-$ индикатор множества $B$. Положим $D_N^*(P) = D_N(J_c^*, P)$, где $J_c^*$ есть семейство подинтервалов на $[\,0,1\,]$ вида $[\,0,u_i)$. Величину $D_N^*(P)$ называют дискрепансом (discrepancy) последовательности $P$.

**Теорема 1** (20, с.18, Koksma, Hlawka)**.** *Если функция* $f(u)$ $(0 \le u \le 1)$ *имеет ограниченную вариацию* $\boldsymbol{V}_a^b(f)$, *то для любых* $0 < x_1 < x_2 < \ldots < x_N < 1$ *мы имеем*

$$\left|\, \frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(x)\, du \,\right| \le \boldsymbol{V}_a^b(f) D_N^*(x_1, \ldots, x_N)\,.$$

### 3. Предположения

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

**Предположения (K).**

$(\mathbf{K}_1)$ $K(x) \geq 0$, *причем* $K(x) = 0,\ x \notin [-1, 1]$. $(\mathbf{K}_2)$ $\int\limits_{-\infty}^{\infty} K(x)\, dx = 1$.

$(\mathbf{K}_3)$ $K(x) = K(-x),\ x \in \mathbf{R}$. $(\mathbf{K}_4)$ $\|K\|_\infty = \sup\limits_{x \in \mathbf{R}} |\, K\,| = \kappa < \infty$.

$(\mathbf{K}_5)$ *Существуют непрерывные ограниченные производные функции* $K(x)$ *до третьего порядка включительно в интервале* $(-1, 1)$.

$(\mathbf{K}_6)$ *Вариация* $\boldsymbol{V}_{-1}^{1}(K)$ *функции* $K(x)$ *ограничена.*

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

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

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

$$\nu_2(K) = \int_{-\infty}^{\infty} x^2 K(x)\, dx\,, \quad \mu_4(K) = \int_{-\infty}^{\infty} x^4 K(x)\, dx\,.$$

Примером функции, удовлетворяющей условиям $(\mathbf{K})$ 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.

**Предположения (H).**

$(\mathbf{H}_1)$ *При* $n \to \infty, \quad h \to 0,\ nh \to \infty$.

Условие $(\mathbf{H}_1)$ означает, что по мере получения большего количества информации из выборки, т.е. при $n \to \infty$, усреднение данных происходит по более узкой области $(h \to 0)$, но в то же время количество «локальной информации» $(nh)$ должно увеличиваться.

Примером числовой последовательности, удовлетворяющей условию $(\mathbf{H}_1)$, является $h = cn^{-1/5}$, $c-$некоторая положительная константа.

**Предположение (F)**.

$(\mathbf{F}_1)$ *Плотность* $f(x)$ *определена на отрезке* $[0, 1]$, *отделима от нуля, т.е* $f(x) \geq c_0 > 0$ *для* $0 \leq x \leq 1$, *и существует третья непрерывная ограниченная производная плотности на интервале* $(0, 1)$, *причем* $f(x) > 0$ *и* $\int_0^1 |\, f'(x)\,|\, dx < \infty$.

Пусть $\mathcal{P}_n = \{u_0, u_1, ..., u_n, u_{n+1}\}$ – упорядоченное разбиение отрезка $[0,1]$, где $u_0 = 0 < u_1 < ... < u_n < 1 = u_{n+1}$.

**Предположение (P)**.

$(\mathbf{P}_1)$ *При* $n \to \infty$, $\max\limits_{k\in\{0,1,\dots,n\}} \max\left\{\left|u_k - \frac{k}{n}\right|, \left|u_{k+1} - \frac{k}{n}\right|\right\} = O(n^{-1}).$

Из условия $(\mathbf{P})$ следует, что $u_k = \frac{k}{n} + O(n^{-1})$, причем последовательность $n\left(u_k - \frac{k}{n}\right)$ равномерно по $0 \le k \le n$ ограничена некоторой константой $C$.

Обозначим $\tilde{q}(u) = K_h(u)$.

**Лемма 1** ([21])**.** *Если выполнены предположения **(K)**, **(H)** , **(F)** , **(P)** то*

$$\sup_{\mathcal{P}_n} \sum_{j=1}^{l} |\ \tilde{q}(u_j) - \ \tilde{q}(u_{j-1})\,| \ = O(h_d^{-1}),$$

*где* sup *берется по всем возможным упорядоченным разбиениям* $\mathcal{P}_n$.

## 4. Результаты

**Теорема 2.** *Пусть выполнены условия регулярности **(K)**, **(H)**, **(F)**, а* $u_i = \frac{i}{n}, i = 1, ..., n$. *Тогда*

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

*Доказательство.* Поскольку вариация ядра (см. [19]) $\boldsymbol{V}_{-h}^{h}\left(K_h\right) = O\left(\frac{\ln n}{h}\right)$, то при $n \to \infty$ мы имеем:

$$\mathbf{E}\left(S_1\right) = \mathbf{E}\left(\frac{1}{nh}\sum_{i=1}^{n} Z_i K\left(\frac{u_i - x}{h}\right)\right) =$$

$$= \frac{1}{nh}\sum_{i=1}^{n}\mathbf{E}\left(Z_i\right)K\left(\frac{u_i - x}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{u_i - x}{h}\right)\cdot\frac{r}{\lambda\left(1-\lambda\left(1-F\left(u_i\right)\right)\right)} =$$

$$= \frac{r}{\lambda h}\int\limits_{-\infty}^{+\infty} K\left(\frac{u-x}{h}\right)\frac{1}{1-\lambda\left(1-F\left(u\right)\right)}\,du + O\left(\frac{\ln n}{nh}\right) =$$

$$= \frac{r}{\lambda}\int\limits_{-x/h}^{(1-x)/h}\frac{K(t)}{1-\lambda\left(1-F\left(x+ht\right)\right)}dt + O\left(\frac{\ln n}{nh}\right) = \frac{r}{\lambda}\int\limits_{-\infty}^{+\infty}\frac{K(t)}{1-\lambda\left(1-F\left(x\right)\right)}\,dt +$$

$$+\frac{r}{\lambda}\int\limits_{-\infty}^{+\infty}K(t)\left(\frac{1}{1-\lambda\left(1-F\left(x+ht\right)\right)}-\frac{1}{1-\lambda\left(1-F\left(x\right)\right)}\right)dt+O\left(\frac{\ln n}{nh}\right)=$$

$$=\frac{r}{\lambda\left(1-\lambda\left(1-F\left(x\right)\right)\right)}+$$

$$+\frac{r}{\lambda}\int_{-\infty}^{+\infty}K(t)\left(\frac{\lambda\left(F\left(x\right)-F\left(x+ht\right)\right)}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)dt+O\left(\frac{\ln n}{nh}\right)=$$

$$=\frac{r}{\lambda\left(1-\lambda\left(1-F\left(x\right)\right)\right)}-$$

$$-r\int_{-\infty}^{+\infty}K(t)\left(\frac{F\left(x+ht\right)-F\left(x\right)}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)dt+O\left(\frac{\ln n}{nh}\right).$$

Оценим последний интеграл. Имеем:

$$\int\limits_{-\infty}^{+\infty}K(t)\left(\frac{F\left(x+ht\right)-F\left(x\right)}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)dt=$$

$$=\int\limits_{-\infty}^{+\infty}K(t)\left(\frac{f(x)ht+1/2f'(x)h^2t^2+1/6f''(x)h^3t^3+1/24f'''(\varsigma)h^4t^4}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)dt,$$

где $\varsigma$ есть «средняя» точка.

Здесь

$$\left|\,h^4\int\limits_{-\infty}^{+\infty}K(t)\left(\frac{f'''(\varsigma)t^4}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)dt\right|\le$$

$$\le h^4\nu_4(K)\cdot\ \sup\frac{f'''(\varsigma)}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\le$$

$$\le\frac{2\tilde{C}h^4}{\left(1-\lambda+\lambda F\left(x\right)-\varepsilon\right)\left(1-\lambda+\lambda F\left(x\right)\right)}\ \text{для}\ n\ge n_1.$$

Преобразуя подынтегральное выражение с учетом этого замечания, получаем

$$\int\limits_{-\infty}^{+\infty}K(t)\left(\frac{f(x)ht+1/2f'(x)h^2t^2}{\left(1-\lambda+\lambda F\left(x+ht\right)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)dt+O(h^4)=$$

$$\int\limits_{-\infty}^{+\infty}K(t)\left(\frac{f(x)ht+1/2f'(x)h^2t^2}{\left(1-\lambda(1-F\left(x\right)-f(x)ht)\right)\left(1-\lambda\left(1-F\left(x\right)\right)\right)+O(h^2)}\right)dt+O(h^4)=$$

$$\frac{1}{\left(1-\lambda\left(1-F(x)\right)\right)^2}\int\limits_{-\infty}^{+\infty}\left(f(x)ht+\frac{1}{2}f'(x)h^2t^2\right)\times$$

$$\times\left(1-\frac{\lambda f(x)}{1-\lambda\left(1-F(x)\right)}ht\right)K(t)\,dt+O(h^3)=$$

$$=\frac{1}{\left(1-\lambda\left(1-F(x)\right)\right)^2}\int\limits_{-\infty}^{+\infty}\left(f(x)ht+\frac{1}{2}f'(x)h^2t^2-\frac{\lambda f^2(x)}{1-\lambda\left(1-F(x)\right)}h^2t^2\right)K(t)\,dt+o(h^2).$$

Учитывая, что $\int_{-\infty}^{+\infty}tK(t)dt=0$, получим следующую оценку для исходного интеграла:

$$\left(\frac{1}{2}\cdot\frac{f'(x)}{\left(1-\lambda\left(1-F(x)\right)\right)^2}-\frac{\lambda f^2(x)}{\left(1-\lambda\left(1-F(x)\right)\right)^3}\right)h^2\nu_2(K)+o(h^2),$$

откуда

$$\mathbf{E}(S_1)=\frac{r}{\lambda\left(1-\lambda\left(1-F\left(x\right)\right)\right)}-$$

$$-r\cdot\left(\frac{1}{2}\cdot\frac{f}{\left(1-\lambda\left(1-F(x)\right)\right)^2}-\frac{\lambda f^2(x)}{\left(1-\lambda\left(1-F(x)\right)\right)^3}\right)h^2\nu_2(K)+o(h^2).$$

Аналогично,

$$\mathbf{D}\left(S_1\right)=\mathbf{D}\left(\frac{1}{nh}\sum_{i=1}^{n}Z_iK\left(\frac{u_i-x}{h}\right)\right)=\frac{1}{n^2h^2}\sum_{i=1}^{n}K^2\left(\frac{u_i-x}{h}\right)D\left(Z_i\right)=$$

$$=\frac{1}{n^2h^2}\sum_{i=1}^{n}K^2\left(\frac{u_i-x}{h}\right)\frac{r(1-F(u_i))}{\left(1-\lambda\left(1-F(u_i)\right)\right)^2}\sim$$

$$\sim\frac{1}{nh^2}\int_{-\infty}^{+\infty}\frac{r(1-F(u))}{\left(1-\lambda\left(1-F(u)\right)\right)^2}K^2\left(\frac{u-x}{h}\right)du=$$

$$=\frac{r}{nh}\int_{-\infty}^{+\infty}\frac{1-F(x+ht)}{\left(1-\lambda\left(1-F(x+ht)\right)\right)^2}K^2\left(t\right)dt\sim\frac{r\left(1-F(x)\right)}{nh(1-\lambda(1-F(x)))^2}\left\|K\right\|^2.$$

Чтобы доказать асимптотическую нормальность статистики $S_1=S_{1n}(x)$ надо проверить условие Ляпунова. Для этого воспользуемся неравенством ($a\geqslant 1$):

$$|x+y|^a\le 2^{a-1}\left(|x|^a+|y|^a\right)\quad\Leftrightarrow\quad\left(\frac{|\,x+y\,|}{2}\right)^a\leqslant\frac{|\,x\,|^a+|\,y\,|^a}{2}.$$

Пусть $\xi_i=\dfrac{1}{nh}Z_iK\left(\dfrac{u_i-x}{h}\right)$. Тогда $S_1=\sum\limits_{i=1}^{n}\xi_i$ и при $a=3$, имеем:

$$|\xi_i-\mathbf{E}\left(\xi_i\right)|^3\le 4\left(|\xi_i|^3+|\mathbf{E}\left(\xi_i\right)|^3\right).$$

Беря математическое ожидание, получим:

$$\mathbf{E}\left(\left(\xi_i-\mathbf{E}\left(\xi_i\right)\right)^3\right)\le 4\left(E\left(\xi_i^3\right)+\left(\mathbf{E}\left(\xi_i\right)\right)^3\right)\le 8\mathbf{E}\left(\xi_i^3\right).$$

Рассмотрим сумму $\sum\limits_{i=1}^{n}\mathbf{E}\left(\xi_i^3\right)$. Имеем:

$$\sum_{i=1}^{n}\mathbf{E}\left(\xi_i^3\right)=\frac{1}{n^3h^3}\sum_{i=1}^{n}\mathbf{E}\left(Z_i^3\right)K^3\left(\frac{u_i-x}{h}\right),$$

$$\mathbf{E}\left(Z_i^3\right)=\frac{r(r+\lambda)\,(r+2\lambda)}{(1-\lambda q)^3}+3\frac{r(r+\lambda)}{(1-\lambda q)^2}+\frac{r}{1-\lambda q},$$

$$\sum_{i=1}^{n}\mathbf{E}\left(\xi_i^3\right)=$$

$$=\frac{1}{n^3h^3}\sum_{i=1}^{n}\left(\frac{r(r+\lambda)(r+2\lambda)}{\left(1-\lambda\left(1-F\left(u_i\right)\right)\right)^3}+3\frac{r(r+\lambda)}{\left(1-\lambda\left(1-F\left(u_i\right)\right)\right)^2}+\frac{r}{\left(1-\lambda\left(1-F\left(u_i\right)\right)\right)}\right)\times$$

$$\times K^3\left(\frac{u_i-x}{h}\right)\sim\frac{1}{n^2h^2}\left(\frac{r(r+\lambda)(r+2\lambda)}{\left(1-\lambda\left(1-F\left(x\right)\right)\right)^3}+3\frac{r(r+\lambda)}{\left(1-\lambda\left(1-F\left(x\right)\right)\right)^2}+\right.$$

$$\left.+\frac{r}{\left(1-\lambda\left(1-F\left(x\right)\right)\right)}\right)\times\int\limits_{-\infty}^{+\infty}K^3(t)dt=\frac{C_1}{n^2h^2},\ \text{где } C_1=const.$$

Тогда для дроби Ляпунова получаем

$$L_n=\frac{\sum_{i=1}^{n}\mathbf{E}\left(\left(\xi_i-\mathbf{E}\left(\xi_i\right)\right)^3\right)}{\left(\sum_{i=1}^{n}\mathbf{D}\xi_i\right)^{3/2}}\leq\frac{C_1n^{3/2}h^{3/2}}{C_2n^2h^2}=\frac{C}{\sqrt{nh}}\underset{n\to\infty}{\longrightarrow}0.$$

Таким образом, условие центральной предельной теоремы Стейна-Тихомирова выполнено и

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

Далее воспользуемся дельта-методом и получим:

$$g(x)=\frac{r}{x},\quad g'(x)=-\frac{r}{x^2},\quad g'\left(\frac{r}{\lambda\left(1-\lambda\left(1-F(x)\right)\right)}\right)=-\frac{\lambda^2\left(1-\lambda(1-F(x))\right)^2}{r},$$

$$\left(g'\left(\frac{r}{\lambda\left(1-\lambda\left(1-F(x)\right)\right)}\right)\right)^2=\frac{\lambda^4\left(1-\lambda(1-F(x))\right)^4}{r^2}.$$

Для оценки $F_n(x)=1-\dfrac{1}{\lambda}+\dfrac{r}{\lambda^2S_1}$, выполнено: $F_n(x)\xrightarrow[n\to\infty]{d}F(x)$, поэтому

$$g\left(\theta_n\right)=g\left(\theta_0\right)+\left(\theta_n-\theta_0\right)g'\left(\theta_0\right)+O\left(\left(\theta_n-\theta_0\right)^2\right)\Rightarrow g\left(\theta_n\right)-g\left(\theta_0\right)=\left(\theta_n-\theta_0\right)g'\left(\theta_0\right),$$

$$\sqrt{nh}\left(g\left(\theta_n\right)-g\left(\theta_0\right)\right)\approx$$

$$\approx\sqrt{nh}\left(\theta_n-\theta_0\right)g'\left(\theta_0\right)\sim N\left(a,\left(g'\left(\theta_0\right)\right)^2\|K\|^2\frac{r\left(1-F(x)\right)}{\left(1-\lambda\left(1-F(x)\right)\right)^2}\right),$$

но

$$(g'(\theta_0))^2 \|K\|^2 \frac{r(1-F(x))}{(1-\lambda(1-F(x)))^2} =$$

$$= \frac{\lambda^4 (1-\lambda(1-F(x)))^4}{r^2} \cdot \|K\|^2 \cdot \frac{r(1-F(x))}{(1-\lambda(1-F(x)))^2} =$$

$$= \frac{\lambda^4 (1-\lambda(1-F(x)))^2 (1-F(x))}{r} \|K\|^2 ,$$

откуда

$$\sqrt{nh}\,(F_n(x) - E(F_n(x))) \xrightarrow[n\to\infty]{d} N\left(0, \frac{\lambda^4 (1-F(x))(1-\lambda(1-F(x)))^2}{r} \|K\|^2\right).$$

□

Определим теперь оценку квантиля $\xi_\alpha$ порядка $0<\alpha<1$ следующим образом:

$$\hat{\xi}_{n,\alpha} = \inf\{x \in R : \hat{F}_n(x) \geqslant \alpha\}.$$

**Теорема 3.** *Пусть $\hat{\xi}_{n,\alpha}$ – оценка квантиля порядка $0<\alpha<1$, выполнены условия регулярности (**K**), (**H**), (**F**), $u_i = \dfrac{i}{n}, i=1,...,n$ и $f(\xi_\alpha)>0$. Тогда*

$$\sqrt{nh}\left(\hat{\xi}_{n,\alpha} - \xi_\alpha - ah^2\right) \xrightarrow[n\to\infty]{d} N\left(0, \frac{(1-\alpha)(1-\lambda(1-\alpha))^2}{rf^2(\xi_\alpha)} \|K\|^2\right),$$

*где*

$$a = \frac{(1-\lambda(1-\alpha))\, f'(\xi_\alpha) - 2\lambda f^2(\xi_\alpha)}{2(1-\lambda(1-\alpha))\sigma} \nu_2(K).$$

*Доказательство.* Пусть

$$\sigma^2 = \frac{(1-\alpha)(1-\lambda(1-\alpha))^2}{r} \|K\|^2 \text{ и } \delta = \delta(x) = \xi_\alpha + \frac{x\sigma}{\sqrt{nh} f(\xi_\alpha)}.$$

Имеем:

$$\mathbf{P}\left(\frac{\sqrt{nh} f(\xi_\alpha)(\hat{\xi}_{n,\alpha} - \xi_\alpha)}{\sigma} \le x\right) = \mathbf{P}\left(\hat{\xi}_{n,\alpha} \le \delta\right) = \mathbf{P}\left(F_n(\delta) \ge \alpha\right) =$$

$$= \mathbf{P}\left(1 - \frac{1}{\lambda} + \frac{r}{\lambda^2 \frac{1}{n}\sum_{i=1}^n z_i \eta_i(\delta)} \ge \alpha\right) =$$

$$= \mathbf{P}\left(\frac{1}{n}\sum_{i=1}^n Z_i K_h(u_i - \delta) \le \frac{r}{\lambda(1-\lambda(1-\alpha))}\right) =$$

$$= \mathbf{P}\left(\frac{1}{n}\sum_{i=1}^n (Z_i K_h(u_i-\delta) - \theta_i) \le \frac{1}{n}\sum_{i=1}^n \left(\frac{r}{\lambda(1-\lambda(1-\alpha))} - \theta_i\right)\right) =$$

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

где
$$\theta_i=\mathbf{E}\left(Z_iK_h(u_i-\delta)\right)=\frac{r}{\lambda(1-\lambda(1-F(u_i)))}K_h(u_i-\delta).$$

Заметим, что $K_h(u-\delta)=\frac{1}{h}K\left(\frac{u-\delta}{h}\right)$, и если $\frac{u-\delta}{h}\notin[-1;1]$, то ядерная функция равна нулю. Учитывая, что $\delta=\xi_\alpha+\frac{x\sigma}{\sqrt{nh}f(\xi_\alpha)}$ и $\frac{1}{\sqrt{nh}}=h^2$, получим отрезок $\ell_\alpha=[\xi_\alpha-h+x\sigma h^2/f(\xi_\alpha);\xi_\alpha+h+x\sigma h^2/f(\xi_\alpha)]$, на котором значение $K_h(u-\delta)$ отлично от нуля. Функция $\frac{r}{\lambda(1-\lambda(1-F(u)))}>0$ является монотонно убывающей на $\ell_\alpha$, $\{u_i=i/n, i=1,2,...,n\}$ и вариация $\boldsymbol{V}\left(\frac{r}{\lambda(1-\lambda(1-F(u)))}\right)<\infty$. Исходя из того, что $K_h(u-\delta)$ имеет ограниченную вариацию и из [20], заключаем:

$$A_n=\frac{1}{n}\sum_{i=1}^{n}\theta_i=\frac{1}{n}\sum_{i=1}^{n}\frac{r}{\lambda(1-\lambda(1-F(u_i)))}K_h(u_i-\delta)=$$
$$=\int_{\ell_\alpha}\frac{r}{\lambda(1-\lambda(1-F(u)))}K_h(u-\delta)du+O\left(\frac{1}{\sqrt{nh}}\right).$$

Сделаем замену $t=\frac{u-\delta}{h}$, где $0\leq u\leq 1$. Тогда

$$A_n=\int_0^1\frac{r}{\lambda(1-\lambda(1-F(u)))}K_h(u-\delta)du=$$
$$=\int_{-\delta/h}^{(1-\delta)/h}\frac{r}{\lambda(1-\lambda(1-F(\xi_\alpha+\rho_1h)))}K(t)dt,$$

где
$\rho_1=t+\frac{x\sigma}{f(\xi_\alpha)}h$ и для достаточно большого $n$ ($n\geq n_1$),

$$A_n=\int_{-1}^{1}\frac{r}{\lambda(1-\lambda(1-F(\xi_\alpha+\rho_1h)))}K(t)dt.$$

Пусть $|x|\leq L$, где $L-$достаточно большое, тогда действуя как при доказательстве теоремы 1, имеем:

$$A_n=\int_{-1}^{1}\frac{r}{\lambda(1-\lambda(1-F(\xi_\alpha+\rho_1h)))}K(t)\,dt=\frac{r}{\lambda(1-\lambda(1-\alpha))}-$$
$$-\frac{r}{(1-\lambda(1-\alpha))^2}\int_{-1}^{1}\left(f(\xi_\alpha)\rho_1h++\frac{1}{2}f'(\xi_\alpha)\rho_1^2h^2-\frac{\lambda f^2(\xi_\alpha)}{1-\lambda(1-\alpha)}\rho_1^2h^2\right)K(t)dt+o(h^2).$$

Теперь, учитывая, что$\rho_1 = t + \dfrac{x\sigma}{f(\xi_\alpha)}h$ и $\int_{-1}^{1} tK(t)dt = 0$, получим:

$$A_n = \frac{r}{\lambda(1-\lambda(1-\alpha))} -$$
$$- \frac{r\sigma h^2}{(1-\lambda(1-\alpha))^2}\left(x + \frac{(1-\lambda(1-\alpha))f'(\xi_\alpha) - 2\lambda f^2(\xi_\alpha)}{2(1-\lambda(1-\alpha))\sigma}\nu_2\right) + o(h^2),$$
$$\frac{(1-\lambda(1-\alpha))^2}{r\sigma h^2}\left(A_n - \frac{r}{\lambda(1-\lambda(1-\alpha))}\right) +$$
$$+\left(x + \frac{(1-\lambda(1-\alpha))f'(\xi_\alpha) - 2\lambda f^2(\xi_\alpha)}{2(1-\lambda(1-\alpha))\sigma}\nu_2(K)\right) = o(1).$$

Отсюда следует, что последовательность

$$\frac{(1-\lambda(1-\alpha))^2}{r\sigma h^2}\left(A_n - \frac{r}{\lambda(1-\lambda(1-\alpha))}\right) +$$
$$+\left(x + \frac{(1-\lambda(1-\alpha))f'(\xi_\alpha) - 2\lambda f^2(\xi_\alpha)}{2(1-\lambda(1-\alpha))\sigma}\nu_2(K)\right)$$

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

Пусть $\Lambda_n(x) = \dfrac{1}{n}\sum\limits_{i=1}^{n}(Z_i K_h(u_i - \delta) - \theta_i)$. Покажем, что

$$\frac{\sqrt{nh}(1-\lambda(1-\alpha))^2}{r\sigma}\cdot \Lambda_n(x) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{\lambda^4}\right).$$

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

$$\mathbf{D}(\Lambda_n(x)) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}\left(Z_i K_h(u_i-\delta)\right) = \frac{1}{n^2}\sum_{i=1}^{n}K_h^2(u_i-\delta)\cdot\mathbf{D}\left(Z_i\right) =$$
$$= \frac{1}{n^2h^2}\sum_{i=1}^{n}K^2\left(\frac{u_i-\delta}{h}\right)\frac{r(1-F(u_i))}{\left(1-\lambda\left(1-F(u_i)\right)\right)^2} =$$
$$= \frac{1}{nh^2}\int_{-1}^{1}\frac{r(1-F(u))}{\left(1-\lambda\left(1-F(u)\right)\right)^2}K^2\left(\frac{u-\delta}{h}\right)du(1+o(1)) \approx \frac{r\left(1-\alpha\right)}{nh(1-\lambda(1-\alpha))^2}\left\|K\right\|^2$$

равномерно по $|x| \le L$, откуда

$$\lim_{n\to\infty}\mathbf{D}\left(\frac{\sqrt{nh}(1-\lambda(1-\alpha))^2}{r\sigma}\cdot\Lambda_n(x)\right) = \frac{1}{\lambda^4}.$$

Условия Ляпунова проверяются как при доказательстве теоремы 1, а следовательно для $|x| \le L$ получаем:

$$\frac{\sqrt{nh}(1-\lambda(1-\alpha))^2}{r\sigma}\cdot \Lambda_n(x) \xrightarrow[n\to\infty]{d} N\left(0; \frac{1}{\lambda^4}\right).$$

$\square$

### Заключение и выводы

В работе рассмотрена задача статистического оценивания неизвестной функции распределения $F(x)$ и квантиля $\xi_\lambda$ порядка $0 < \lambda < 1$ функции распределения $F(x)$ в модели отрицательной $\lambda$-биномиальной регрессии, которая как частные случаи включает модель отрицательной биномиальной регрессии и пуассоновской регрессии. Эти оценки состоятельны и асимптотически нормальны, что позволяет при больших объемах выборки рассчитывать доверительные интервалы для значений функции распределения, однако оценка квантиля (эффективных доз) в основном сохраняет свойства смещения и дисперсии. Однако предложенная здесь оценка более устойчива при оценивании экстремальных квантилей.

## Список литературы

[1] Razzaghi M. Statistical Models in Toxicology. New York: Chapman and Hall/CRC, 2020. 288 p.

[2] Криштопенко С.В., Тихов М.С., Попова Е.Б. Доза-Эффект. М.: Медицина, 2008. 228 с.

[3] Вараксин А.Н., Казмер Ю.И., Кацнельсон Б.А., Панов В.Г., Привалов Л.И. Применение бинарной теории к оценке типа комбинированного действия загрязнителей атмосферного воздуха на респираторное здоровье детей // Медицинская информатика. 2011. № 3. С. 36–44.

[4] Михайлюк А.Н., Солод Р.А. Уточнение размеров наступления половой зрелости у пиленгаса Liza haematocheilus с использованием расширенных возможностей пробит-метода // Водные биоресурсы и среда обитания. 2019. Т. 2, № 1. С. 47–52.

[5] 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.

[6] 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

[7] Хардле В. Прикладная непараметрическая регрессия. М.: Мир, 1993. 349 с.

[8] Тихов М.С., Криштопенко Д.С. Выбор параметра сглаживания несмещенной оценки функции эффективностим в зависимости доза-эффект // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр.. Пермь: Пермский университет, 2010. С. 85–95.

[9] Тихов М.С., Криштопенко Д.С. Оценивание распределений в зависимости доза-эффект при фиксированном плане эксперимента // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр.. Пермь: Пермский университет, 2006. С. 66–77.

[10] Okumura H., Naito K. Weighted kernel estimators in nonparametric binomial regression // Journal of Nonparametric Statistics. 2004. Vol. 16, № 2. Pp. 39–62.

[11] Tikhov M.S. Negative binomial regression in dose-effect relationship // The 5th International Conference on Stochastic Methods, ICSM-5. Peoples Friendship University of Russia, 2020. Pp. 205–208.

[12] Абрамовиц М., Стиган И. Справочник по специальным функциям. М.: Наука, 1979. 832 с.

[13] Кендалл М., Стьюарт А. Теория распределений. М.: Наука, 1966. 588 с.

[14] Джонсон Н., Коц С., Кемп А. Одномерные дискретные распределения. М.: БИНОМ, 2010. 560 с.

[15] Stein Ch. Approximation Computation of Expectations. Vol. 7. Hayward: Institute of Mathematical Statistics, 1986. 180 p.

[16] Форманов Ш.К. О методе Стейна-Тихомирова и его приложениях в неклассических предельных теоремах // Дискретная математика. 2007. Т. 19, № 1. С. 27–39.

[17] Stein C. A bound for the error in the normal approximation to the distribution of a sum ofdependent random variables // Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. 1972. Pp. 586–602.

[18] Chen L., Goldstein L., Shao Q.M. Normal Approximation by Stein's Method. New York: Springer, 2011. 405 p.

[19] Натансон И.П. Теория функций вещественной переменной. М.: Наука, 1974. 480 с.

[20] Niederreiter H. Random Number Generation and Quasi-Monte Carlo-Method. Pennsylvania: SIAM Philadelphia, 1992. 241 p.

[21] Тихов М.С. Непараметрическое оценивание эффективных доз по данным бинарных откликов // Уфимский математический журнал. 2013. Т. 5, № 2. С. 94–108.

## Образец цитирования

Тихов М.С. Отрицательная $\lambda$-биномиальная регрессия в зависимости доза-эффект // Вестник ТвГУ. Серия: Прикладная математика. 2022. № 4. С. 53–75. https://doi.org/10.26456/vtpmk649

## Сведения об авторах

1. **Тихов Михаил Семенович**
профессор кафедры теории вероятностей и анализа данных 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: mikhail.tikhov@itmm.unn.ru*

## NEGATIVE $\lambda$-BINOMIAL REGRESSION IN DOSE-EFFECT RELATIONSHIP

**Tikhov Mikhail Semenovich**
Professor of Probability Theory and Data Analysis department,
Institute of Information Technology, Mathematics and Mechanics,
Lobachevsky State University of Nizhniy Novgorod
*Russia, 603950, Nyzhniy Novgorod, 23 Gagarin av., UNN.*
*E-mail: mikhail.tikhov@itmm.unn.ru*

*Received 14.09.2022, revised 12.12.2022.*

This paper is concern to the problem of estimating the distribution function and its quantiles in the dose-effect relationships with nonparametric negative $\lambda$-binomial regression. Here, a kernel-based estimators of the distribution function are proposed, of which kernel is weighted by the negative $\lambda$-binomial random variable at each covariate. Our estimates are consistent, that is, they converge to their optimal values in probability as $n$, the number of observations, grow to infinity. It is shown that these estimates have a smaller asymptotic variance in comparison, in particular, with estimates of the Nadaray-Watson type and other estimates. Nonparametric quantiles estimators obtained by inverting a kernel estimator of the distribution function are offered. It is shown that the asymptotic normality of this bias-adjusted estimator holds under some regularity conditions. In the first part, the relations between the moments of the negative $\lambda$-binomial distribution are analyzed. A new characterization of the Poisson distribution is obtened.

**Keywords:** negative $\lambda$-binomial response model, effective dose level, nonparametric estimate.

### Citation

Tikhov M.S., “Negative $\lambda$-binomial regression in dose-effect relationship”, *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2022, № 4, 53–75 (in Russian). https://doi.org/10.26456/vtpmk649

### References

[1] Razzaghi M., *Statistical Models in Toxicology*, Chapman and Hall/CRC, New York, 2020, 288 pp.

[2] Krishtopenko S.V., Tikhov M.S., Popova E.B., *Dosa-Effect [Dose-Effect]*, Medicina Publ., Moscow, 2008 (in Russian), 228 pp.

[3] Varaksin A.N., Kazmer Yu.I., Katsnelson B.A., Panov V.G., Privalov L.I., "Application of binary theory to the assessment of the type of combined effect of atmospheric pollutants on the respiratory health of children", *Meditsinskaya informatika [Medical Informatics]*, 2011, № 3, 36–44 (in Russian).

[4] Mikhajlyuk A.N., Solod R.A., "Clarification of the size of the onset of puberty in the hare Liza haematocheilus using extended capabilities of the probit method", *Vodnye bioresursy i sreda obitaniya [Aquatic bioresources and habitat]*, **2**:1 (2019), 47–52 (in Russian).

[5] 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).

[6] 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.

[7] Härdle W., *Applied Nonparametric Regression*, Cambridge University Press, New York, 1990, 333 pp.

[8] Tikhov M.S., Krishtopenko D.S., "On the choice of smoothing parameter for unbiased estimates in dose-effect dependence with fixed experiment plan", *Statisticheskie metody otsenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. [Statistical methods of evaluation and hypothesis testing: inter-university collection of scientific works]*, Perm University, Perm, 2010, 85–95 (in Russian).

[9] Tikhov M.S., Krishtopenko D.S., "Estimation of distribution under dose-effect dependence with fixed experiment plan", *Statisticheskie metody otsenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. [Statistical methods of evaluation and hypothesis testing: inter-university collection of scientific works]*, Perm University, Perm, 2006, 66–77 (in Russian).

[10] Okumura H., Naito K., "Weighted kernel estimators in nonparametric binomial regression", *Journal of Nonparametric Statistics*, **16**:2 (2004), 39–62.

[11] Tikhov M.S., "Negative binomial regression in dose-effect relationship", *The 5th International Conference on Stochastic Methods*, ICSM-5, Peoples Friendship University of Russia, 2020, 205–208.

[12] Abramowitz M., Stegun I., *Handbook of Mathematical Functions*, Dover Publications, New York, 1965, 1046 pp.

[13] Kendall M., Stuart A., *The Advanced Theory of Statistics*, Charles Griffin, London, 1945, 457 pp.

[14] Johnson N., Kotz S., Kemp A., *Univariate Discrete Distributions*, John Wiley and Sons, New York, 2005, 672 pp.

[15] Stein Ch., *Approximation Computation of Expectations*. V. 7, Institute of Mathematical Statistics, Hayward, 1986, 180 pp.

[16] Formanov Sh.K., “On the Stein-Tikhomirov method and its applications in nonclassical limit theorems”, *Diskretnaya matematika [Discrete mathematics]*, **19**:1 (2007), 27–39 (in Russian).

[17] Stein C., “A bound for the error in the normal approximation to the distribution of a sum ofdependent random variables”, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1972, 586–602.

[18] Chen L., Goldstein L., Shao Q.M., *Normal Approximation by Stein's Method*, Springer, New York, 2011, 405 pp.

[19] Natanson I.P., *Theory of Functions of a Real Variable*, Dover Publications, New York, 2016, 544 pp.

[20] Niederreiter H., *Random Number Generation and Quasi-Monte Carlo-Method*, SIAM Philadelphia, Pennsylvania, 1992, 241 pp.

[21] Tikhov M.S., “Neparametricheskoe otsenivanie effektivnykh doz po dannym binarnykh otklikov”, *Ufimskij matematicheskij zhurnal [Ufa Mathematical Journal]*, **5**:2 (2013), 94–108 (in Russian).