

Ядерные оценки симметричной функции распределения в модели бинарного отклика

Серикова А.(Казахстан),
Тихов М.С. (Россия)

г. Нижний Новгород, Россия

**Нижегородский Государственный Университет им. Н.И.Лобачевского,
Институт Информационных Технологий, Математики и Механики,
Кафедра Теории Вероятностей и Анализа Данных**

E-mail: september415@inbox.ru, tikhovm@mail.ru
mikhail.tikhov@itmm.unn.ru

Рассматривается задача оценивания ф. р. $F(x)$ в модели «доза-эффект». Именно, пусть $X_1, X_2, \dots, X_n$ – независимые и одинаково распределенные (н.о.р.) случайные величины (с.в.) с неизвестной непрерывной функцией распределения (ф.р.) $F(x)$ и плотностью $f(x)$, $U_1, U_2, \dots, U_n$ – н.о.р. с.в., независимые от X_i – ов, с неизвестной ф.р. $G(x)$ и плотностью распределения $g(x) > 0$. Мы наблюдаем $U^{(n)} = \{(u_i, w_i), i = 1, 2, \dots, n\}$, где $W_i = I(X_i < U_i)$ есть индикатор события $(X_i < U_i)$. Рассматривается задача оце-

нивания функции распределения $F(x)$ для фиксированного значения x на основе интервальных цензурированных данных $U^{(n)}$.

Отметим, что если X и U независимы, то

$$E(W | U = x) = F(x), \quad (1)$$

т.е. задача оценивания сводится в этом случае к задаче оценивания регрессии. Если X и U зависимы, то функция (1) может быть и немонотонной (**парадоксальная токсичность**).

В качестве оценки $F(x)$ предлагается статистика

$$F_n(x) = \frac{S_{2n}(x)}{S_{1n}(x)}, \quad S_{2n}(x) = \frac{1}{nh} \sum_{j=1}^n w_j K\left(\frac{u_j - x}{h}\right),$$
$$S_{1n}(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{u_j - x}{h}\right).$$

*Это – ядерная оценка типа **Надарая-Ватсона (NW)**.*

При некоторых условиях она является состоятельной и асимптотически нормальной оценкой.

ТЕОРЕМА 1. *Предположим $h = n^{-1/5}$, $F(x)$ и $g(x) > 0$ – трижды непрерывно дифференцируемы. Тогда*

(i) $F_n(x)$ *сходится по вероятности в точке x к ф. р. $F(x)$;*

(ii) $\frac{\sqrt{nh} (F_n(x) - F(x))}{b(x)\|K\|}$ *сходится по распределению*

в точке x к нормальной с. в. со средним $a(x) \neq 0$ и единичной дисперсией, где

$$b^2(x) = \frac{F(x)(1-F(x))\|K\|^2}{g(x)}, \quad a(x) = \frac{f'(x) + 2f(x)g'(x)}{g(x)},$$

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

Сдвиг $a(x)$ можно уменьшить за счет построения двухшаговых оценок (см. M.S.Tikhov, I.S.Dolgih, *Asymptotically unbiased estimates of a distribution function under dose-effect dependence. Journal of Mathematical Sciences. 2015. Vol. 205. № 1. С. 113-120.*)

В предельной дисперсии $b(x) = \frac{F(x)(1-F(x))\|K\|^2}{g(x)}$

в знаменателе стоит плотность $g(x)$.

В качестве ядра K обычно берут

1) $K_1(x) = \frac{3}{4}(1-x^2)I(|x| \leq 1)$, (Ядро Епанечникова)

2) $K_2(x) = \frac{15}{16}(1-x^2)^2 I(|x| \leq 1)$. (Квартическое ядро)

3) $K_3(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right) I(|x| \leq 1)$. (Косинус ядро)

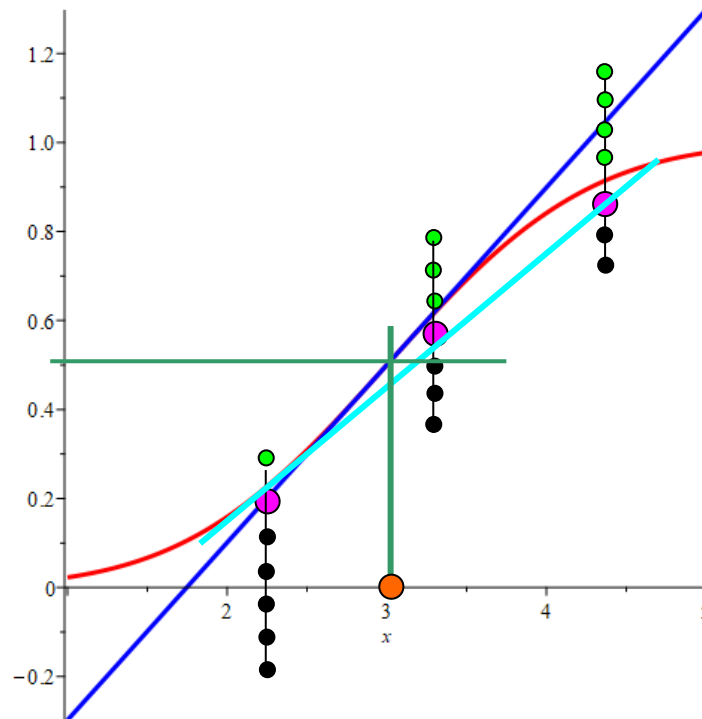
4) $K_4(x) = \exp\left(\frac{x^2}{x^2-1}\right) I(|x| \leq 1)$. (Ядро Соболева)

Если в выборке $U^{(n)} = \{(W_i, u_i), i = 0, 1, \dots, n\}$, $W_i = I(X_i < u_i)$ величины u_i не случайны (фиксированный план испытаний), то $\sigma_1^2(x) = F(x)(1-F(x))\|K\|^2$.

Одна из **интерпретаций** данной ситуации следующая: величина U — введенная в организм доза какого-либо вещества, X — нижняя граница, с которой начинается реакция организма. Если $(U > X)$, то есть реакция (эффект) и $W = I(U > X) = 1$, в противном случае $W = 0$.

Как оценивается медианная доза?

(см. Л.Закс. Статистическое оценивание. М.: Статистика, 1976. — 598 с., 2.3, с.205.)



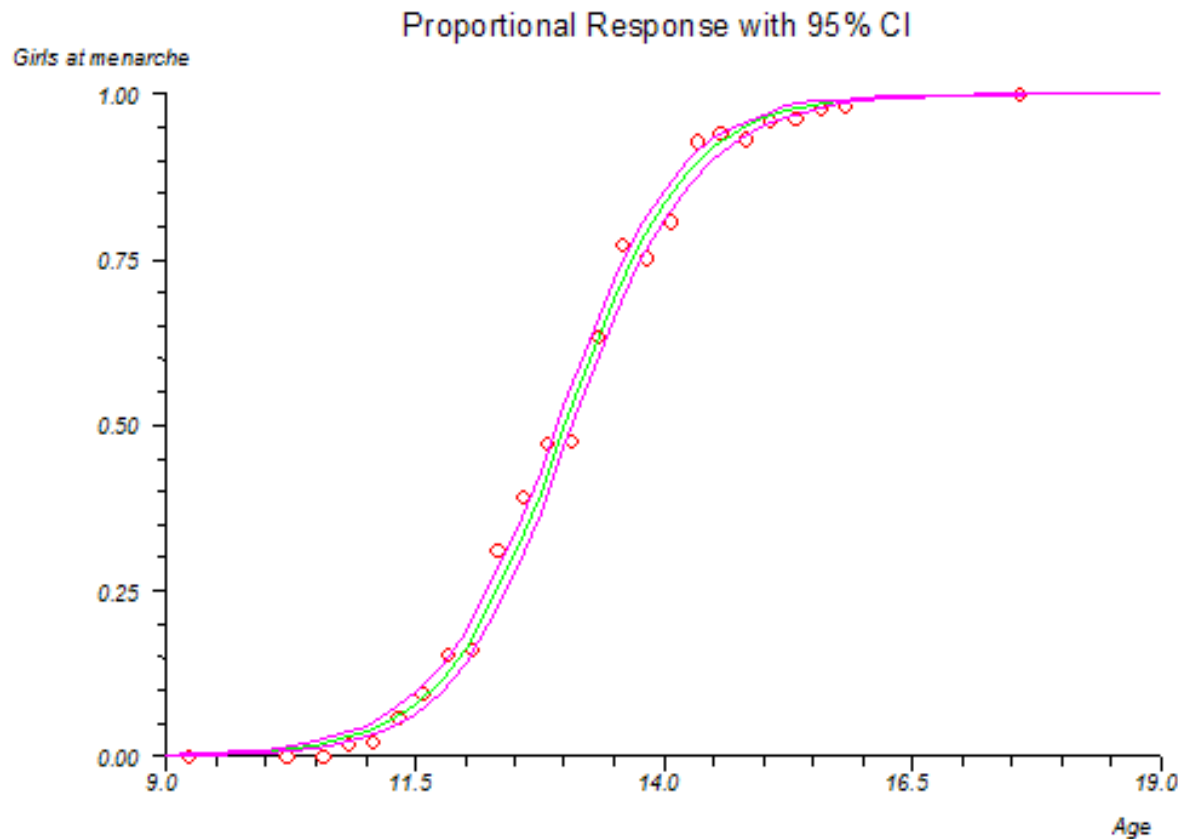
Пример Finney D. Probit-Analysis (1971, p. 98). ($N = 3918$)

Age $U := [9.21, 10.21, 10.58, 10.83, 11.08, 11.33, 11.58, 11.83, 12.08, 12.33, 12.58, 12.83, 13.08, 13.33, 13.58, 13.83, 14.08, 14.33, 14.58, 14.83, 15.08, 15.33, 15.58, 15.83, 17.58]$

Menses $W := [0, 0, 0, 2, 2, 5, 10, 17, 16, 29, 39, 51, 47, 67, 81, 88, 79, 90, 113, 95, 117, 107, 92, 112, 1049];$

Girls $M := [376, 200, 93, 120, 90, 88, 105, 111, 100, 93, 100, 108, 99, 106, 105, 117, 98, 97, 120, 102, 122, 111, 94, 114, 1049];$

Probit analysis - logit sigmoid curve **Median Dose** = 13.006622. **Confidence interval** = 12.930535 to 13.082483



NW-ядерная оценка. Медианная доза

$$\hat{m}_e = 12.952231,$$

Доверительный интервал (12.936231, 12.968231).

Длина $l_K = 0.032$

Probit: **Median Dose** = 13.006622.

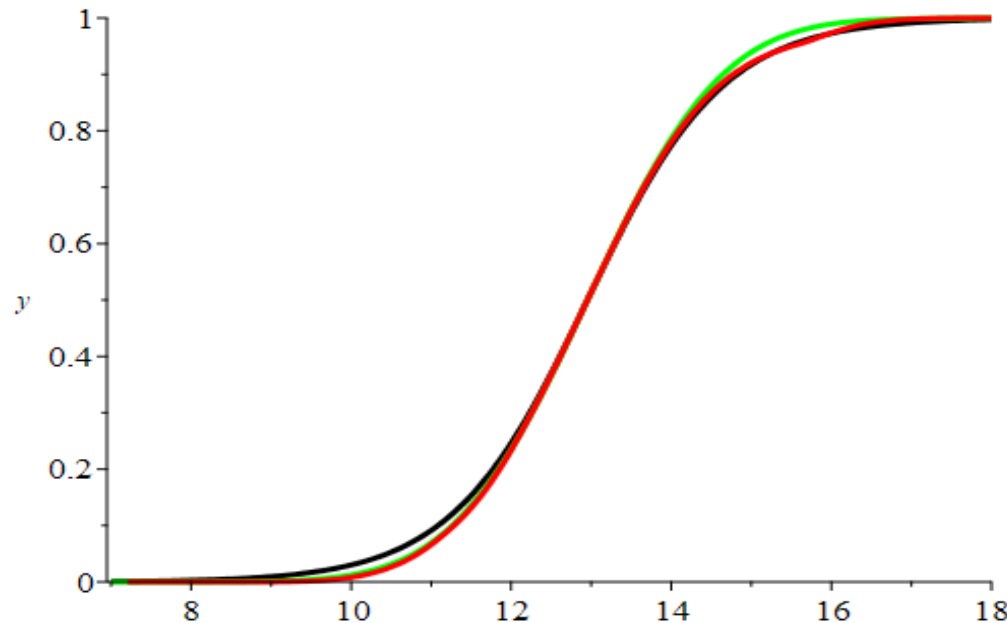
Confidence interval = 12.930535 to 13.082483

Длина $l_P = 0.152$.

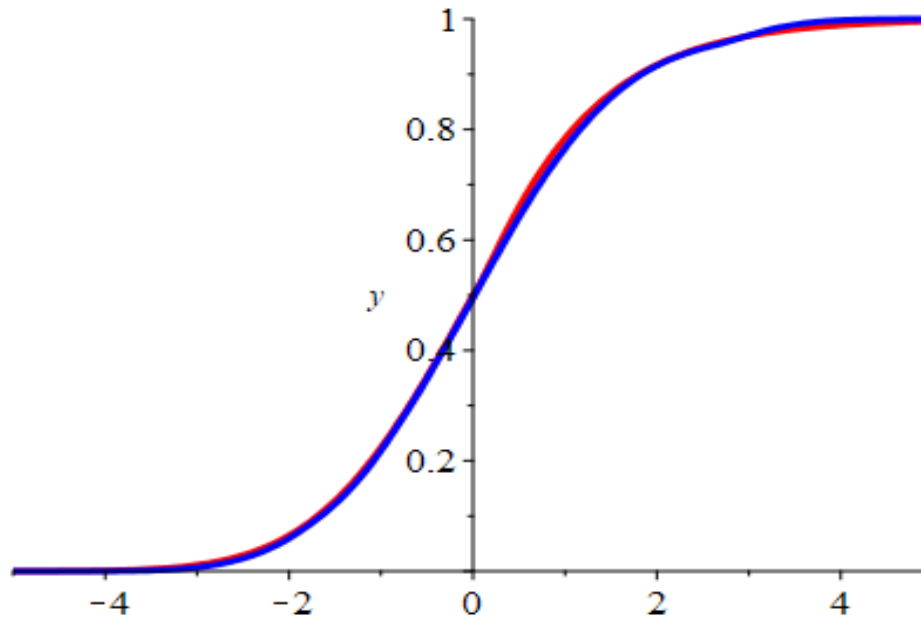
Есть предположение, что предельное

Распределение является нормальным, а для него математическое ожидание является и медианой, оно симметрично относительно медианы.

Распределение симметрично относительно точки a , если $F(x+a) = 1 - F(a-x)$, $x \in \mathbf{R}$.



Красная линия – оценка $F_n(x)$. Черная линия –
 $GL(x) = \left(1 + \exp\left(-(x - 12.952) / 0.85\right)\right)^{-1}$ (логистическое распределение)
Зеленая линия $N(12.952; 1.32)$ (нормальное распределение)



Красная линия – оценка $F_n(x)$. $\sigma^2 = 1.32$, $a = 1.2$, $b = 0.95$.

Синяя линия –
$$\frac{1}{2} \left(\frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \cdot I(x < 0) + \frac{x^{a-1}}{b^a \Gamma(a)} \cdot \exp\left(-\frac{x}{b}\right) \cdot I(x \geq 0) \right)$$

Если распределение симметрично (относительно нуля), а u_i — неслучайны, то в качестве оценки ф.р. возьмем

$$F_n^*(x) = \frac{1}{2} (S_{2n}(x) + 1 - S_{2n}(-x)).$$

Имеем

$$\mathbf{E}(F_n^*(x)) = F(x) + \frac{h^2}{2} \mu_2(K) f'(x) + O(h^4),$$

$$\mu_2(K) = \int_{-1}^1 z^2 K(z) dz,$$

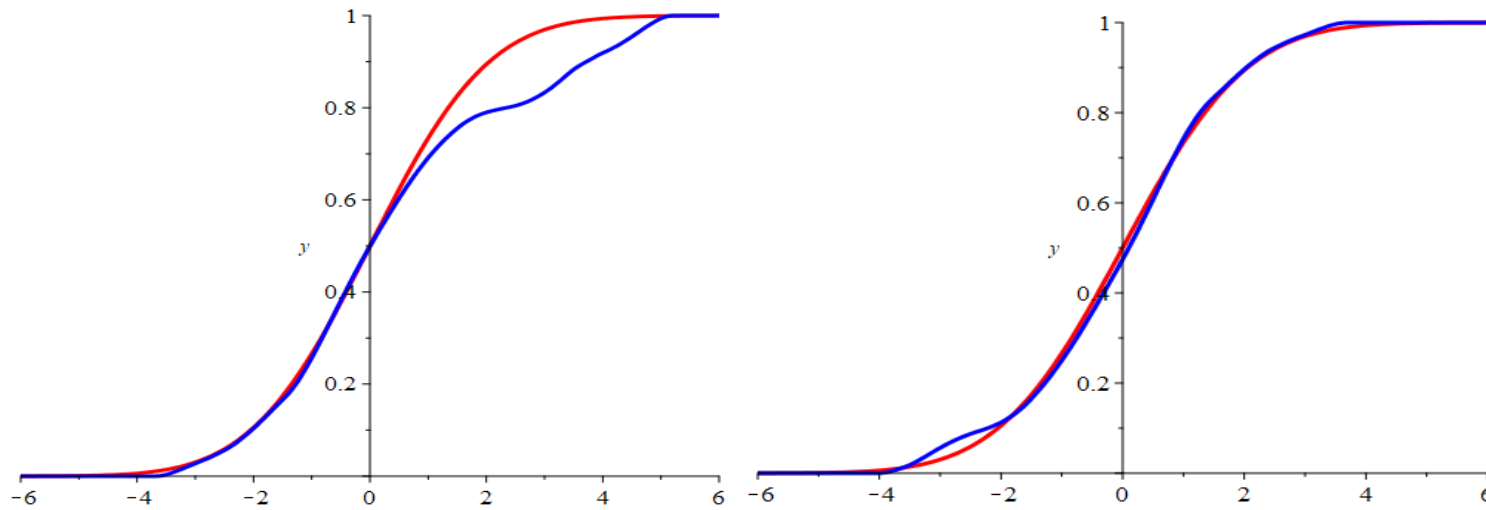
$$\mathbf{D}(F_n^*(x)) = \frac{\sigma_0^2(x)}{\sqrt{nh}} (1 + o(1)), \quad \sigma_0^2(x) = \frac{1}{2} \sigma^2(x).$$

ТЕОРЕМА. Пусть выполнены условия теоремы 4.2.1 [1]. Тогда

$$(i) \quad F_n^{(*)}(x) \xrightarrow[n \rightarrow \infty]{p} F(x);$$

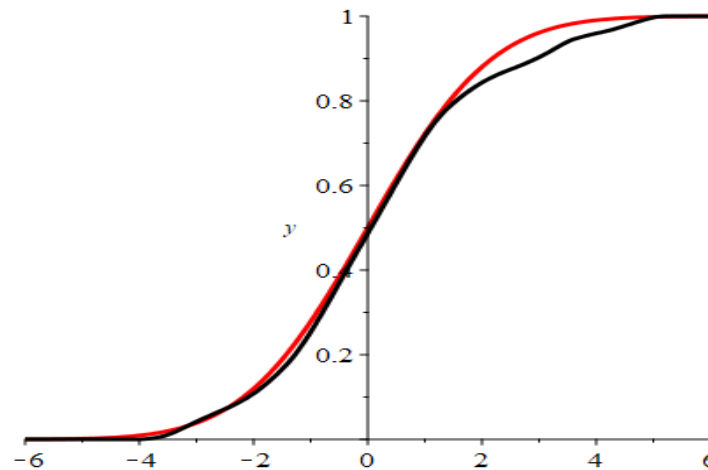
$$(ii) \quad \sqrt{nh} (F_n^{(*)}(x) - \mathbf{E}(F_n^{(*)}(x))) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_0^2(x)).$$

- [1] M.S. Tikhov, *Statistical estimation based on interval censored data*, Parametric and semiparametric models with applications to reliability, survival analysis and quality of life, Stat. Ind. Technol., Birkhauser Boston, MA, 2004, 211-218.



1. Оценка $S_{2n}(x)$

2. Оценка $1 - S_{2n}(-x)$

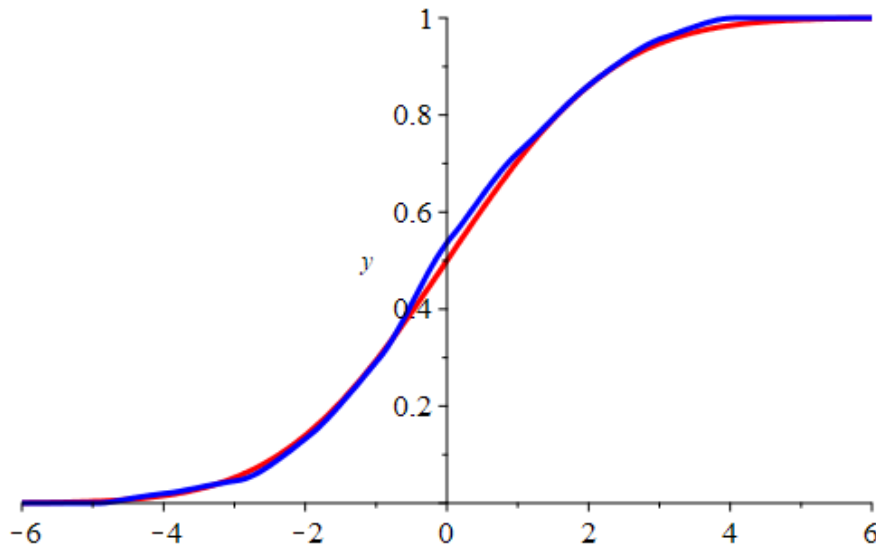


3. Оценка $F_n^*(x) = (S_{2n}(x) + 1 - S_{2n}(-x)) / 2$. (Моделированные данные)

Т а б л и ц а 10 .

Соотношение зрелых и незрелых лещей каждого возраста в уловах
 Рыбинского водохранилища по Остроумову, 1955 (1956)

	В о з р а с т														
	3 +	4 +	5 +	6 +	7 +	8 +	9 +	10 +	11 +	12 +	13 +	15 +	17 +	Итого	<i>n</i>
Самцы:															
незрелые	1	5	1	4	27	12	6	6	6	—	—	—	—	100,0	68
зрелые	—	—	—	—	1	4	3	11	38	13	5	2	1	100,0	78
Самки:															
незрелые	—	5	1	5	27	18	11	9	18	4	—	—	—	100,0	98
зрелые	—	—	—	—	1	1	4	6	19	7	3	1	1	100,0	39



**Чугунова Н.И. Руководство по изучению возраста и роста
рыб. М.: Изд-во АН СССР, 1959 – 164 с.**

САМЦЫ Медиана $me = 10.0$

Доверительный интервал 10.0 ± 3.9

Длина $L = 0.111 + 5.88t - 0.187t^2$

САМКИ Медиана $me = 9.5$

Доверительный интервал 9.5 ± 3.5

Ю. Герасимов, С. Бражник, А. Стрельников, Динамика структурных показателей популяции леща *ABRAMIS BRAMA (CYPRINIDAE)* Рыбинского водохранилища за период 1954-2007гг. **Вопросы ихтиологии**, 2010, т.50, №4, с.515-525.

**Возраст половозрелых особей (медиана)
95% доверительный интервал**

<i>Возраст, лет</i>	Годы			
	1950-е	1970-е	1990-е	2000-е
3+ – 14+	10.2 ± 3.9	8.2 ± 3.9	4.0 ± 3.1	5.1 ± 4.3

СПИСОК ЛИТЕРАТУРЫ.

- [1] M.S. Tikhov, *Statistical estimation based on interval censored data*, Parametric and semiparametric models with applications to reliability, survival analysis and quality of life, Stat. Ind. Technol., Birkhauser Boston, MA, 2004, 211-218.
- [2] M. Tikhov, D. Krishtopenko, M. Yaroschuk, Estimation of distributions under dose-effect dependence with fixed experiment plan, Journal Math. Sciences, **220**:6 (2017), 753-762.
- [3] H. Niderreiter, Random number generation and quasi-Monte Carlo method, SIAM, Philadelphia, 1992, 241 p.

**СПАСИБО
ЗА ВНИМАНИЕ!**