ОЦЕНИВАНИЕ ПРОИЗВОДНЫХ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ПО НЕПОЛНЫМ НАБЛЮДЕНИЯМ

К. С. Сагидуллаев

Национальный университет Узбекистана имени Мирзо Улугбека
Ташкент, Узбекистан
E-mail: khalmurza@mail.ru

В работе в модели случайного цензурирования справа установлены свойства равномерной строгой состоятельности ядерных оценок производных функции распределения.

Ключевые слова: случайное цензурирование, степенные оценки, ядерные оценки.

Пусть $\{X_j, j \geq 1\}$ – последовательность независимых и одинаково распределённых (н.о.р.) случайных величин (с.в.) с функцией распределения (ф.р.) F. Пусть существуют непрерывные и ограниченные производные $F^{(m)}(t) = d^mF(t)/dt^m$, $m = 1, 2, \ldots, r + 1$ и $f = F^{(1)}$ – плотность. Рассмотрим случай, когда исходная последовательность с.в. подвергается случайному цензурированию справа другой независящей от неё последовательностью $\{Y_j, j \geq 1\}$ – н.о.р.с.в. с ф.р. G. Наблюдению доступна выборка $C^{(n)} = \{(Z_j, \delta_j), j = 1, \ldots, n\}$, где $Z_j = \min(X_j, Y_j)$, $\delta_j = I(X_j \leq Y_j)$ и $I(A)$ – индикатор события A. В выборке $C^{(n)}$, представляющие интерес с.в. X_j наблюдаемы лишь при $\delta_j = 1$. Такая статистическая модель встречается в медико-биологических и инженерных испытаниях, в страховом деле, где X_j означают продолжительность жизни испытываемого на выживаемость объекта (индивидуума или технического устройства), а Y_j – время прекращения испытания по другим причинам. В такой практической ситуации задача заключается в оценивании по выборке $C^{(n)}$ функциональных характеристик ф.р. F при мешающей ф.р. G. В данной работе рассмотрим задачу оценивания производных $F^{(m)}$ по выборке $C^{(n)}$. Следует отметить, что в случае полной выборки (т.е. при $P(Y_j = +\infty) = 1$ или $G(t) \equiv 0$) такие задачи рассматривались авторами [3–5]. В специальной модели случайного цензурирования справа – модели пропорциональных интенсивностей, когда ф.р. G степенным образом выражается через ф.р. F оценки функций $F^{(m)}$ были исследованы в работе [2]. В данной работе установлены свойство равномерной строгой состоятельности ядерных оценок производных $F^{(m)}$. Для построения ядерных оценок используем степенную оценку из [1]. Для введения этой оценки нам необходимы некоторые определения и обозначения. Далее предположим, что ф.р. G также является непрерывной. Легко видеть, что ф.р. $H(t) = P(Z_j \leq t)$ имеет представление

$$H(t) = 1 - (1 - F(t))(1 - G(t)).$$

Определим субраспределения $H^{(l)}(t) = P(Z_j \leq t, \delta_j = l)$, $l = 0, 1$. Тогда легко видеть, что $H^{(0)}(t) + H^{(1)}(t) = H^{(t)}$, $t \in R$ и $H^{(0)}(t) = \int_{-\infty}^{t} \left(1 - F(u)\right) dG(u)$, $H^{(1)}(t) = \int_{-\infty}^{t} \left(1 - G(u)\right) dF(u)$. Через Λ_F, Λ_G и Λ_H обозначим интегральные функции интенсивности (и.ф.и.), соответствующие ф.р. F, G и H, где

$$\Lambda_F(t) = -\log\left(1-F(t)\right) = \int_{-\infty}^{t} \frac{dH^{(1)}(u)}{1-H(u)},$$
$$\Lambda_G(t) = -\log\left(1-G(t)\right) = \int_{-\infty}^{t} \frac{dH^{(0)}(u)}{1-H(u)}, \qquad (1)$$
$$\Lambda_H(t) = -\log\left(1-H(t)\right) = \Lambda_F(t) + \Lambda_G(t).$$

Пусть

$$H_n^{(l)}(t) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \le t,\ \delta_j = l),\ \ l = 0,\ 1,$$
$$H_n(t) = H_n^{(0)}(t) + H_n^{(1)}(t) = \frac{1}{n}\sum_{j=1}^{n} I\left(Z_j \le t\right)$$

– эмпирические оценки, соответствующие $H^{(l)}(t)$, l = 0, 1 и $H(t)$. Тогда естественными оценками и.ф.и. (1) соответственно являются

$$\Lambda_F^{(n)}(t) = \int_{-\infty}^{t} \frac{dH_n^{(1)}(u)}{1-H_n(u)},\ \ \Lambda_G^{(n)} = \int_{-\infty}^{t} \frac{dH_n^{(0)}(u)}{1-H_n(u)},$$
$$\Lambda_H^{(n)}(t) = \Lambda_F^{(n)} + \Lambda_G^{(n)} = \int_{-\infty}^{t} \frac{dH_n(u)}{1-H_n(u)}.$$

Автором [1] была предложена и исследована следующая степенная оценка для F:

$$F_n(t) = 1-[1-H_n(t)]^{R_n(t)} = \begin{cases} 0, & t < Z_{(1)}, \\ 1-\left(\dfrac{n-k}{n}\right)^{R_n(t)}, & Z_{(k)} \le t < Z_{(k+1)}, \quad k = 1,\dots,n-1, \\ 1, & t \ge Z_{(n)}, \end{cases}$$

где $Z_{(1)} \le \ldots \le Z_{(n)}$ – вариационный ряд, соответствующий наблюдениям $\{Z_j,\ j = 1,\ \ldots,\ n\}$ и $R_n(t) = \Lambda_F^{(n)}(t) \big/ \Lambda_H^{(n)}(t)$.

Пусть $K(t)$ – абсолютно непрерывная ф.р. (ядро) с плотностью $k(t)$ и существуют непрерывные производные $K^{(m)}(t) = d^m K(t)/dt^m$, m = 1, 2, …, r + 1, где $k = K^{(1)}$ и $k^{(m-1)} = K^{(m)}$. Введём также и последовательность «ширины окна» $\{a_n,\ n \ge 1\}$. Нами будут использованы условия:

(У1) При $n \to \infty$ $a_n \downarrow 0$, $na_n^{2m} \to \infty$ при каждом m = 1, …, r + 1;

(У2) $\alpha = \int_{-\infty}^{\infty} |t|\, dK(t) < \infty$;

(У3) Функции $\{K^{(m)},\ m = 1,\ \ldots,\ r+1\}$ непрерывны и имеют ограниченные вариации $\{v_m = Var\{K^{(m)}\},\ m = 1,\ \ldots,\ r+1\}$;

(У4) $\lambda_m = \int_{-\infty}^{\infty} |K^{(m)}(t)|\, dt < \infty,\ m = 1,\ \ldots,\ r+1$.

Для производных $F^{(m)}(t)$ рассмотрим следующие «усечённые» оценки ядерной структуры:

$$F_n^{(m)}(t) = \int_{-\infty}^{\infty} I(\tau_n \ge y) a_n^{-m} K^{(m)}\left(\frac{t-y}{a_n}\right) dF_n(y),\quad t \in R, \qquad (2)$$

где $\{\tau_n, n \geq 1\}$ – последовательность чисел, стремящихся к $+\infty$ с определённой скоростью при $n \to \infty$.

Для исследования оценок (2) введём следующий вспомогательный функционал вида свёртки ф.р. K и F:

$$\tilde{F}_n(t) = \int_{-\infty}^{\infty} K\left(\frac{t-y}{a_n}\right) dF(y), \quad t \in R, \tag{3}$$

и его усечённый вариант

$$\tilde{F}_{n\tau_n}(t) = \int_{-\infty}^{\infty} I(\tau_n \geq y) K\left(\frac{t-y}{a_n}\right) dF(y), \quad t \in R. \tag{4}$$

Нетрудно заметить, что ввиду условия (У2), интегрированием по частям при $n \to \infty$ имеем

$$\sup_{t\in R} |\tilde{F}_n(t) - F(t)| \leq \sup_{t\in R} \int_{-\infty}^{\infty} |F(t - a_n u) - F(t)| k(u)\, du \leq \alpha \sup_{t\in R} |f(t)| a_n = \mathrm{O}(a_n). \tag{5}$$

Учитывая близость функций $\tilde{F}_n$ и F в (5), в качестве промежуточных приближений для $F^{(m)}$, $m = 1, \ldots, r+1$, рассмотрим производные функций (3) и (4):

$$\tilde{F}_n^{(m)}(t) = \frac{d^m \tilde{F}_n(t)}{dt^m} = \int_{-\infty}^{\infty} a_n^{-m} K^{(m)}\left(\frac{t-y}{a_n}\right) dF(y), \tag{6}$$

$$\tilde{F}_{n\tau_n}^{(m)}(t) = \frac{d^m \tilde{F}_{n\tau_n}(t)}{dt^m} = \int_{-\infty}^{\infty} I(\tau_n \geq y) a_n^{-m} K^{(m)}\left(\frac{t-y}{a_n}\right) dF(y). \tag{7}$$

Следующий результат из [1] утверждает закон повторного логарифма (ЗПЛ) для оценки F_n.

Теорема 1 ([1])**.** Пусть $b^{-1}(\tau) = 1 - H(\tau) > 0$ для $\tau < T_H = \inf\{t \in R: H(t) = 1\}$. Тогда с вероятностью 1 при $n \to \infty$

$$\sup_{t<\tau} |F_n(t) - F(t)| \leq C \cdot b^2(\tau) \cdot \left(\frac{\log\log n}{n}\right)^{1/2}. \tag{8}$$

Следующая теорема устанавливает свойство равномерной состоятельности оценки (2).

Теорема 2. Пусть выполнены условия (У1) – (У4), а последовательности $\{a_n, n \geq 1\}$ и $\{\tau_n, n \geq 1\}$ также удовлетворяют условиям

$$\lim_{n\to\infty}\left(a_n + \left(\frac{\log\log n}{n}\right)^{1/2} \frac{b^2(\tau_n)}{a_n^m}\right) = 0. \tag{9}$$

Тогда при $m = 1, \ldots, r+1$,

$$P\left(\lim_{n\to\infty} \sup_{t<\tau_n} |F_n^{(m)}(t) - F^{(m)}(t)| = 0\right) = 1. \tag{10}$$

Доказательство теоремы 2. Справедливо неравенство при $t \leq \tau_n$ и $m = 1, \ldots, r+1$:

$$\left|F_n^{(m)}(t) - F^{(m)}(t)\right| \leq \left|F_n^{(m)}(t) - \tilde{F}_n^{(m)}(t)\right| +$$

$$+|\tilde{F}_n^{(m)}(t)-F^{(m)}(t)|+\left|\int_{-\infty}^{\infty} I(y>\tau_n)a_n^{-m}K^{(m)}\left(\frac{t-y}{a_n}\right)dF(y)\right|. \qquad (11)$$

Поскольку $M\left[a_n^{-m}K^{(m)}\left(\frac{t-X_j}{a_n}\right)\right]=\tilde{F}_n^{(m)}(t)$, тогда согласно лемме 2.3 в [5] существует положительное постоянное C, независящее от a_n и для всех $m = 1, \ldots, r + 1$:

$$\sup_{t\in R}\left|\tilde{F}_n^{(m)}(t)-F^{(m)}(t)\right|\le C\cdot a_n. \qquad (12)$$

Кроме того, учитывая доказательство леммы 2.2 в [5] с учетом условия (У4) и интегрируя по частям получаем при $m = 1, \ldots, r + 1$:

$$\sup_{t\in R}\left|\int_{-\infty}^{\infty} I(y>\tau_n)a_n^{-m}K^{(m)}\left(\frac{t-y}{a_n}\right)dF(y)\right|\le 2v_m(1-F(\tau_n))=2\lambda_{m+1}(1-F(\tau_n)). \qquad (13)$$

Из (11) – (13) получаем оценку

$$\sup_{t\le\tau_n}|F_n^{(m)}(t)-F^{(m)}(t)|\le\sup_{t\le\tau_n}|F_n^{(m)}(t)-\tilde{F}_{n\tau_n}^{(m)}(t)|+\mathrm{O}(\max\{a_n,\ (1-F(\tau_n))\}),\ m=1,\ldots,r+1. \qquad (14)$$

Первое слагаемое в правой части (14) оценим сперва интегрированием по частям, а затем применением ЗПЛ (8) следующими соотношениями:

$$\sup_{t\le\tau_n}|F_n^{(m)}(t)-\tilde{F}_{n\tau_n}^{(m)}(t)|=a_n^{-m}\sup_{t\le\tau_n}\left|\int_{-\infty}^{\infty} I(\tau_n\ge y)\cdot K^{(m)}\left(\frac{t-y}{a_n}\right)d(F_n(y)-F(y))\right|=$$

$$=a_n^{-m}\cdot\sup_{t\le\tau_n}\left|\left[(F_n(y)-F(y))K^{(m)}\left(\frac{t-y}{a_n}\right)\right]_{-\infty}^{\tau_n}-\int_{-\infty}^{\infty}(F_n(y)-F(y))\,dK^{(m)}\left(\frac{t-y}{a_n}\right)\right|\le$$

$$\le 2v_m a_n^{-m}\cdot\sup_{t\le\tau_n}|F_n(t)-F(t)|=\mathrm{O}\left(\frac{b^2(\tau)}{a_n^m}\left(\frac{\log\log n}{n}\right)^{1/2}\right)=o(1),\quad n\to\infty. \qquad (15)$$

Теперь с учётом условия (8) утверждение теоремы следует из соотношений (14) и (15).

ЛИТЕРАТУРА

1. *Абдушукуров, А. А.* Статистика неполных наблюдений / А. А. Абдушукуров. Ташкент : Университет, 2009. 269 с.

2. *Абдушукуров, А. А.* Ядерное оценивание производных распределения в модели пропорциональных интенсивностей./ А. А. Абдушукуров, К. С. Сагидуллаев // Сб. «Статистические методы оценивания и проверки гипотез». Пермский госуниверситет. Пермь, 2013. Вып. 23. С. 4–10.

3. *Дмитриев, Ю. Г.* Непараметрическое оценивание функционалов по стационарным выборкам / Ю. Г. Дмитриев [и др.] // Томск : Изд-во ТГУ, 1974. С. 63–74.

4. *Alekseev, V. G.* Estimation of a probability density function and its derivatives / V. G. Alekseev // Математические заметки. 1972. Т. 12. № 5. С. 621–626.

5. *Schusler, E. F.* Estimation of a probability density function and its derivatives / E. F. Schusler // Ann. Math. Statist. 1969. V.40. № 4. P. 1187–1195.