Estimation of Distribution Function Based on Presmoothed Relative – Risk Function

Abdushukurov A. A.[1], **Bozorov S. B.**[2]
[1]Moscow State University named after M. V. Lomonosov, Tashkent Branch, Tashkent, Uzbekistan;
a_abdushukurov@rambler.ru
[2]Gulistan State University, Gulistan, Uzbekistan
suxrobbek_8912@mail.ru

Introduction. Censored data occur in survival analysis, bio-medical trials, industrial experiments. There are several schemas of censoring (from the right, left, both sides, mixed with competing risks and others). However, in statistical literature right random censoring is a very spreading, in so far as it is easy described from the methodological point of view. Here we also consider this kind of censorship in order to comparing our results with others. Let $X_1, X_2, ...$ and $Y_1, Y_2, ...$ be two independent sequences of independent and identically distributed (i.i.d.) random variables (r.v.-s) with common unknown continous distribution functions (d.f.-s) F and G, respectively. Let the X_j be censored on the right by Y_j, so that the observations available for us at the $n-$ th stage consist of the sample $C^{(n)} = \{(Z_j, \delta_j), 1 \leq j \leq n\}$, where $Z_j = \min(X_j, Y_j)$ and $\delta_j = I(X_j \leq Y_j)$ with $I(A)$ meaning the indicator of the event A. The main problem is consist of nonparametrical estimating of d.f. F with nuisance d.f. G based on censored sample $C^{(n)}$. In [1] we proposed following estimator of $F(t)$:

$$F_n^{PR}(t) = 1 - (1 - H_n(t))^{R_n^p(t)}. \tag{1}$$

where $R_n^p(t) = (\Lambda_n(t))^{-1} \cdot \int_{-\infty}^t p_n(u)\, d\Lambda_n(u), \Lambda_n(t) = \frac{1}{n}\sum_{j=1}^n \frac{I(z_j \leq t)}{1 - H_n(z_j) + \frac{1}{n}},\ H_n(t) = \frac{1}{n}\sum_{j=1}^n I(Z_j \leq t)$ and

$$p_n(t) = \left[\frac{1}{nh(n)}\sum_{j=1}^n k\left(\frac{t - Z_j}{h(n)}\right)\right]^{-1} \cdot \left[\frac{1}{nh(n)}\sum_{j=1}^n \delta_j k\left(\frac{t - Z_j}{h(n)}\right)\right],$$

where the kernel $k(\cdot)$ is a given probability density function and $\{h = h(n), n \geq 1\}$ is a bandwith sequence. In [1] the strong uniform consistency and asymptotic normality of estimator $F_n^{PR}(t)$ are established. In order to formulate these results we need some definitions and conditions. Let's denote

$$r(n) = h^2(n) + (nh(n))^{-1/2}(\log n)^{1/2}.$$

Consider following conditions.

(C1) $(F, G) \in K = \{(F, G) : N_F \bigcap N_G \neq \emptyset, P(X_j \leq Y_j) \in (0, 1)\}$, where $N_F = \{t : 0 < F(t) < 1\}$ and $N_G = \{t : 0 < G(t) < 1\}$;

(C2) Numbers α, β and γ are such that: $\min\{H(\alpha), 1 - H(\beta)\} \geq \gamma(0,1),\ \alpha > \tau_H = \sup\{t : H(t) = 0\}$ and $\beta < T_H = \inf\{t : H(t) = 1\}, [\alpha, \beta] \neq \emptyset$;

(C3) For all $n \geq 1 : P(0 < \nu_n < n) = 1$;

(C4) k is a symmetric, continuous, twice continuously differentiable and of bounded variation density function with compact support;

(C5)Density $q(t) = H'(t)$ exists, is four times continuously differentiable at $t \in [\alpha, \beta]$ and $\sup\limits_{\alpha \le t \le \beta} q(t) > 0$;

(C6)$p(t)$ is four times continuously differentiable at $t \in [\alpha, \beta]$;

(C7)$n^{1-\varepsilon} \cdot h(n) \to \infty$ for some $\varepsilon > 0$, $\sum_{n=1}^{\infty} h^{\lambda}(n) < \infty$ for some $\lambda > 0$ and $h^2(n) = o\left((nh(n))^{-1/2} \cdot \left(\log\left(\frac{1}{h(n)}\right)\right)^{1/2} \right)$;

Consider random functions

$$\varphi_1(t;z) = \frac{p(t)}{1 - H(t)} \left(I(Z \le t) - H(t) \right), \varphi_2(t;z) = \int_{-\infty}^{t} \frac{I(Z \le u) - H(u)}{1 - H(u)} p'(u)\, du,$$

$$\varphi_3(t; z, \delta) = \int_{-\infty}^{t} k\left(\frac{u - Z}{h}\right) \cdot \frac{(\delta - p(u))}{1 - H(u)} du.$$

In the next theorem, the difference ${F_n}^{PR}(t) - F(t)$ can be approximated by summ of i.i.d. random functions on t with the rate for the remainder term tending to zero at $n \to \infty$ almost surely.

Theorem 1. Under the conditions (C1) $-$ (C7),

$${F_n}^{PR}(t) - F(t) = (1 - F(t))\, \Psi_n(t) + \mathrm{Q}_n(t),$$

where $\Psi_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left[\varphi_1(t; Z_i) - \varphi_2(t; Z_i) + \varphi_3(t; Z_i, \delta_i) \right]$, with

$$\sup_{\alpha \le t \le \beta} |\mathrm{Q}_n(t)| \overset{a.s.}{=} \mathrm{O}\left(\max\{(r(n)\log n)^2, \frac{\log n}{n}\} \right).$$

Theorem 2. Under assumptions of theorem 1, at $n \to \infty$

$$\sup_{\alpha \le t \le \beta} \left| F_n^{PR}(t) - F(t) \right| \overset{a.s.}{=} \mathrm{O}\left(\max\{r(n)\log n, (\frac{\log n}{n})^{1/2}\} \right).$$

Theorem 3. Under the assumptions of theorem 1 and if

(C8) $nh^2(n)(\log n)^{-6} \to \infty$, $nh^8(n)(\log n)^4 \to 0$ and $h^3(n)(\log n)^5 \to 0$ as $n \to \infty$for any $t \in [\alpha, \beta]$:

1. If $nh^4(n) \to 0$, then $n^{1/2}\left(F_n^{PR}(t) - F(t)\right) \xrightarrow{d} N\left(0, \sigma^2(t)\right)$,

2. If $nh^4(n) \to C^4$, then $n^{1/2}\left(F_n^{PR}(t) - F(t)\right) \xrightarrow{d} N\left(b(t), \sigma^2(t)\right)$,

where $b(t) = C^2 (1 - F(t))\, \alpha(t)\, d(k)$, $d(k) = \int u^2 k^2(u)\, du$, $\sigma^2(t) = (1 - F(t))^2 \gamma(t)$, $\alpha(t) = \int_{-\infty}^{t} \frac{\left(\frac{1}{2} p''(u) q(u) + p'(u) q'(u)\right) du}{1 - H(u)}$, $\gamma(t) = \int_{-\infty}^{t} \mu(u)\, du$, $\mu(t) = \frac{p(t) q(t)}{(1 - H(t))^2}$. We will discuss other properties and aplication of estimator (1).

REFERENCES

1. Abdushukurov A.A. , Bozorov S.B. , Nurmukhamedova N.S. Noparametric Estimation of Distribution Function Under Right Random Censoring Based on Presmoothed Relative - Risk Function // ISSN 1995-0802, Lobachevskii journal of mathematics, 2021, Vol. 42, No. 2, 257-268. Pleiades Publishing, Ltd., 2021.