

# Estimation of Distribution Function Based on Presmoothed Relative-Risk Function

Abdurakhim Akhmedovich Abdushukurov<sup>1</sup>, Sukhrob Bakhodirovich Bozorov<sup>2</sup>,  
Dilshod Ravilovich Mansurov<sup>3</sup>

<sup>1</sup>Moscow State University Tashkent Branch, Tashkent, Uzbekistan

<sup>2</sup>Guliston State University, Gulistan, Uzbekistan

<sup>3</sup>Navoi State Pedagogical Institute, Navoi, Uzbekistan

Email: a\_abdushukurov@rambler.ru, suxrobbek\_8912@mail.ru, mathematicianmd@gmail.com

**How to cite this paper:** Abdushukurov, A.A., Bozorov, S.B. and Mansurov, D.R. (2022) Estimation of Distribution Function Based on Presmoothed Relative-Risk Function. *Applied Mathematics*, 13, 191-204. <https://doi.org/10.4236/am.2022.132015>

**Received:** December 29, 2021

**Accepted:** February 22, 2022

**Published:** February 25, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In this article, the lifetime data subjecting to right random censoring is considered. Nonparametric estimation of the distribution function based on the conception of presmoothed estimation of relative-risk function and the properties of the estimator by using methods of numerical modeling are discussed. In the model under consideration, the estimates were compared using numerical methods to determine which of the estimates is actually better.

## Keywords

Random Censorship, Product-Limit, Relative Risk, Presmoothed, Proportional Hazards, Asymptotic Representation, Strong Consistency, Asymptotic Normality

## 1. 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 wide spread, in so far as it was easily described from the methodological point of view. Here we consider also this kind of censorship in order to compare our results with others.

Let  $X_1, X_2, \dots$  and  $Y_1, Y_2, \dots$  be two independent sequences of independent and identically distributed (i.i.d.) random variables (r.v.-s) with common unknown continuous 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 consists of a non-parametrical estimating of d.f.  $F$  with nuisance d.f.  $G$  based on the censored sample  $C^{(n)}$ , where the number of observed  $X_j - s$ ,  $v_n = \delta_1 + \dots + \delta_n$  is random.

Kaplan and Meier [1] were the first to suggest the product-limit (PL) estimator  $F_n^{PL}$  of  $F$  defined as

$$F_n^{PL}(t) = \begin{cases} 1 - \prod_{\{j: Z_{(j)} \leq t\}} \left(1 - \frac{\delta_{(j)}}{n - j + 1}\right), & t \leq Z_{(n)}, \\ 1, & t > Z_{(n)}, \delta_{(n)} = 1, \\ \text{undefined,} & t > Z_{(n)}, \delta_{(n)} = 0, \end{cases} \tag{1}$$

where  $Z_{(1)} \leq \dots \leq Z_{(n)}$  are the order statistics of  $Z$ -sample  $\{Z_j, 1 \leq j \leq n\}$  and  $\{\delta_{(j)}, 1 \leq j \leq n\}$  the sequence of indicators adjunct to the ordered  $Z$ -sample. There are different versions of PL-estimators. However, those do not coincide, if the largest  $Z_j$  is a censoring time. There is an enormous set of works on investigating several properties of PL-estimators and their application on statistical problems, specially in the case of right random censorship. However,  $F_n^{PL}$  is not a unique estimator of d.f.  $F$ . Abdushukurov [2] [3] proposed another estimator of  $F$ , of relative-risk power type:

$$F_n^{RR}(t) = 1 - (1 - H_n(t))^{R_n(t)} = \begin{cases} 0, & t < Z_{(1)}, \\ 1 - \left(\frac{n-j}{n}\right)^{R_n(t)}, & Z_{(j)} \leq t < Z_{(j+1)}, 1 \leq j \leq n-1, \\ 1, & t \geq Z_{(n)}, \end{cases} \tag{2}$$

where  $H_n(t) = \frac{1}{n} \sum_{j=1}^n I(Z_j \leq t), t \in \mathbb{R}^1 \equiv (-\infty; +\infty)$  is an empirical estimator of d.f.  $P(Z_j \leq t) = 1 - (1 - F(t))(1 - G(t)) \equiv H(t)$  and  $R_n(t) = (\Lambda_n(t))^{-1} \Lambda_{1n}(t)$  is an estimator of relative-risk function  $R(t) = (\Lambda(t))^{-1} \Lambda_1(t), t \in \mathbb{R}^1$ . Here cumulative hazard functions (c.h.f.-s)  $\Lambda, \Lambda_0$  and  $\Lambda_1$  corresponding to d.f.-s  $H, G$  and  $F$  defined as

$$\begin{aligned} \Lambda(t) &= \int_{-\infty}^t \frac{dH(u)}{1 - H(u-)} = \Lambda_0(t) + \Lambda_1(t), \\ \Lambda_0(t) &= \int_{-\infty}^t \frac{dG(u)}{1 - G(u-)} = \int_{-\infty}^t \frac{dH_0(u)}{1 - H(u-)}, \\ \Lambda_1(t) &= \int_{-\infty}^t \frac{dF(u)}{1 - F(u-)} = \int_{-\infty}^t \frac{dH_1(u)}{1 - H(u-)}, \end{aligned} \tag{3}$$

with subdistribution functions  $H_0(t) + H_1(t) = H(t), t \in \mathbb{R}^1$ ,

$$\begin{aligned} H_0(t) &= P(Z_j \leq t, \delta_j = 0) = \int_{-\infty}^t (1 - F(u-)) dG(u), \\ H_1(t) &= P(Z_j \leq t, \delta_j = 1) = \int_{-\infty}^t (1 - G(u-)) dF(u). \end{aligned}$$

The corresponding estimators of c.h.f.-s (3) are

$$\Lambda_n(t) = \int_{-\infty}^t \frac{dH_n(u)}{1-H_n(u-)} = \Lambda_{0n}(t) + \Lambda_{1n}(t),$$

where

$$\Lambda_{kn}(t) = \int_{-\infty}^t \frac{dH_{kn}(u)}{1-H_n(u-)}, \quad k = 0, 1; \quad H_{kn}(t) = \frac{1}{n} \sum_{j=1}^n I(Z_j \leq t, \delta_j = k)$$

are empirical counterparts of  $H_k(t), k = 0, 1$  with  $H_{0n}(t) + H_{1n}(t) = H_n(t), t \in \mathbb{R}^1$ .

In [2] [3] [4] it was shown that both estimators (1) and (2) have similar asymptotic properties tending to the same limiting Gaussian process. However, the relative-risk power estimator (2) has some small-sample advantages with respect to PL-estimator (1). For example, it is not sensitive to censoring in last observed point  $Z_{(n)}$ , since  $F_n^{RR}(Z_{(n)}) = 1$  and it is identifiable with the model:  $(1 - F_n^{RR}(t))(1 - G_n^{RR}(t)) = 1 - H_n(t), n \geq 1, t \in \mathbb{R}^1$ , where  $G_n^{RR}(t) = 1 - (1 - H_n(t))^{1-R_n(t)}$  is a corresponding estimator of d.f.  $G(t)$ . In [4] it was proposed several extended versions of estimator (2) in generalized models of incomplete observations mixed with competing risks. These estimators were also extensively studied in some statistical problems. It is not difficult to observe that estimator (2) is a natural extension of well-know ACL-(Abdushukurov-Cheng-Lin) estimator of F in simple Proportional Hazards Model (PHM):

$$F_n^{ACL}(t) = 1 - (1 - H_n(t))^{p_n}, \quad t \in \mathbb{R}^1,$$

where  $p_n = \frac{V_n}{n}$  is an estimator of probability  $p = P(\delta_j = 1)$ , which is value of the constant relative-risk function  $R(t) \equiv p$  (so far as in PHM,

$\Lambda_1(t) = p\Lambda(t), t \in \mathbb{R}^1$ ). Note that  $F_n^{ACL}$  was independently proposed and studied by Abdushukurov [5] and Chen, Lin [6] (for more information, see also Csörgő [7]). This estimator was studied, extended and used by many other authors up a present. The main property of PHM is its characterization by independence of subsamples  $\{Z_1, \dots, Z_n\}$  and  $\{\delta_1, \dots, \delta_n\}$ . This property is equivalent to relation  $1 - G(t) = (1 - F(t))^\beta, t \in \mathbb{R}^1$  for some positive  $\beta$ . In PHM,

$p = \frac{1}{1 + \beta}$  and therefore  $\beta$  is a censoring parameter. Estimator  $F_n^{ACL}$  in

PHM is asymptotically efficient with respect to  $F_n^{PL}$ . This advantage of the estimator is well preserved for plug-in estimators of many functionals (see [2] [5] [7]). That is why, in this framework, the conditional probability that datum is not censored given its observed value

$$p(t) = P(\delta_j = 1 / Z_j = t) = E[\delta_j / Z_j = t], \quad t \in \mathbb{R}^1, \tag{4}$$

is a very important function, which in PHM is constant  $p(t) = \frac{1}{1 + \beta}, t \in \mathbb{R}^1$ .

Moreover, the key role of probability (4) takes part in expressing c.h.f.  $\Lambda_1$  via

$\Lambda$  as

$$\Lambda_1(t) = \int_{-\infty}^t p(u) d\Lambda(u), t \in \mathbb{R}^1,$$

and, therefore, a relative-risk function given as

$$R(t) = (\Lambda(t))^{-1} \int_{-\infty}^t p(u) d\Lambda(u), t \in \mathbb{R}^1.$$

Probability (4) is a regression of  $\delta_j$  on  $Z_j$ . Hence, it can be estimated by some regression statistics. We have used the following nonparametric regression estimator of Nadaraya [8] and Watson [9]:

$$p_n(t) = \left[ \frac{1}{nh(n)} \sum_{j=1}^n k\left(\frac{t-Z_j}{h(n)}\right) \right]^{-1} \left[ \frac{1}{nh(n)} \sum_{j=1}^n \delta_j k\left(\frac{t-Z_j}{h(n)}\right) \right], \tag{5}$$

where the kernel  $k(\cdot)$  is a given probability density function and  $\{h = h(n), n \geq 1\}$  is a bandwidth sequence such that:  $h \downarrow 0, n \rightarrow \infty$ . In case of dependence of probability (4) on unknown parameters, it may be estimated parametrically (see, Dikta [10] in this context). Cao et al. [11] proposed following presmoothed PL-estimator of d.f. F by replacing the censoring indicators  $\delta_{(j)}$  in the expression of PL-estimator (1) by the estimator (5) at the observed data points:

$$F_n^p(t) = 1 - \prod_{\{j: Z_{(j)} \leq t\}} \left( 1 - \frac{p_n(Z_{(j)})}{n-j+1} \right), t \in \mathbb{R}^1. \tag{6}$$

Some asymptotic properties of estimator (6) were investigated in [11] [12]. Taking into account some advantages of estimator (2) with respect to (1), we propose a new presmoothed relative-risk power (PRRP) estimator:

$$F_n^{PRR}(t) = 1 - (1 - H_n(t))^{R_n^p(t)} = \begin{cases} 0, & t < Z_{(1)}, \\ 1 - \left(\frac{n-j}{n}\right)^{R_n^p(t)}, & Z_{(j)} \leq t < Z_{(j+1)}, 1 \leq j \leq n-1, \\ 1, & t \geq Z_{(n)}, \end{cases} \tag{7}$$

were

$$R_n^p(t) = (\Lambda_n(t))^{-1} \Lambda_n^p(t) = (\Lambda_n(t))^{-1} \int_{-\infty}^t p_n(u) d\Lambda_n(u), t \in \mathbb{R}^1,$$

is a partially presmoothed analogue of estimator  $R_n(t)$ . For probability mass function (4) smooth estimator (5) is used in formula for c.h.f.  $\Lambda_1(t)$ . But the estimator (7) is not smooth. We can see that estimator (7) is also well defined in whole line without any conditions on censorship.

## 2. Asymptotic Properties of PRRP Estimator

Let's denote  $r(n) = h^2(n) + (nh(n))^{-1/2} (\log n)^{1/2}$ . In order to investigate the properties of estimator (7) we need the following conditions:

$$(C1) \quad (F, G) \in K = \{(F, G) : N_F \cap N_G \neq \emptyset, P(X_j \leq Y_j) \in (0, 1)\}, \text{ where}$$

$$N_F = \{t : 0 < F(t) < 1\} \text{ and } N_G = \{t : 0 < G(t) < 1\};$$

(C2) Numbers  $\alpha, \beta$  and  $\gamma$  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  $n \geq 1$  there takes place  $P(0 < v_n < n) = 1$ ;

(C4)  $k$  is a symmetric, twice continuously differentiable and 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  $\sup q(t) > 0$ ;

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

(C7)  $n^{1-\varepsilon} \cdot h(n) \rightarrow \infty$  for some  $\varepsilon > 0$ ,  $\sum_{n=1}^{\infty} h^\lambda(n) < \infty$  for some  $\lambda > 0$

$$\text{and } h^2(n) = o\left((nh(n))^{-1/2} \cdot \left(\log\left(\frac{1}{h(n)}\right)\right)^{1/2}\right).$$

Consider random functions

$$\varphi_1(t; z) = \frac{p(t)}{1 - H(t)}(I(Z \leq t) - H(t)),$$

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

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

In the next theorem, we will show that PRRP estimator can be approximated by summ of i.i.d. random functions on  $t$  with the rate for the remainder term tending to zero at  $n \rightarrow \infty$  almost surely.

**Theorem 1.** If the conditions (C1)-(C7) are fulfilled, then there holds

$$F_n^{PR}(t) - F(t) = (1 - F(t))\Omega_n(t) + Q_n(t) \tag{8}$$

with  $\sup_{\alpha \leq t \leq \beta} |Q_n(t)| \stackrel{a.s.}{=} O\left(\max\left\{(r(n) \log n)^2, \frac{\log n}{n}\right\}\right)$ , where

$$\Omega_n(t) = \frac{1}{n} \sum_{i=1}^n [\varphi_1(t; Z_i) - \varphi_2(t; Z_i) + \varphi_3(t; Z_i, \delta_i)].$$

The following Lemmas allow us to prove Theorem 1.

**Lemma 1.** (Corollary from lemma 3.2 in [12]) Assume that the conditions (C1)-(C7) are fulfilled. Then following estimate holds

$$\sup_{\alpha \leq t \leq \beta} |\Lambda_{1n}^p(t) - \Lambda_1(t)| \stackrel{a.s.}{=} O(r(n) \log n). \tag{9}$$

**Lemma 2.** (Theorem 3.4 in [12]). Assume that the conditions (C1)-(C7) are fulfilled. Then for  $t \in [\alpha, \beta]$  it is true that

$$\Lambda_{1n}^p(t) - \Lambda_1(t) \stackrel{a.s.}{=} \Omega_n(t) + O(r^2(n)). \tag{10}$$

**Lemma 3.** (Dworetzky-Kiefer-Wolfowitz inequality with tight constant  $d = 2$  from [13]). For all  $n \geq 1$ , some  $\gamma > 0$  and  $\varepsilon = ((1 + \gamma)(2n)^{-1} \log n)^{1/2}$  the following estimate holds

$$P\left(\sup_{-\infty < t < \infty} |H_n(t) - H(t)| > \varepsilon\right) \leq 2n^{-(1+\gamma)} \tag{11}$$

**Lemma 4.** (Lemma on page 53 [14]). For  $\gamma > 0$  there is true following estimate

$$P\left(\sup_{\alpha \leq t \leq \beta} \left| \int_{-\infty}^t \frac{(H_n(u) - H(u))d(H_n(u) - H(u))}{(1 - H(u))^2} \right| > A(1 - H(\beta))^{-2} n^{-1} \log n \right) \leq Bn^{-(1+\gamma)} \tag{12}$$

where  $A = A(\gamma)$  and  $B$  are some positive constants.

**Lemma 5.** 1) If conditions (C1)-(C3) are fulfilled, then there hold estimates

$$\begin{aligned} \text{a) } & \sup_{\alpha \leq t \leq \beta} \left| \frac{\Lambda(t)}{\Lambda_n(t)} - 1 \right| \stackrel{a.s.}{=} O\left(\left(\frac{\log n}{n}\right)^{1/2}\right); \\ \text{b) } & \sup_{\alpha \leq t \leq \beta} \left| \frac{\Lambda_1(t)}{\Lambda_n(t)} - R(t) \right| \stackrel{a.s.}{=} O\left(\left(\frac{\log n}{n}\right)^{1/2}\right); \end{aligned}$$

2) If the conditions (C4)-(C7) are additionally required fulfilled, then the following estimate is also valid

$$\text{c) } \sup_{\alpha \leq t \leq \beta} |R_n^p(t) - R(t)| \stackrel{a.s.}{=} O\left(\max\left\{r(n) \log n, \left(\frac{\log n}{n}\right)^{1/2}\right\}\right).$$

*Proof of Lemma 5.* Observe that

$$\begin{aligned} \sup_{\alpha \leq t \leq \beta} \left| \frac{\Lambda(t)}{\Lambda_n(t)} - 1 \right| & \leq \sup_{\alpha \leq t \leq \beta} (H_n(\alpha))^{-1} |\Lambda_n(t) - \Lambda(t)| \\ & \leq 3 \left[ H_n(\alpha)(1 - H_n(\beta)(1 - H(\beta))) \right]^{-1} \sup_{-\infty < t < \infty} |H_n(t) - H(t)| \\ & \stackrel{a.s.}{=} O\left(\left(\frac{\log n}{n}\right)^{1/2}\right), \end{aligned} \tag{13}$$

where the last equality follows from (11) and Borel-Cantelli's lemma. Further, it is clear that b) is consequence of (13) and

$$\begin{aligned} \sup_{\alpha \leq t \leq \beta} \left| \frac{\Lambda_1(t)}{\Lambda_n(t)} - R(t) \right| & = \sup_{\alpha \leq t \leq \beta} R(t) \left| \frac{\Lambda(t)}{\Lambda_n(t)} - 1 \right| \leq \sup_{\alpha \leq t \leq \beta} \left| \frac{\Lambda(t)}{\Lambda_n(t)} - 1 \right| \\ & \stackrel{a.s.}{=} O\left(\left(\frac{\log n}{n}\right)^{1/2}\right). \end{aligned}$$

For c) we have

$$\begin{aligned} \sup_{\alpha \leq t \leq \beta} |R_n^p(t) - R(t)| & = \sup_{\alpha \leq t \leq \beta} \left| R_n^p(t) - \frac{\Lambda_1(t)}{\Lambda_n(t)} \right| + \sup_{\alpha \leq t \leq \beta} \left| \frac{\Lambda_1(t)}{\Lambda_n(t)} - R(t) \right| \\ & \leq \sup_{\alpha \leq t \leq \beta} (\Lambda_n(t))^{-1} \left[ \sup_{\alpha \leq t \leq \beta} |\Lambda_n^p(t) - \Lambda_1(t)| + \sup_{\alpha \leq t \leq \beta} |\Lambda_n(t) - \Lambda(t)| \right] \\ & \stackrel{a.s.}{=} (H_n(\alpha))^{-1} \left[ O(r(n)) + O\left(\left(\frac{\log n}{n}\right)^{1/2}\right) \right] \\ & \stackrel{a.s.}{=} O\left(\max\left\{r(n) \log n, \left(\frac{\log n}{n}\right)^{1/2}\right\}\right), \end{aligned}$$

where we have used Lemma 1 and the estimate (13). Lemma 5 is proved.

*Proof of Theorem 1.* By two-term Taylor expansion for difference  $F_n^{PR}(t) - F(t)$  we obtain

$$F_n^{PR}(t) - F(t) = (1 - F(t))M_n(t) + L_n(t), \tag{14}$$

where

$$M_n(t) = -\left[ R_n^p(t) \log(1 - H_n(t)) R(t) \log(1 - H(t)) \right],$$

$$L_n(t) = \frac{1}{2} M_n^2(t) \exp(-\chi_n(t)), \quad \chi_n(t) \text{ lies between } \log(1 - F_n^{PR}(t)) \text{ and } \Lambda_1(t) = -\log(1 - F(t)).$$

For  $M_n(t)$  we have representation

$$M_n(t) = \left( \Lambda_{ln}^p(t) - \Lambda_1(t) \right) \frac{\Lambda(t)}{\Lambda_n(t)} - \left( \Lambda_n(t) - \Lambda(t) \right) \frac{\Lambda_1(t)}{\Lambda_n(t)} + \left[ -\log(1 - H_n(t)) + \log(1 - H(t)) \right] R_n^p(t). \tag{15}$$

Now we will show that in (15) the first summand is main term and the sum of other two terms tends (at  $n \rightarrow \infty$ ) to zero. Consider first term, which can be decomposed as

$$\left( \Lambda_{ln}^p(t) - \Lambda_1(t) \right) \frac{\Lambda(t)}{\Lambda_n(t)} = \left( \Lambda_{ln}^p(t) - \Lambda_1(t) \right) + S_n(t), \tag{16}$$

where

$$S_n(t) = \left( \Lambda_{ln}^p(t) - \Lambda_1(t) \right) \left( \frac{\Lambda(t)}{\Lambda_n(t)} - 1 \right).$$

Hence by (9) and Lemma 5 (condition (a)), we have

$$\sup_{\alpha \leq t \leq \beta} |S_n(t)| = O\left( r(n) (\log n)^{3/2} n^{-1/2} \right). \tag{17}$$

Then from (10), (16) and (17) we obtain for all  $t \in [\alpha, \beta]$

$$\left( \Lambda_{ln}^p(t) - \Lambda_1(t) \right) \frac{\Lambda(t)}{\Lambda_n(t)} \stackrel{a.s.}{=} \Omega_n(t) + O(r_1(n)).$$

For other two terms of (15) we have

$$\begin{aligned} & -\left( \Lambda_n(t) - \Lambda(t) \right) \frac{\Lambda_1(t)}{\Lambda_n(t)} + \left[ -\log(1 - H_n(t)) + \log(1 - H(t)) \right] \\ &= R_n^p(t) \left\{ -\left( \Lambda_n(t) - \Lambda_1(t) \right) + \left[ -\log(1 - H_n(t)) + \log(1 - H(t)) \right] \right. \\ & \quad \left. - \left( \Lambda_n(t) - \Lambda(t) \right) \left( \frac{\Lambda_1(t)}{\Lambda_{ln}^p(t)} - 1 \right) \right\} \\ &= R_n^p(t) [A_n(t) + B_n(t) + C_n(t)]. \end{aligned} \tag{18}$$

Hence by (9) and (13), we obtain

$$\begin{aligned} & \sup_{\alpha \leq t \leq \beta} R_n^p(t) |C_n(t)| \leq \sup_{\alpha \leq t \leq \beta} |C_n(t)| \\ & \leq \sup_{\alpha \leq t \leq \beta} (\Lambda_{1n}^p(t))^{-1} \sup_{\alpha \leq t \leq \beta} |\Lambda_n(t) - \Lambda(t)| \sup_{\alpha \leq t \leq \beta} |\Lambda_{1n}^p(t) - \Lambda_1(t)| \quad (19) \\ & \stackrel{a.s.}{=} O\left(r(n)(\log n)^{3/2} n^{-1/2}\right). \end{aligned}$$

Now by simple algebra and integrating by parts for  $A_n(t)$  and taking into account Taylor expansion for  $B_n(t)$  we get chain of equalities

$$\begin{aligned} & A_n(t) + B_n(t) = -(\Lambda_n(t) - \Lambda(t)) + [-\log(1 - H_n(t)) + \log(1 - H(t))] \\ & = -\left[ \int_{-\infty}^t \frac{dH_n(u)}{1 - H_n(u-)} - \int_{-\infty}^t \frac{dH(u)}{1 - H(u)} \right] + \left[ \frac{H_n(t) - H(t)}{1 - H(t)} + \frac{1}{2} \frac{(H_n(t) - H(t))^2}{\theta_n^2(t)} \right] \\ & = \left\{ -\left[ \int_{-\infty}^t \frac{dH_n(u)}{1 - H_n(u-)} - \int_{-\infty}^t \frac{dH_n(u)}{1 - H_n(u)} \right] - \left[ \int_{-\infty}^t \frac{dH_n(u)}{1 - H_n(u)} - \int_{-\infty}^t \frac{dH(u)}{1 - H(u)} \right] \right. \\ & \quad \left. + \left[ \frac{H_n(t) - H(t)}{1 - H(t)} + \frac{1}{2} \frac{(H_n(t) - H(t))^2}{\theta_n^2(t)} \right] \right\} \\ & = \left\{ -\left[ \int_{-\infty}^t \frac{(H_n(u) - H(u))dH(u)}{(1 - H_n(u))(1 - H(u))} + \frac{H_n(t) - H(t)}{1 - H(t)} \right. \right. \\ & \quad - \int_{-\infty}^t \frac{(H_n(u) - H(u))dH(u)}{(1 - H(u))^2} + \int_{-\infty}^t \frac{(H_n(u) - H(u))d(H_n(u) - H(u))}{(1 - H_n(u))(1 - H(u))} \\ & \quad \left. \left. - D_{1n}(t) \right] + \left[ \frac{H_n(t) - H(t)}{1 - H(t)} + D_{2n}(t) \right] \right\} \quad (20) \\ & = -\int_{-\infty}^t \frac{(H_n(u) - H(u))dH(u)}{(1 - H_n(u))(1 - H(u))} + \int_{-\infty}^t \frac{(H_n(u) - H(u))dH(u)}{(1 - H(u))^2} \\ & \quad + \int_{-\infty}^t \frac{(H_n(u) - H(u))d(H_n(u) - H(u))}{(1 - H_n(u))(1 - H(u))} + D_{1n}(t) + D_{2n}(t), \end{aligned}$$

where

$$\begin{aligned} D_{1n}(t) &= \int_{-\infty}^t \frac{(H_n(u) - H_n(u-))dH_n(u)}{(1 - H_n(u-))(1 - H_n(u))}, \quad D_{2n}(t) = \frac{1}{2} \frac{(H_n(t) - H(t))^2}{\theta_n^2(t)}, \\ \theta_n(t) &\in [\min\{H_n(t), H(t)\}, \max\{H_n(t), H(t)\}]. \end{aligned}$$

Hence, using (11) we have

$$\sup_{\alpha \leq t \leq \beta} |D_{1n}(t)| \stackrel{a.s.}{=} O\left(\frac{1}{n}\right), \quad \sup_{\alpha \leq t \leq \beta} |D_{2n}(t)| \stackrel{a.s.}{=} O\left(\frac{\log n}{n}\right). \quad (21)$$

Consider equality

$$\frac{1}{1 - H_n(u)} = \frac{1}{1 - H(u)} + \frac{H_n(u) - H(u)}{(1 - H_n(u))(1 - H(u))} \quad (22)$$

and its integral form

$$\begin{aligned} & \int_{-\infty}^t \frac{(H_n(u) - H(u))dH(u)}{(1 - H_n(u))(1 - H(u))} \\ &= \int_{-\infty}^t \frac{(H_n(u) - H(u))dH(u)}{(1 - H(u))^2} + \int_{-\infty}^t \frac{(H_n(u) - H(u))^2 dH(u)}{(1 - H(u))^2 (1 - H_n(u))}. \end{aligned} \tag{23}$$

Using (22) and (23) in the third and first integrals in (20) and taking into account also (21) we obtain

$$\begin{aligned} & A_n(t) + B_n(t) \\ & \stackrel{a.s.}{=} - \int_{-\infty}^t \frac{(H_n(u) - H(u))^2 dH(u)}{(1 - H(u))^2 (1 - H_n(u))} + \int_{-\infty}^t \frac{(H_n(u) - H(u))^2 d(H_n(u) - H(u))}{(1 - H_n(u))(1 - H(u))} \\ & \quad + \int_{-\infty}^t \frac{(H_n(u) - H(u))d(H_n(u) - H(u))}{(1 - H(u))^2} + O\left(\frac{\log n}{n}\right). \end{aligned} \tag{24}$$

Application of estimator (11) to the first and second integrals and (12) to the third integral in (24) gives that

$$\begin{aligned} & \sup_{\alpha \leq t \leq \beta} \left| \int_{-\infty}^t \frac{(H_n(u) - H(u))^2 dH(u)}{(1 - H(u))^2 (1 - H_n(u))} \right| \\ & \leq \left[ (1 - H(\beta))^2 (1 - H_n(\beta)) \right]^{-1} \left[ \sup_{-\infty < t < \infty} |H_n(u) - H(u)| \right]^2 \stackrel{a.s.}{=} O\left(\frac{\log n}{n}\right), \end{aligned} \tag{25}$$

$$\begin{aligned} & \sup_{\alpha \leq t \leq \beta} \left| \int_{-\infty}^t \frac{(H_n(u) - H(u))^2 d(H_n(u) - H(u))}{(1 - H_n(u))(1 - H(u))} \right| \\ & \leq 2 \left[ (1 - H(\beta))(1 - H_n(\beta)) \right]^{-1} \left[ \sup_{-\infty < t < \infty} |H_n(u) - H(u)| \right]^2 \stackrel{a.s.}{=} O\left(\frac{\log n}{n}\right), \end{aligned} \tag{26}$$

$$\sup_{\alpha \leq t \leq \beta} \left| \int_{-\infty}^t \frac{(H_n(u) - H(u))d(H_n(u) - H(u))}{(1 - H(u))^2} \right| \stackrel{a.s.}{=} O\left(\frac{\log n}{n}\right). \tag{27}$$

Thus, adding (18), (19) and (24)-(27), we derive

$$\begin{aligned} & \sup_{\alpha \leq t \leq \beta} R_n^p(t) |A_n(t) + B_n(t) + C_n(t)| \\ & \stackrel{a.s.}{=} O\left( \max \left\{ r(n) \cdot (\log n)^{3/2} n^{-1/2} \cdot \frac{\log n}{n} \right\} \right) \end{aligned}$$

Then, by virtue of (9) and (16), from (15) we have

$$\sup_{\alpha \leq t \leq \beta} |M_n(t)| \stackrel{a.s.}{=} O\left( \max \left\{ r(n) \log n, \frac{\log n}{n} \right\} \right)$$

and, consequently,

$$\sup_{\alpha \leq t \leq \beta} |L_n(t)| \leq \left[ \sup_{\alpha \leq t \leq \beta} |M_n(t)| \right]^2 \stackrel{a.s.}{=} O\left( \left[ \max \left\{ r(n) \log n, \frac{\log n}{n} \right\} \right]^2 \right). \tag{28}$$

Finally, the desired result (8) follows, from (15)-(17) and (28). The proof is completed.

Now as a consequence the strong uniform consistency of PRRP estimator can be obtained.

**Theorem 2.** *Let the assumptions of Theorem 1 are fulfilled. Then at  $n \rightarrow \infty$  there holds*

$$\sup_{\alpha \leq t \leq \beta} |F_n^{PRR}(t) - F(t)| = O \left( \max \left\{ r(n) \log n, \left( \frac{\log n}{n} \right)^{1/2} \right\} \right). \tag{29}$$

*Proof of Theorem 2.* Using inequality  $|u - v| \leq |\log u - \log v|$ ,  $0 < u, v \leq 1$ , we have a chain of following relations:

$$\begin{aligned} \sup_{\alpha \leq t \leq \beta} |F_n^{PRR}(t) - F(t)| &\leq \sup_{\alpha \leq t \leq \beta} |-(1 - F_n^{PRR}(t)) + \log(1 - F(t))| \\ &= \sup_{\alpha \leq t \leq \beta} |-R_n^P(t) \log(1 - H_n(t)) + R(t) \log(1 - H(t))| \\ &\leq (1 - H(\beta))^{-1} \left[ \sup_{\alpha \leq t \leq \beta} |R_n^P(t) - R(t)| + (1 - H(\beta))^{-1} \sup_{\alpha \leq t \leq \beta} |H_n(t) - H(t)| \right] \\ &\stackrel{a.s.}{=} O \left( \max \left\{ r(n) \log n, \left( \frac{\log n}{n} \right)^{1/2} \right\} \right), \end{aligned}$$

where the last equality is obtained by using of Lemmas 3 and 5 (condition (c)) and this completes the proof of Theorem 2.

The approximating sequence of normalized sum of random functions  $\Omega_n(t)$  in Theorem 1 is the same that for presmoothed PL-estimator (6). Therefore, from theorem 3.7 in [12] follows the asymptotic normality of PRRP estimator, under taking into account the representation (8).

**Theorem 3.** *Let the assumptions of Theorem 1 be fulfilled and*

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

*Then there hold*

- 1) If  $nh^4(n) \rightarrow 0$ , then  $n^{1/2} (F_n^{PRR}(t) - F(t)) \xrightarrow{d} N(0, \sigma^2(t))$ ,
- 2) If  $nh^4(n) \rightarrow C^4$ , then  $n^{1/2} (F_n^{PRR}(t) - F(t)) \xrightarrow{d} N(b(t), \sigma^2(t))$ ,

*where*

$$\begin{aligned} b(t) &= C^2 (1 - F(t)) \alpha(t) d(k), \\ d(k) &= \int_{-\infty}^{\infty} u^2 k^2(u) du, \\ \alpha(t) &= \int_{-\infty}^t \frac{\left( \frac{1}{2} p''(u) q(u) + p'(u) q'(u) \right) du}{1 - H(u)}, \\ \sigma^2(t) &= (1 - F(t))^2 \gamma(t), \\ \gamma(t) &= \int_{-\infty}^t \mu(u) du, \quad \mu(t) = \frac{p(t) q(t)}{(1 - H(t))^2}. \end{aligned}$$

### 3. Numerical Study of Estimators

In this section, we investigate the above estimates using numerical methods. By python programming language we are preparing a high-quality sample. We select  $F(t, c) = 1 - e^{-t^c}, c = 1.79 (t \geq 0)$  and get a sample of volume  $n = 500$ . This sample is censored from the right with r.v.-s having a d.f.  $G(t) = 1 - e^{-t} (t \geq 0)$ . The resulting sample has a degree of censorship 47%. We will study the above estimates on the resulting sample.

The red line in the figure shows the theoretical d.f.  $F(t, c)$  and the green line shows the Kaplan-Meier estimate (Figure 1). One disadvantage of this estimate is that it may not matter at this endpoint.

Now we draw the evaluation graph (Figure 2) of estimator proposed by Abdushukurov (2). In the figure, the red line shows the theoretical d.f., the blue line shows Abdushukurov's estimate. It can be seen from the graphs drawn that both estimates are very good. But in practice, it is difficult for us to see on the graph which score is better. Therefore, we study the sum

$$\sum_{i=1}^{n-1} (F_n(Z_i) - F(Z_i))^2.$$

Let's make the appropriate tables for it.

From the table (Table 1) above, it can be concluded that the estimate (2) proposed by Abdushukurov is closer to the d.f.  $F(t, c)$ .

Now we draw the estimates (6) (Figure 3) and (7) (Figure 4).

As can be seen from the graph, despite the high level of censorship, both estimates are very close to the theoretical d.f. The table below shows that the price actually depends on the selected bandwidth sequence.

From the table (Table 2) above, we can conclude that the  $F_n^{PR}$ -estimator is better than  $F_n^P$ -estimator.

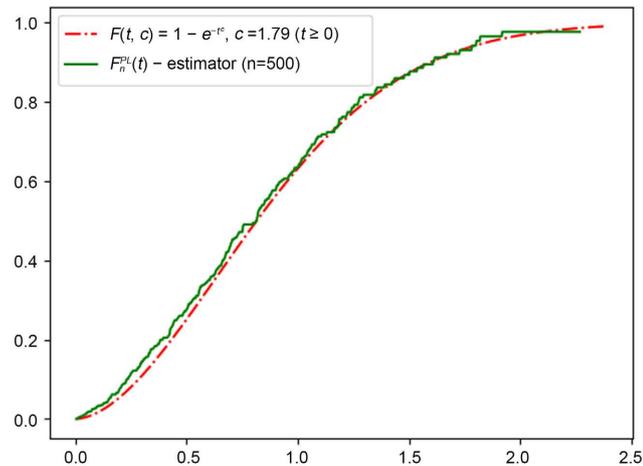


Figure 1.  $F_n^{PL}$ -Estimator (Kaplan-Meier).

Table 1. Comparison of  $F_n^{PL}(t)$ -estimate with  $F_n^{RR}(t)$ -estimate.

Sum	$\sum_{i=1}^{n-1} (F_n^{PL}(Z_i) - F(Z_i))^2$	$\sum_{i=1}^{n-1} (F_n^{RR}(Z_i) - F(Z_i))^2$	Relative
Value	0.204858	0.123901	0.604814

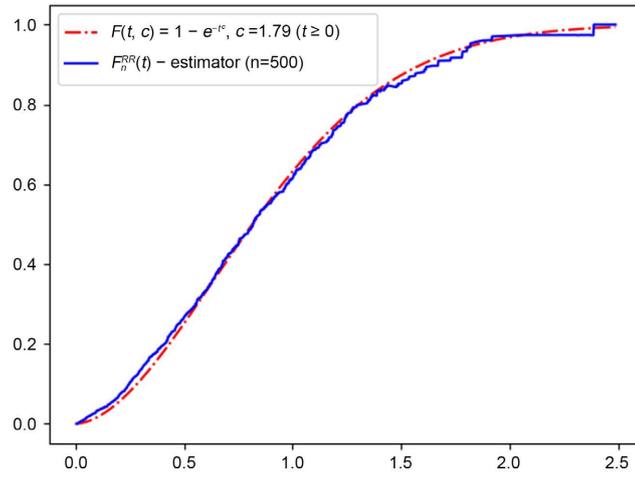


Figure 2.  $F_n^{RR}$ -Estimator (Abdushukurov).

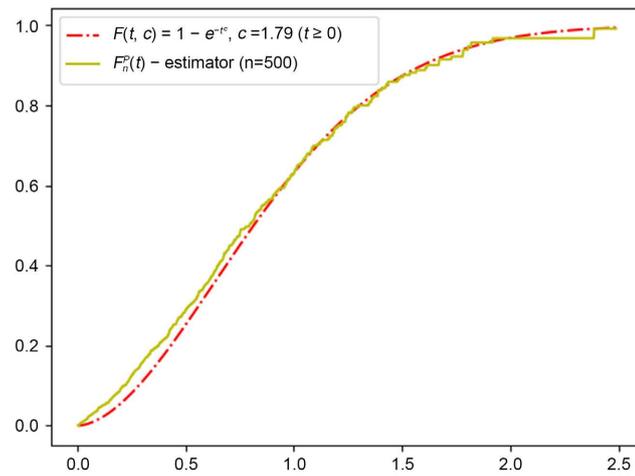


Figure 3.  $F_n^P$ -Estimator.

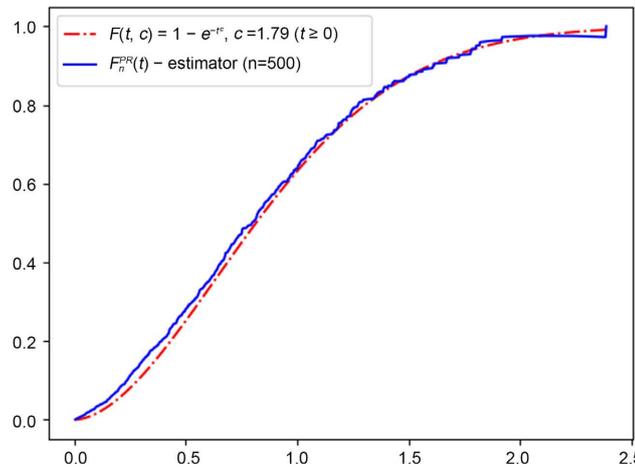


Figure 4.  $F_n^{PR}$ -Estimator.

**Table 2.** Comparison of  $F_n^p(t)$ -estimate with  $F_n^{PR}(t)$ -estimate.

Bandwith	$\sum_{i=1}^n (F_n^p(Z_i) - F(Z_i))^2$	$\sum_{i=1}^n (F_n^{PR}(Z_i) - F(Z_i))^2$	Relative
$h(n) = \frac{1}{\sqrt{n}}$	0.294402	0.277757	0.943462
$h(n) = \frac{1}{\sqrt[3]{n}}$	0.23152	0.210734	0.910219
$h(n) = \frac{1}{\sqrt[5]{n}}$	0.181307	0.178650	0.985345

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observation. *Journal of the American Statistical Association*, **53**, 457-481. <https://doi.org/10.1080/01621459.1958.10501452>
- [2] Abdushukurov, A.A. (1998) Nonparametric Estimation of the Distribution Function Based on Relative-Risk Function. *Communications in Statistics: Theory and Methods*, **27**, 1991-2012. <https://doi.org/10.1080/03610929808832205>
- [3] Abdushukurov, A.A. (1999) On Nonparametric Estimation of Reliability Indices by Censored Samples. *Theory of Probability & Its Applications*, **43**, 3-11. <https://doi.org/10.1137/S0040585X97976702>
- [4] Abdushukurov, A.A. (2011) Estimation of Unknown Distributions from Incomplete Observations and their Properties. LAP Lambert Academic, Saarbrücken. (in Russian)
- [5] Abdushukurov, A.A. (1987) Nonparametric Estimation in Proportional Hazards Model of Random Censorship. VINITI 3448 (B87).
- [6] Cheng, P.E. and Lin, G.D. (1987) Maximum Likelihood Estimation of Survival Function under the Koziol-Green Proportional Hazards Model. *Statistics & Probability Letters*, **5**, 75-80. [https://doi.org/10.1016/0167-7152\(87\)90030-7](https://doi.org/10.1016/0167-7152(87)90030-7)
- [7] Csörgő, S. (1988) Estimation in the Proportional Hazards Model of Random Censorship. *Statistics*, **19**, 437-463. <https://doi.org/10.1080/02331888808802115>
- [8] Nadaraya, E.A. (1964) On Estimating Regression. *Probability Theory and Related Fields*, **61**, 405-415.
- [9] Watson, G.S. (1964) Smooth Regression Analysis. *Sankhya: The Indian Journal of Statistics, Series A*, **26**, 359-372.
- [10] Dikta, J. (1998) On Semiparametric Random Censorship Models. *Journal of Statistical Planning and Inference*, **66**, 253-279. [https://doi.org/10.1016/S0378-3758\(97\)00091-8](https://doi.org/10.1016/S0378-3758(97)00091-8)
- [11] Cao, R., Lopez-de-Ullibarri, I., Janssen, P. and Veraverbeke, N. (2005) Presmoothed Kaplan-Meier and Nelson-Aalen Estimators. *Journal of Nonparametric Statistics*, **17**, 31-56. <https://doi.org/10.1080/10485250410001713981>
- [12] Jacome, M.A. and Cao, R. (2007) Almost Sure Asymptotic Representation for the

Presmoothed Distribution and Density Estimators for Censored Data. *Statistics*, **41**, 517-534. <https://doi.org/10.1080/02331880701529522>

- [13] Massart, P. (1990) The Tight Constant in the Dworetzky-Kiefer-Wolfowitz Inequality. *Annals of Probability*, **18**, 1269-1283. <https://doi.org/10.1214/aop/1176990746>
- [14] Burke, M.D., Csörgő, S. and Horvath, L. (1988) A Correction to and Improvement of “Strong Approximations of Some Biometric Estimates under Random Censorship”. *Probability Theory and Related Fields*, **79**, 51-57. <https://doi.org/10.1007/BF00319103>