МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL MODELLING

DOI: 10.18287/2541-7525-2023-29-3-72-78

УДК 512.531; 519.7

Дата: поступления статьи: 24.07.2023
после рецензирования: 31.08.2023
принятия статьи: 30.10.2023

А.А. Абдушукуров
Московский государственный университет, филиал в г. Ташкенте, Ташкент, Узбекистан
E-mail: a_abdushukurov@rambler.ru. ORCID: https://orcid.org/0000-0002-0994-8127
С.Б. Бозоров
Гулистанский государственный университет, Гулистан, Узбекистан
E-mail: suxrobbek_8912@mail.ru. ORCID: https://orcid.org/0009-0001-8133-4963

СРАВНЕНИЕ НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК ФУНКЦИИ ВЫЖИВАНИЯ

АННОТАЦИЯ

В статье проводится сравнение трех видов оценок: экспоненциальной, множительной и степенной структур для функции выживания при случайном цензурировании наблюдений справа. Ранее было установлено, что все эти три оценки при растущем объеме выборки эквивалентны, т. е. при одинаковой центровке и нормировке сходятся к одному и тому же гауссовскому процессу. Конкретно в выборке показано, что степенные оценки определены на всей прямой в отличие от экспоненциальной и множительных оценок. Следовательно, степенные оценки являются лучше, чем остальные две. Подвергнутые цензуре данные используются при анализе выживаемости, в биомедицинских испытаниях, в промышленных экспериментах. Существует несколько схем цензурирования (справа, слева, с обеих сторон, в сочетании с конкурирующими рисками и другими). Однако в статистической литературе широко распространено правостороннее случайное цензурирование, поскольку его легко описать с методологической точки зрения. В статье также рассмотрен этот вид цензурирования, чтобы сравнить наши результаты с другими исследованиями.

Ключевые слова: оценки; случайное цензурирование справа; функция выживания; доверительные полосы.

Цитирование. Абдушукуров А.А., Бозоров С.Б. Сравнение непараметрических оценок функции выживания // Вестник Самарского университета. Естественнонаучная серия / Vestnik of Samara University. Natural Science Series. 2023. Т. 29, № 3. С. 72–78. DOI: http://doi.org/10.18287/2541-7525-2023-29-3-72-78.

Информация о конфликте интересов: авторы и рецензенты заявляют об отсутствии конфликта интересов.

Абдурахим Ахмедович Абдушукуров — профессор кафедры прикладной математики и информатики, Московский государственный университет имени М.В. Ломоносова, филиал в г. Ташкент, 100060, Узбекистан, г. Ташкент, пр. Амира Темура, 22.
Сухроб Баходирович Бозоров — докторант кафедры математики факультета информационных технологий, Гулистанский государственный университет, 120100, Узбекистан, Сырдарьинская область, г. Гулистан, 4 микрорайон, 1.

1. Предварительные сведения

Исследования непараметрических оценок, экспоненциальной, множительной и степенной структур показывают их асимптотическую эквивалентность (при $n \to \infty$). Некоторые отличительные свойства этих оценок проявляются при фиксированном объеме выборки, и они проведены в монографии [1].

Пусть $\{Z_j,\, j \geqslant 1\}$ и $\{Y_j, j \geqslant 1\}$ — взаимонезависимые последовательности, независимые и одинаково распределенные случайная величина с непрерывными функциями распределения H и G соответственно. Наблюдается выборка объема n:

$$C^{(n)} = \{(\xi_j, \Delta_j),\ \ 1 \leqslant j \leqslant n\},$$

где

$$\xi_j = \min\left(Z_j; Y_j\right),$$
$$\Delta_j = \ \ I(Z_j \leqslant Y_j)$$

($I\,(A)$ — это индикатор события A.

1. Если $Z_j \leqslant Y_j$, то $\xi_j = \min\left(\mathrm{Z_j}; \mathrm{Y_j}\right) = \mathrm{Z_j}$, $\Delta_j = 1$, и в этом случае мы можем наблюдать Z;

2. Если $Y_j \leqslant Z_j$, то $\xi_j = \min\left(\mathrm{Z_j}; \mathrm{Y_j}\right) = \mathrm{Y_j}$, $\Delta_j = 0$, это будет случай цензурирования.

Задача состоит в оценивании функции выживания $1-H(x)$ по выборке $C^{(n)}$ при мешающей функции распределения G. Для $1 - H$ справедливо представление [2]:

$$1 - H(x) = \exp(-\Lambda(x; 1)),$$

где

$$\Lambda\left(x;1\right) = \int\limits_{\left(-\infty;x\right]} \left(1 - H\left(u-\right)\right)^{-1} dH\left(u\right) = \int\limits_{\left(-\infty;x\right]} \left(1 - N\left(u-\right)\right)^{-1} dM\left(u;1\right),$$

$$N\left(x\right) = P\left(\xi_j \leqslant x\right) = 1 - \left(1 - H\left(x\right)\right)\left(1 - G\left(x\right)\right) = M\left(x;1\right) + M\left(x;0\right),$$

$$M\left(x;1\right) = P\left(\xi_j \leqslant x,\, \Delta_j = i\right),\ i = 0;1.$$

$$\begin{gathered} H_{1n}\left(x\right) = 1 - \prod_{u \leqslant x} \exp\left\{-\tfrac{M_n(u;1) - M_n(u-;1)}{1 - N_n(u-)}\right\} = 1 - \exp\left(-\Lambda_n\left(x;1\right)\right), \\ H_{2n}\left(x\right) = 1 - \prod_{u \leqslant x} \exp\left\{1 - \tfrac{M_n(u;1) - M_n(u-;1)}{1 - N_n(u-)}\right\}, \\ H_{3n}\left(x\right) = 1 - \left(1 - N_n\left(x\right)\right)^{R_n(x)}, \end{gathered} \tag{1}$$

где

$$R_n(x) \ \ = \Lambda_n(x;1)(\Lambda_n(x))^{-1},$$

$$\Lambda_n\left(x;1\right) = \int\limits_{\left(-\infty;x\right]} \left(1 - N_n\left(u-\right)\right)^{-1} dM_n\left(u;1\right),$$

$$\Lambda_n\left(x\right) = \int\limits_{\left(-\infty;x\right]} \left(1 - N_n\left(u-\right)\right)^{-1} dN_n\left(u\right),$$

$$N_n\left(x\right) = M_n\left(x;1\right) + M_n\left(x;0\right) = \frac{1}{n}\sum_{j=1}^{n} I\left(\xi_j \leqslant x\right),$$

$$M_n\left(x;i\right) = \frac{1}{n}\sum_{j=1}^{n} I\left(\xi_j \leqslant x, \Delta_j = i\right),\ i = 0, 1.$$

Таким образом, рассматриваемая модель является моделью случайного цензурирования справа Z_j при помощи Y_j, где Z_j наблюдаемы лишь при $\Delta_j = \ 1$.

Пусть $G_{1n}(x)$, $G_{2n}(x)$ и $G_{3n}(x)$ соответствующие оценки мешающей функции распределения $G(x)$, определяемые формулами (1) с заменой $M_n(x;1)$ на $M_n(x;0)$. В рассматриваемой модели $1 - N(x) = (1 - H(x))(1 - G(x))$ для всех $x \in \mathbb{R}$. Однако для этих трех типов оценок имеем:

I.

$$(1 - H_{1n}(x))(1 - G_{1n}(x)) \ \ = \exp(-\Lambda_n(x)) \neq 1 - N_n(x)$$

и при

$$x \geqslant \xi_{(n)} = \max_{1 \leqslant i \leqslant n}\ \{\xi_i\},$$

$$\max\left(H_{1n}(x); G_{1n}(x)\right) < 1.$$

II.

$$(1 - H_{2n}(x))(1 - G_{2n}(x)) \neq 1 - N_n(x)$$

и при

$$x \geqslant \xi_{(n)}$$

оценки $H_{2n}(x)$ и $G_{2n}(x)$ неопределенны.

III. Для степенных оценок

$$(1-H_{3n}(x))(1-G_{3n}(x))=1-N_n(x)$$

и, следовательно, при $x\geqslant\xi_{(n)}$, $H_{2n}(x)=G_{2n}(x)=1$.

Таким образом, для случая непрерывных распределений H и G, только оценки степенной структуры H_{3n} и G_{3n} являются идентифицируемыми с моделью. Для демонстрации свойств оценок (1) рассмотрим выборку объема $n=97$ из работ [3; 5]. Это данные из центра уединения Ченнинг Хаус (Channing House) в г. Пало Альто (Palo Alto) в Калифорнии (США). Вариационный ряд, построенный по этим данным, есть:

(777;1), (781;0), (843;0), (866;0), (869;1), (872;1), (876;1), (893;1), (894;1), (895;0), (898;1), (906;0), (907;1), (909;1), (911;1), (911;0), (914;0), (927;1), (932;1), (936;0), (940;0), (942,5;0), (943;0), (945;1), (945;0), (948;1), (951;0), (953;0), (956;0), (957;1), (957;0), (959;0), (960;0), (966;1), (966;0), (969;1), (970;0), (971;1), (972;0), (973;0), (977;0), (983;1), (984;0), (985;1), (989;1), (992,5;1), (993;1), (996;1), (998;1), (1001;0), (1002;0), (1005;0), (1006;0), (1009;1), (1011,5;1), (1012;1), (1012;0), (1013;0), (1015;0), (1016;0), (1018;0), (1022;1), (1023;0), (1025;1), (1027;0), (1029;1), (1031;1), (1031;0), (1031,5;0), (1033;1), (1036;1), (1043;1), (1043;0), (1044;1), (1044;0), (1045;0), (1047;0), (1053;1), (1055;1), (1058;0), (1059;1), (1060;1), (1060;0), (1064;0), (1070;0), (1073;0), (1080;1), (1085;1), (1093;0), (1093,5;1), (1094;1), (1106;0), (1107;0), (1118;0), (1128;1), (1139;1), (1153;0).

Здесь данные представлены в месяцах, причем находящееся с рядом число 1 в парах означает нецензурирование (т. е. смерть), а 0 — цензурирование. При этом 46 человек умерли с начала открытия центра в 1964 году по 1 июля 1975 года ко дню сбора данных. Это нецензурированные данные. Из остальных данных о 51 человеке 5 были выписаны из центра, а 46 еще были живы к 1 июля 1975 года. Это цензурированные данные. По этим 97 данным приведены графики оценок $H_{m;97}(x)$, $m=1,2,3$ на рис. 1–3 по отдельности и на рис. 4 вместе:

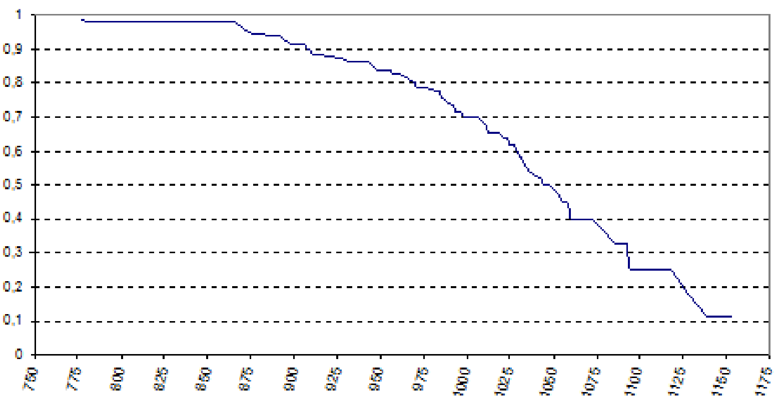

Рис. 1. Оценка $1-H_{1;97}(x)$
Fig. 1. Estimator $1-H_{1;97}(x)$

Из рисунков видно, что в отличие от экспоненциальных и множительных оценок только степенные оценки определены на всей прямой. Теперь при помощи оценок (1) построим доверительные полосы для неизвестной функции $1-H(x)$. Для этого будем следовать работам [3; 4] и используем доверительные полосы вида

$$M^*_{mn}(x,\mu_1,\mu_2)=\left[\hat{M}^{(1)}_{mn}(x,\mu_1,\mu_2);M^{(2)}_{mn}(x,\mu_1,\mu_2)\right],$$

где $m=1,2,3$,

$$\hat{M}^{(1)}_{mn}(x,\mu_1,\mu_2)=H_{mn}(x)-n^{-\frac{1}{2}}(1-H_{mn}(x))\left(\mu_1 d_n^{\frac{1}{2}}(T)+\mu_2\cdot\frac{d_n(x)}{d_n^{\frac{1}{2}}(T)}\right),$$

$$M^{(2)}_{mn}(x,\mu_1,\mu_2)=\frac{H_{mn}(x)+n^{-\frac{1}{2}}\left(\mu_1 d_n^{\frac{1}{2}}(T)+\mu_2\frac{d_n(x)}{d_n^{\frac{1}{2}}(T)}\right)}{1+n^{-\frac{1}{2}}\left(\mu_1 d_n^{\frac{1}{2}}(T)+\mu_2\frac{d_n(x)}{d_n^{\frac{1}{2}}(T)}\right)},$$

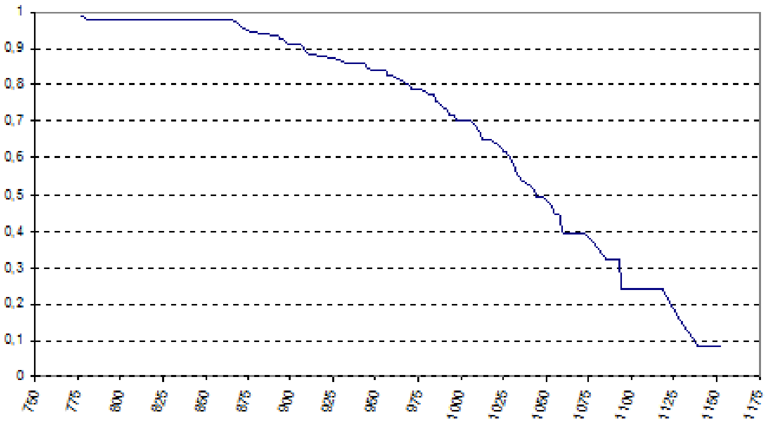

Рис. 2. Оценка $1-H_{2;97}(x)$
Fig. 2. Estimator $1-H_{2;97}(x)$

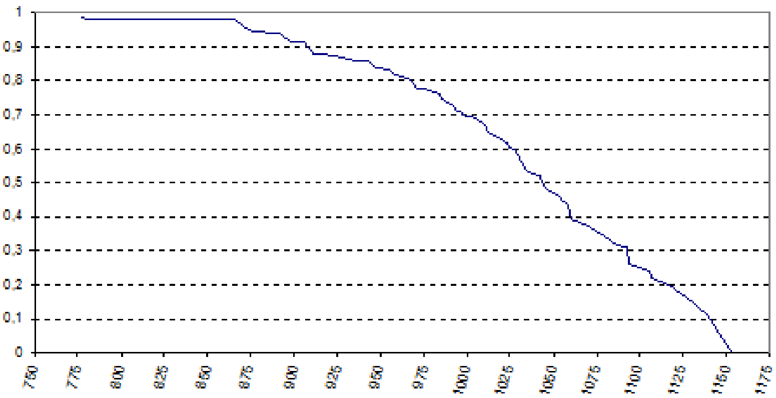

Рис. 3. Оценка $1-H_{3;97}(x)$
Fig. 3. Estimator $1-H_{3;97}(x)$

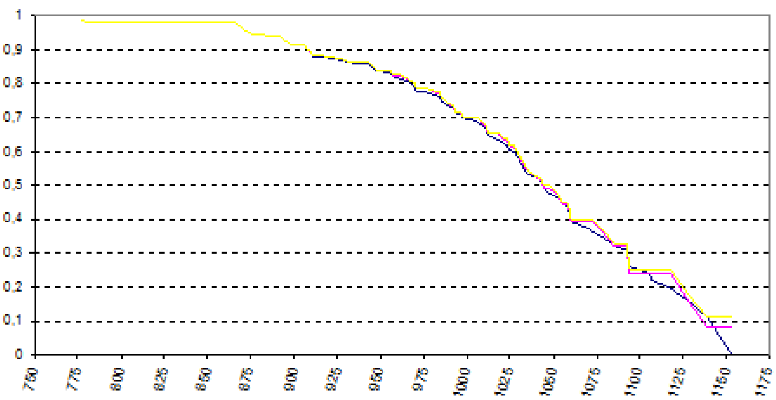

Рис. 4. Оценка $1-H_{m;97}(x)$, $m=\ 1,2,3$
Fig. 4. Estimator $1-H_{m;97}(x)$, $m=\ 1,2,3$

$T=1128$; $\mu_1=1$; $\mu_2=\ 1,37$ и $d_n\left(x\right)=\int\limits_{(-\infty;x]}\left(1-N_n\left(u-\right)\right)^{-2}dM_n\left(u;1\right)$. Эти полосы для данных объема n=97 с использованием оценок (1) приведены на рис. 5–7.

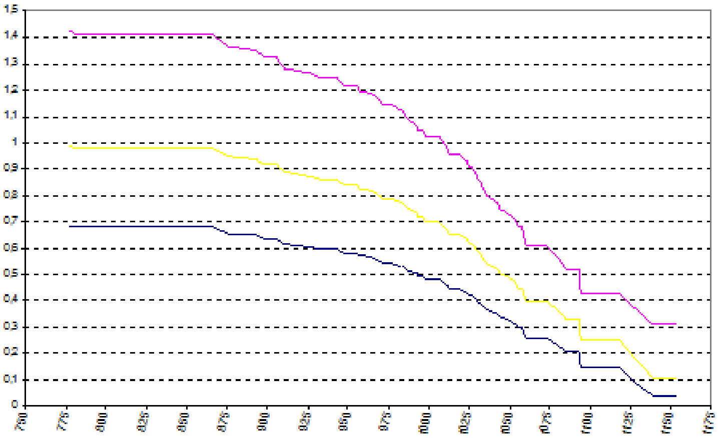

Рис. 5. Доверительные полосы $M_1{}^*{}_{;97}(x; 1; 1,37)$
Fig. 5. Confidence bands $M_1{}^*{}_{;97}(x; 1; 1,37)$

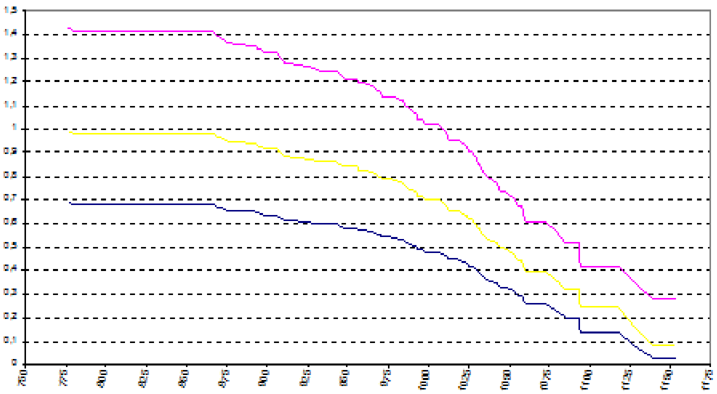

Рис. 6. Доверительные полосы $M_2{}^*{}_{;97}(x;\ 1;\ 1,37)$
Fig. 6. Confidence bands $M_2{}^*{}_{;97}(x;\ 1;\ 1,37)$

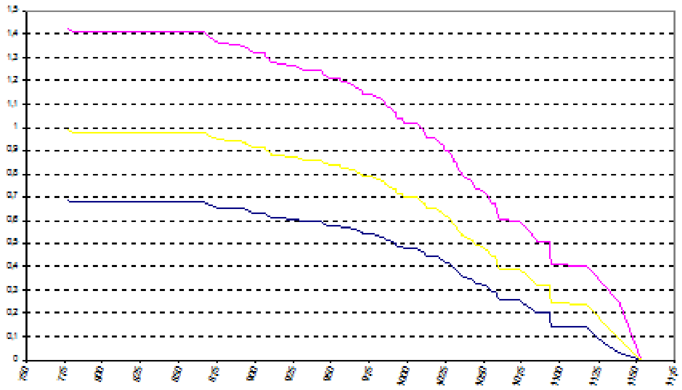

Рис. 7. Доверительные полосы $M_3{}^*{}_{;97}(x;\ 1;\ 1,37)$
Fig. 7. Confidence bands $M_3{}^*{}_{;97}(x;\ 1;\ 1,37)$

Заключение

Сравнивают три вида оценок: экспоненциальной, множительной и степенной для функции выживания при случайном цензурировании справа. Ранее была установлена асимптотическая эквивалентность этих трех видов оценок при растущем объеме выборки в смысле сходимости к одному и тому же гауссовскому процессу. Для конкретной конечной выборки объема $n = 97$ показаны некоторые преимущества степенной оценки по сравнению с остальными двумя. Следовательно, эта оценка лучше, чем остальные. Имеются численные примеры демонстрации результатов.

Литература

[1] Абдушукуров А.А. Статистика неполных наблюдений. Ташкент: Университет, 2009. 269 с.

[2] Abdushukurov A.A., Bozorov S.B., Nurmukhamedova N.S. Nonparametric Estimation of Distribution Function Under Right Random Censoring Based on Presmoothed Relative — Risk Function // Lobachevskii Journal of Mathematics, 2021, vol. 42, no. 2, pp. 257–268. DOI: https://doi.org/10.1134/S1995080221020049.

[3] Csörgő S. Estimating in the proportional hazards model of random censorship // Statistics. 1988. Vol. 19, Issue 3. Pp. 437–463. DOI: https://doi.org/10.1080/02331888808802115.

[4] Csörgő S., Horvath L. Confidence bands from censored samples // Canadian Journal of Statistics-revue Canadienne De Statistique. 1986. Vol. 14, Issue 2. Pp. 131–144. DOI: https://doi.org/10.2307/3314659.

[5] Efron B. Censored Data and the Bootstrap // Journal of the American Statistical Association, 1981, vol. 76, № 374, pp. 312–319. DOI: http://doi.org/10.2307/2287832.

Scientific article

DOI: 10.18287/2541-7525-2023-29-3-72-78

Submited: 24.07.2023
Revised: 31.08.2023
Accepted: 30.10.2023

A.A. Abdushukurov
Lomonosov Moscow State University, Tashkent branch, Tashkent, Uzbekistan
E-mail: a_abdushukurov@rambler.ru. ORCID: https://orcid.org/0000-0002-0994-8127
S.B. Bozorov
Gulistan State University, Gulistan, Uzbekistan
E-mail:suxrobbek_8912@mail.ru. ORCID: https://orcid.org/0009-0001-8133-4963

COMPARISON OF NONPARAMETRIC ESTIMATES OF THE SURVIVAL FUNCTIONS

ABSTRACT

The article compares three types of estimates: exponential, multiplying and power structures for the survival function of three random censoring observations on the right. It was previously established that all these three estimates are equivalent with a growing sample size, i.e. three with the same centering and normalization converge to the same Gaussian process. For specific samples, it is shown that power estimates are defined on the entire line, in contrast to exponential and multiply estimates. Therefore, power estimates are better than the other two. Censored data is used in survival analyses, biomedical trials, and industrial experiments. There are several censoring schemes (right, left, both sides, combined with competing risks, and others). However, right-sided random censoring is common in the statistical literature because it is easy to describe from a methodological point of view. Here we also consider this type of censoring, to compare our results with others.

Key words: estimators; random censorship from the right; survival function; confidence bands.

Citation. Abdushukurov A.A., Bozorov S.B. Comparison of nonparametric estimates of the survival functions. *Vestnik Samarskogo universiteta. Estestvennonauchnaya seriya / Vestnik of Samara University. Natural Science Series*, 2023, vol. 29, no. 3, pp. 72–78. DOI: http://doi.org/10.18287/2541-7525-2023-29-3-72-78. (In Russ.)

Information about the conflict of interests: authors and reviewers declare no conflict of interests.

Abdurakhim A. Abdushukurov — professor of the Department of Applied Mathematics and Informatics, Lomonosov Moscow State University, Tashkent branch, 22, Amir Temur Street, Tashkent, 100060, Uzbekistan.

Sukhrob B. Bozorov — Doctoral student of the Department of Mathematics, Faculty of Information Technology, Gulistan State University, 4th House of Saodat Street in the neighborhood of Gulistan City Mevazor, Gulistan, 120100, Uzbekistan.

References

[1] Abdushukurov A.A. Statistics of incomplete observations. Tashkent: Universitet, 2009, 269 p. (In Russ.)

[2] Abdushukurov A.A., Bozorov S.B., Nurmukhamedova N.S. Nonparametric Estimation of Distribution Function Under Right Random Censoring Based on Presmoothed Relative - Risk Function. *Lobachevskii Journal of Mathematics*, 2021, vol. 42, no. 2, pp. 257–268. DOI: https://doi.org/10.1134/S1995080221020049.

[3] Csörgő S. Estimating in the proportional hazards model of random censorship. *Statistics*, 1988, vol. 19, issue 3, pp. 437–463. DOI: https://doi.org/10.1080/02331888808802115.

[4] Csörgő S., Horvath L. Confidence bands from censored samples. *Canadian Journal of Statistics-revue Canadienne De Statistique*, 1986, vol. 14, № 2, pp. 131–144. DOI: https://doi.org/10.2307/3314659.

[5] Efron B. Censored Data and the Bootstrap. *Journal of the American Statistical Association*, 1981, vol. 76, no. 374, pp. 312–319. DOI: https://doi.org/10.2307/2287832.