ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

DOI: 10.18698/1812-3368-2015-6-68-84
УДК 519.248

ПРИМЕНЕНИЕ ОЦЕНОК КАПЛАНА – МЕЙЕРА ДЛЯ ПРОВЕРКИ СТЕПЕННОЙ ГИПОТЕЗЫ КОКСА ПО ДВУМ ПРОГРЕССИВНО ЦЕНЗУРИРОВАННЫМ ВЫБОРКАМ

В.И. Тимонин, **Н.Д. Тянникова**

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация
e-mail: timoninmgtu52@mail.ru; tiannikova@yandex.ru

Рассмотрена задача проверки степенной гипотезы Кокса для двух прогрессивно цензурированных выборок. В качестве статистики для проверки данной гипотезы предложен критерий типа Колмогорова – Смирнова, основанный на сравнении оценок Каплана – Мейера функции надежности по каждой выборке. Разработан метод вычисления точных распределений статистики, основанный на модели случайного блуждания частицы по двумерному массиву ячеек. Метод позволяет получать точные вероятности для очень значительных объемов выборок, что дает возможность оценить их необходимый объем, для которого точные вероятности можно заменить асимптотическими. Рассчитаны таблицы значений вероятностей точных распределений предложенной статистики для широкого набора возможных значений объемов выборок. Показана сходимость асимптотического распределения данной статистики к стандартному распределению Колмогорова – Смирнова при условии справедливости проверяемой гипотезы. Исследованы статистические свойства оценки параметра модели Кокса в случае справедливости проверяемой гипотезы методом Монте-Карло. В качестве оценки рассмотрено значение параметра, минимизирующее предлагаемую статистику критерия. Показана состоятельность оценки.

***Ключевые слова*:** непараметрическая статистика, гипотеза Кокса, критерий типа Колмогорова – Смирнова, оценка Каплана – Мейера.

APPLICATION OF KAPLAN – MEIER ESTIMATES TO TESTING THE COX POWER HYPOTHESIS FOR TWO PROGRESSIVE LYCENSORED SAMPLES

V.I. Timonin, **N.D. Tyannikova**

Bauman Moscow State Technical University, Moscow, Russian Federation
e-mail: timoninmgtu52@mail.ru; tiannikova@yandex.ru

The paper considers the problem of testing the Cox power hypothesis for two progressively censored samples. To test the hypothesis, the authors propose the Kolmogorov – Smirnov criterion as some statistics based on comparison of the Kaplan – Meier estimates of a reliability function for each sample. A method for calculating the exact distributions of statistics is based on the model of a particle random walk on a two-dimensional cell array. The method allows obtaining accurate probabilities for a considerable amount of samples. It enables one to estimate their required amount for which the exact probabilities can be replaced by the asymptotic *ones. Tables of probability distributions for the proposed accurate statistics are calculated considering a wide range of possible values of the sample amount. The authors show that the statistics asymptotic distribution converges to the Kolmogorov–Smirnov standard distribution, provided the tested hypothesis is valid. The statistical characteristics for estimating the Cox model parameter are evaluated, in case the hypothesis is valid, if tested with the help of the Monte-Carlo method. The value that minimizes the proposed test statistics is considered estimation. The estimation consistency is shown.*

Keywords: nonparametric statistics, Cox hypothesis, Kolmogorov – Smirnov criterion, Kaplan – Meier estimator.

Введение. При испытаниях технических систем часто возникает задача проверки степенной гипотезы

$$H_0 : P_1(t) = (P_2(t))^k \,, \tag{1}$$

где $P_i(t) = 1 - F_i(t)$, $i = 1, 2$ — функции надежности для первой и второй независимых выборок. Выполнение этой гипотезы эквивалентно пропорциональности интенсивностей отказов. Впервые модель (1) была рассмотрена Д. Коксом [1]. В настоящей работе исследована задача проверки гипотезы (1) в случае, когда обе выборки являются прогрессивно цензурированными. Эта работа представляет собой обобщение работы [2], в которой была предложена статистика для проверки гипотезы однородности, что эквивалентно случаю $k = 1$.

Постановка задачи. Пусть имеется n_1 систем, состоящих из m_1 последовательно соединенных элементов, которые работают в режиме ε_1, и n_2 систем, состоящих из m_2 последовательно соединенных элементов, функционирующих в режиме ε_2. При испытаниях последовательных систем в случае отказа одного из изделий в системе, оставшиеся $(m_j - 1)$ изделий цензурируются $(j = 1, 2)$. Таким образом, по результатам испытаний двух групп систем имеются две прогрессивно цензурированные выборки $\Theta_1 = (\theta_1^1, \ldots, \theta_1^{n_1})$, $\Theta_2 = (\theta_2^1, \ldots, \theta_2^{n_2})$, где $\theta_1^i = \min\left\{\xi_1^i, \xi_2^i, \ldots, \xi_{m_1}^i\right\}$, $i = \overline{1, n_1}$, $\theta_2^j = \min\left\{\xi_1^j, \xi_2^j, \ldots, \xi_{m_2}^j\right\}$, $j = \overline{1, n_2}$ — минимумы наработок до отказа элементов систем, работающих в режимах ε_1 и ε_2 [3, 4].

Требуется проверить гипотезу о том, что интенсивности отказов $\lambda_1(t), \lambda_2(t)$ элементов для двух режимов пропорциональны $\lambda_1(t) = k\lambda_2(t)$, где $k \geq 1$ — известное фиксированное число. Обозначим $P_1(t)$ функцией надежности наработок до отказа элементов в режиме ε_1, а $P_2(t)$ функцией надежности наработок до отказа в режиме ε_2. Тогда проверяемая гипотеза имеет вид, аналогичный виду модели (1): $H_0 : P_1(t) = (P_2(t))^k$, $k \geq 1$.

В настоящей работе для проверки гипотезы (1) предложен критерий типа Колмогорова – Смирнова, основанный на сравнении оценок Каплана – Мейера $\overset{\frown}{P}_{\theta_1}(t)$, $\overset{\frown}{P}_{\theta_2}(t)$ функций надежности $P_1(t)$, $P_2(t)$ по прогрессивно цензурированным выборкам Θ_1 и Θ_2 [5, 6]:

$$\widehat{P}_{\theta_1}(t) = \begin{cases} 1, d_1(t) = 0, \\ \prod\limits_{i=1}^{d_1(t)} \left(1 - \dfrac{1}{m_1(n_1 - i + 1)}\right), \quad 1 \le d_1(t) \le (n_1 - 1), \\ 0, d_1(t) = n_1; \end{cases}$$

$$\widehat{P}_{\theta_2}(t) = \begin{cases} 1, d_2(t) = 0, \\ \prod\limits_{j=1}^{d_2(t)} \left(1 - \dfrac{1}{m_2(n_2 - j + 1)}\right), 1 \le d_2(t) \le (n_2 - 1), \\ 0, d_2(t) = n_2. \end{cases}$$

Здесь $d_1(t), d_2(t)$ — число элементов выборок Θ_1и Θ_2, меньших значения t.

Отметим, что выборки $(\theta_1^1, \ldots, \theta_1^{n_1}), (\theta_2^1, \ldots, \theta_2^{n_2})$ можно рассматривать и как полные (нецензурированные) независимые выборки из совокупностей с функциями распределения $F^1(t) = 1 - (P_1(t))^{m_1}, F^2(t) = 1 - (P_2(t))^{m_2}$. При этом функции распределения $F^1(t), F^2(t)$ можно оценить обычными эмпирическими функциями распределения $\widehat{F}^1 = \dfrac{d_1(t)}{n_1}, \widehat{F}^2 = \dfrac{d_2(t)}{n_2}$.

Для проверки гипотезы (1) предложена статистика вида

$$T = \frac{m_1 m_2 \sqrt{\rho n_2}}{\sqrt{k^2 \rho m_1^2 + m_2^2}} \times$$
$$\times \max_t \frac{\left(k_2 \left(1 - \widehat{F}^1\right)^{\frac{1}{m_1}} + k_1 \left(1 - \widehat{F}^2\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - 1}}{k_2 \left(k_2 \left(1 - \widehat{F}^1\right)^{\frac{1}{m_1}} + k_1 \left(1 - \widehat{F}^2\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - m_1} + k_1} \times$$
$$\times \left| \widehat{P}_{\theta_1}(t) - \left(\widehat{P}_{\theta_2}(t)\right)^k \right|, \quad (2)$$

где$\rho = \dfrac{n_1}{n_2}$; $k_2 = \dfrac{m_2^2}{\rho m_1^2 k^2 + m_2^2}$; $k_1 = \dfrac{\rho m_1^2 k^2}{\rho m_1^2 k^2 + m_2^2}$. При этом для случая $\dfrac{m_2}{k} - 1 < 0$ и $k_2 \left(1 - \widehat{F}^1\right)^{\frac{1}{m_1}} + k_1 \left(1 - \widehat{F}^2\right)^{\frac{k}{m_2}} = 0$ примем

$$\frac{\left(k_2\left(1-\overset{\frown}{F}^{1}\right)^{\frac{1}{m_1}}+k_1\left(1-\overset{\frown}{F}^{2}\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k}-1}}{k_2\left(k_2\left(1-\overset{\frown}{F}^{1}\right)^{\frac{1}{m_1}}+k_1\left(1-\overset{\frown}{F}^{2}\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k}-m_1}+k_1}=0.$$

Точные распределения *T*. Для вычисления точных распределений статистики (2) рассмотрим более общую постановку задачи. Пусть испытания проводят следующим образом. Имеется два режима работы: в режиме ε_1 работает N_1 элементов, в режиме ε_2 — N_2 элементов. При i-м отказе одного из изделий в режиме ε_1 с испытаний снимаются (цензурируются) r_i^1 изделий $(i=\overline{1,q_1})$, в режиме ε_2 при j-м отказе — r_j^2 изделий $(j=\overline{1,q_2})$. Такая схема цензурирования называется прогрессивным цензурированием [4, 7, 8], причем $N_1=\sum\limits_{i=1}^{q_1} r_i^1+q_1$, $N_2=\sum\limits_{j=1}^{q_2} r_j^2+q_2$. По результатам испытаний имеется две прогрессивно цензурированные выборки Q_1, Q_2 объемами q_1, q_2. Уточним, что параметры r_i^1, r_j^2 известны заранее и не являются случайными величинами.

Введем вектор $\overrightarrow{Z}=(z_1,z_2,\ldots,z_{q_1+q_2})$, который состоит из q_1 единиц и q_2 нулей, причем

$$z_i=\begin{cases}1, \text{ если отказ из выборки } Q_1,\\ 0, \text{ если отказ из выборки } Q_2,\end{cases} \quad i=1,\ldots,(q_1+q_2)\,.$$

Рассмотрим следующую модель случайного блуждания [9]. Пусть $\{a_{ij}\}=A$, $i=\overline{0,q_1}$, $j=\overline{0,q_2}$ — двумерный массив ячеек. Частица на первом шаге выходит из ячейки $a_{0,0}$ и на (q_1+q_2)-м шаге заканчивает блуждание в ячейке a_{q_1,q_2}, совершая q_1 скачков "вниз" и q_2 скачков "вправо". Траектории частицы будут находиться во взаимно однозначном соответствии с вектором $\overrightarrow{Z}$. Равенство $z_k=1$, $k=1,\ldots,(q_1+q_2)$, в векторе $\overrightarrow{Z}$ соответствует скачку "вниз" на k-м шаге, появление $z_k=0$, $k=1,\ldots,(q_1+q_2)$ — скачку "вправо". Схема массива ячеек, по которым происходит случайное блуждание частицы, показана на рис. 1.

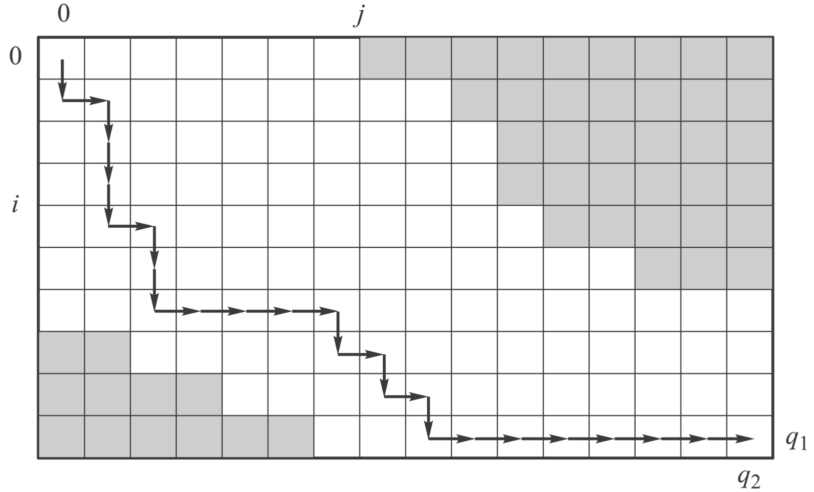

Рис. 1. Схема массива ячеек, по которым происходит случайное блуждание частицы

Теорема 1. *Распределение вероятностей вектора* $\overrightarrow{Z} = (z_1, z_2, \ldots$ $\ldots, z_{q_1+q_2})$ *определяется по следующему выражению*:

$$p\left(\overrightarrow{Z}\right) = \prod_{s=1}^{q_1+q_2} \left(\frac{\left(\left(N_1 - \sum\limits_{l=1}^{V_{s-1}} r_l^1 - V_{s-1}\right) k\right)^{z_s} \left(N_2 - \sum\limits_{l=1}^{U_{s-1}} r_l^2 - U_{s-1}\right)^{1-z_s}}{\left(N_1 - \sum\limits_{l=1}^{V_{s-1}} r_l^1 - V_{s-1}\right) k + N_2 - \sum\limits_{l=1}^{U_{s-1}} r_l^2 - U_{s-1}} \right),$$

где $V_j = \sum\limits_{i=1}^{j} z_i$, $V_0 = 0$ — *число единиц в векторе* $\overrightarrow{Z}$ *до* j*-го места включительно*; $U_j = j - \sum\limits_{i=1}^{j} z_i$, $U_0 = 0$ — *число нулей в векторе* $\overrightarrow{Z}$ *до* j*-го места включительно*.

◄ Утверждение теоремы следует из вида условных вероятностей перехода “вниз” или “вправо” в схеме случайного блуждания частицы [9, 10]. ►

Обозначим через P вероятность невыхода траектории случайного блуждания из некоторого подмножества A_0 множества $\{a_{ij}\} = A$, $i = \overline{0, q_1}$, $j = \overline{0, q_2}$.

Теорема 2. *Вероятность* P *равна* π_{q_1,q_2}, *которую можно получить повторным применением соотношения*

$$\pi_{ij}=\begin{cases} 1\,\chi_{ij}, \text{ если } i=0, j=0; \\ \left(\dfrac{N_2-\sum\limits_{l=1}^{j-1} r_l^2 - j+1}{\left(N_1-\sum\limits_{l=1}^{i} r_l^1 - i\right)k+N_2-\sum\limits_{l=1}^{j-1} r_l^2 - j+1}\pi_{i,j-1}\right)\chi_{ij}, \\ \qquad i=0,\ 1\le j\le q_2; \\ \left(\dfrac{\left(N_1-\sum\limits_{l=1}^{i-1} r_l^1 - i+1\right)k}{\left(N_1-\sum\limits_{l=1}^{i-1} r_l^1 - i+1\right)k+N_2-\sum\limits_{l=1}^{j} r_l^2 - j}\pi_{i-1,j}\right)\chi_{ij}, \\ \qquad 1\le i\le q_1,\ j=0; \\ \left(\begin{array}{l}\dfrac{\left(N_1-\sum\limits_{l=1}^{i-1} r_l^1 - i+1\right)k}{\left(N_1-\sum\limits_{l=1}^{i-1} r_l^1 - i+1\right)k+N_2-\sum\limits_{l=1}^{j} r_l^2 - j}\pi_{i-1,j}+ \\ +\dfrac{N_2-\sum\limits_{l=1}^{j-1} r_l^2 - j+1}{\left(N_1-\sum\limits_{l=1}^{i} r_l^1 - i\right)k+N_2-\sum\limits_{l=1}^{j-1} r_l^2 - j+1}\pi_{i,j-1}\end{array}\right)\chi_{ij}, \\ \qquad 1\le i\le q_1,\ 1\le j\le q_2. \end{cases}$$

Здесь $\chi_{ij}=\begin{cases} 1, & \forall\,(i,j):\ a_{ij}\in A_0; \\ 0, & \forall\,(i,j):\ a_{ij}\notin A_0. \end{cases}$

◄ Пусть ω — некоторая траектория случайного блуждания частицы. Тогда из теоремы 1 получим, что вероятность ω можно записать в виде

$$p(\omega)=$$
$$=\prod_{s=1}^{q_1+q_2}\left(\frac{\left(\left(N_1-\sum\limits_{l=1}^{V_{s-1}} r_l^1 - V_{s-1}\right)k\right)^{z_s}\left(N_2-\sum\limits_{l=1}^{U_{s-1}} r_l^2 - U_{s-1}\right)^{1-z_s}}{\left(N_1-\sum\limits_{l=1}^{V_{s-1}} r_l^1 - V_{s-1}\right)k+N_2-\sum\limits_{l=1}^{U_{s-1}} r_l^2 - U_{s-1}}\right)=$$
$$=\prod_{s=1}^{q_1+q_2}\lambda_s(\omega).$$

Обозначим через ω_t часть траектории ω, определяемую первыми t шагами (“частичная” траектория). Пусть ω_{ij} — множество “частичных” траекторий, оканчивающихся в ячейке a_{ij} (соответствующий вектор $\overrightarrow{Z}$ имеет i единиц и j нулей до $(i+j)$-го места). Обозначим через $q_t = \prod\limits_{s=1}^{t} \lambda_s(\omega)$ первые t сомножителей выражения для $p(\omega)$. Вероятность любой траектории, совершающей скачок на $(i+j)$-м шаге $a_{i-1,j} \to a_{ij}$ $(z_{i+j}=1)$, имеет множитель

$$\frac{\left(N_1 - \sum\limits_{l=1}^{i-1} r_l^1 - i + 1\right) k}{\left(N_1 - \sum\limits_{l=1}^{i-1} r_l^1 - i + 1\right) k + N_2 - \sum\limits_{l=1}^{j} r_l^2 - j}.$$

Если происходит скачок $a_{i,j-1} \to a_{ij}$ $(z_{i+j}=0)$, то соответствующий множитель равен

$$\frac{N_2 - \sum\limits_{l=1}^{j-1} r_l^2 - j + 1}{\left(N_1 - \sum\limits_{l=1}^{i} r_l^1 - i\right) k + N_2 - \sum\limits_{l=1}^{j-1} r_l^2 - j + 1}.$$

Пусть $\pi_{ij} = \sum\limits_{\omega_t \in \omega_{ij}} q_{i+j}$. Тогда утверждение теоремы 2 следует из того, что в ячейку a_{ij} за один скачок можно попасть только из ячейки $a_{i-1,j}$ или $a_{i,j-1}$. Множитель χ_{ij} обеспечивает обращение в нуль вероятностей траекторий, не лежащих в подмножестве $A_0 \subset A$. ▶

Для испытаний последовательных систем имеем $r_i^1 = m_1 - 1$, $i = \overline{1, n_1}$, $r_j^2 = m_2 - 1$, $j = \overline{1, n_2}$, $N_1 = n_1 m_1$, $N_2 = n_2 m_2$, $q_1 = n_1$, $q_2 = n_2$. В таком случае при прохождении блуждания через ячейку a_{ij} функция

$$\frac{m_1 m_2 \sqrt{\rho n_2}}{\sqrt{k^2 \rho m_1^2 + m_2^2}} \frac{\left(k_2 \left(1 - \overset{\frown}{F}{}^1\right)^{\frac{1}{m_1}} + k_1 \left(1 - \overset{\frown}{F}{}^2\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - 1}}{k_2 \left(k_2 \left(1 - \overset{\frown}{F}{}^1\right)^{\frac{1}{m_1}} + k_1 \left(1 - \overset{\frown}{F}{}^2\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - m_1} + k_1} \times$$

$$\times \left| \overset{\frown}{P}_{\theta_1}(t) - \left(\overset{\frown}{P}_{\theta_2}(t)\right)^k \right|$$

принимает значение t_{ij}, равное

$$t_{ij} = \frac{m_1 m_2 \sqrt{\rho n_2}}{\sqrt{k^2 \rho m_1^2 + m_2^2}} \times$$

$$\times \frac{\left(k_2 \left(1 - \dfrac{i}{n_1}\right)^{\frac{1}{m_1}} + k_1 \left(1 - \dfrac{j}{n_2}\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - 1}}{k_2 \left(k_2 \left(1 - \dfrac{i}{n_1}\right)^{\frac{1}{m_1}} + k_1 \left(1 - \dfrac{j}{n_2}\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - m_1} + k_1} \Delta_{ij},$$

где

$$\Delta_{ij} = \left| \prod_{s_1=1}^{i} \left(1 - \frac{1}{m_1 (n_1 - s_1 + 1)}\right) - \right.$$

$$\left. - \left(\prod_{s_2=1}^{j} \left(1 - \frac{1}{m_2 (n_2 - s_2 + 1)}\right)\right)^k \right|.$$

При этом для $m_2/k - 1 < 0$ и

$$k_2 \left(1 - \frac{i}{n_1}\right)^{\frac{1}{m_1}} + k_1 \left(1 - \frac{j}{n_2}\right)^{\frac{k}{m_2}} = 0$$

сохраняется условие

$$\frac{\left(k_2 \left(1 - \dfrac{i}{n_1}\right)^{\frac{1}{m_1}} + k_1 \left(1 - \dfrac{j}{n_2}\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - 1}}{k_2 \left(k_2 \left(1 - \dfrac{i}{n_1}\right)^{\frac{1}{m_1}} + k_1 \left(1 - \dfrac{j}{n_2}\right)^{\frac{k}{m_2}}\right)^{\frac{m_2}{k} - m_1} + k_1} = 0.$$

Отметим, что параметр t_{ij} не зависит от траектории частицы. Основываясь на описанной модели случайного блуждания предложен метод вычисления точных распределений статистики T [9].

Теорема 3. *Вероятность* $P\{T < h\}$ *равна* $\pi_{n_1,n_2}(h)$, *которую можно получить повторным применением соотношения*

$$
\pi_{ij}(h) = \begin{cases} 1, \text{ если } i=0, j=0; \\ \left(\dfrac{m_2(n_2-j+1)}{km_1(n_1-i)+m_2(n_2-j+1)}\pi_{i,j-1}(h)\right)\chi_{ij}(h), \\ \qquad i=0,\ 1\le j\le n_2; \\ \left(\dfrac{km_1(n_1-i+1)}{km_1(n_1-i+1)+m_2(n_2-j)}\pi_{i-1,j}(h)\right)\chi_{ij}(h), \\ \qquad 1\le i\le n_1,\ j=0; \\ \left(\begin{array}{l} \dfrac{km_1(n_1-i+1)}{km_1(n_1-i+1)+m_2(n_2-j)}\pi_{i-1,j}(h)+ \\ +\dfrac{m_2(n_2-j+1)}{km_1(n_1-i)+m_2(n_2-j+1)}\pi_{i,j-1}(h) \end{array}\right)\chi_{ij}(h), \\ \qquad 1\le i\le n_1,\ 1\le j\le n_2. \end{cases}
$$

Здесь $\chi_{ij}(h) = \begin{cases} 1, & t_{ij} < h; \\ 0, & t_{ij} \ge h. \end{cases}$

◀ Теорема 3 является следствием теоремы 2, при $r_i^1 = m_1 - 1$, $i = \overline{1, n_1}$, $r_j^2 = m_2 - 1, j = \overline{1, n_2}$, $N_1 = n_1 m_1$, $N_2 = n_2 m_2$, $q_1 = n_1$, $q_2 = n_2$. При этом множитель $\chi_{ij}(h)$ обеспечивает обращение в нуль вероятностей траекторий, для которых $T > h$. ▶

Значения вероятности точного распределения статистики T для квантилей $h = 1{,}22$, $1{,}36$ и $1{,}63$, которые являются квантилями уровней 0,8981, 0,9505 и 0,9901 асимптотического распределения Колмогорова – Смирнова, приведены в таблице.

Значения точной вероятности $P(T < h)$ в случае равных объемов выборок при $m_1 = m_2 = 2$ для $k = 1{,}5$ (числитель) и 3,0 (знаменатель)

$n_1 = n_2$	$P(T < h)$		
	$h = 1{,}22$	$h = 1{,}36$	$h = 1{,}63$
100	0,9108/0,8916	0,9572/0,9442	0,9913/0,9864
300	0,9060/0,9014	0,9551/0,9518	0,9911/0,9901
500	0,9046/0,9025	0,9542/0,9530	0,9909/0,9906
700	0,9041/0,9028	0,9536/0,9530	0,9908/0,9906
900	0,9033/0,9024	0,9531/0,9529	0,9908/0,9906
1100	0,9029/0,9021	0,9530/0,9528	0,9907/0,9906
1300	0,9023/0,9023	0,9528/0,9526	0,9907/0,9906
1500	0,9020/0,9020	0,9527/0,9525	0,9906/0,9906
∞	0,8981/0,8981	0,9505/0,9505	0,9901/0,9901

Асимптотическое распределение *T*. Вид предлагаемой статистики T связан с тем, что ее асимптотическое распределение может быть приближено классическим распределением Колмогорова – Смирнова.

Без ограничения общности можно полагать, что $P_1(t) = 1 - t$, $P_2(t) = (1-t)^{\frac{1}{k}}$, $0 \le t \le 1$. Рассмотрим сначала процесс $X_1(t) = \sqrt{n_1}\left(\widehat{P}_{\theta_1}(t) - (1-t)\right)$. В работе [2] была доказана следующая теорема, дающая выражение для асимптотической ковариации процесса $X_1(t)$.

Теорема 4. *Пусть* $0 \le s \le t \le \Delta < 1$, *тогда*

$$K_1(s,t) = EX_1(s)\,X_1(t) - \\ - EX_1(s)\,EX_1(t) \underset{n_1\to\infty}{\longrightarrow} (1-t)(1-s)\left(\frac{1-(1-s)^{m_1}}{m_1^2\,(1-s)^{m_1}}\right).$$

Далее рассмотрим процесс $\tilde{X}_2(t) = \sqrt{n_2}\left(\widehat{P}_{\theta_2}(t) - (1-t)^{\frac{1}{k}}\right)$, $0 \le t \le 1$.

Метод получения асимптотической ковариации процесса $\tilde{X}_2(t)$ полностью аналогичен методу определения ковариации $X_1(t)$, приведенному в работе [2], и поэтому вид асимптотической ковариации $\tilde{K}_2(s,t) = E\tilde{X}_2(s)\,\tilde{X}_2(t) - E\tilde{X}_2(s)\,E\tilde{X}_2(t)$ приводится без доказательства.

Теорема 5. *Пусть* $0 \le s \le t \le \Delta < 1$, *тогда*

$$\tilde{K}_2(s,t) = E\tilde{X}_2(s)\,\tilde{X}_2(t) - \\ - E\tilde{X}_2(s)\,E\tilde{X}_2(t) \underset{n_2\to\infty}{\longrightarrow} (1-t)^{\frac{1}{k}}(1-s)^{\frac{1}{k}}\left(\frac{1-(1-s)^{\frac{m_2}{k}}}{m_2^2\,(1-s)^{\frac{m_2}{k}}}\right).$$

Для того чтобы получить асимптотическую функцию ковариации процесса $X_2(t) = \sqrt{n_2}\left(\left(\widehat{P}_{\theta_2}(t)\right)^k - (1-t)\right), 0 \le t \le 1$, воспользуемся следующим результатом [11].

Теорема 6. *Пусть* $h_n(x) = \sqrt{n}\left(\left((1-t)^{1/k} + \dfrac{x}{\sqrt{n}}\right)^k - (1-t)\right)$, $h(x) = k(1-t)^{1-1/k}x$, $x \in D[0,1]$, $D[0,1]$ — *пространство функций без разрывов второго рода на* $[0,\ 1]$. *Тогда последовательность* $h_n(x)$ *сходится в метрике Скорохода к* $h(x)$ *равномерно на ограниченных множествах из* $D[0,1]$.

Учитывая, что асимптотически $X_2(t) = h(\tilde{X}_2(t))$, получаем асимптотическую ковариацию процесса $X_2(t)$ в виде

$$K_2(s,t) = EX_2(s)\,X_2(t) - \\ - EX_2(s)\,EX_2(t) \underset{n_2\to\infty}{\longrightarrow} (1-t)(1-s)\left(\frac{k^2\left(1-(1-s)^{\frac{m_2}{k}}\right)}{m_2^2\,(1-s)^{\frac{m_2}{k}}}\right).$$

Рассмотрим случайный процесс $Z_n(t)=\sqrt{n_1}\left(\widehat{P}_{\theta_1}(t)-\left(\widehat{P}_{\theta_2}(t)\right)^k\right)$, $0 \le t < 1$, определяющий статистику (2). Обозначим через $K_n(s,t)$ функцию ковариации $Z_n(t)$. Очевидно, что справедливо представление

$$Z_n(t) = \sqrt{n_1}\left(\left(\widehat{P}_{\theta_1}(t) - (1-t)\right) - \left(\left(\widehat{P}_{\theta_2}(t)\right)^k - (1-t)\right)\right) = \\ = X_1(t) - \sqrt{\rho}X_2(t),\ 0 \le t < 1.$$

Из этого представления следует теорема.

Теорема 7. *Пусть* $0 \le s \le t \le \Delta < 1$, *тогда*

$$K_n(s,t) \underset{\substack{n_1\to\infty\\ n_2\to\infty}}{\longrightarrow} K(s,t) = K_1(s,t) + \rho K_2(s,t) =$$

$$= (1-t)(1-s)\left(\frac{1-(1-s)^{m_1}}{m_1^2\,(1-s)^{m_1}} + \frac{\rho k^2\left(1-(1-s)^{\frac{m_2}{k}}\right)}{m_2^2\,(1-s)^{\frac{m_2}{k}}}\right) =$$

$$= (1-t)\left(\frac{1-(1-s)^{m_1}}{m_1^2\,(1-s)^{m_1-1}} + \frac{\rho k^2\left(1-(1-s)^{\frac{m_2}{k}}\right)}{m_2^2\,(1-s)^{\frac{m_2}{k}-1}}\right).$$

Обозначим $Y_n(t) = \dfrac{m_1 m_2}{\sqrt{k^2\rho m_1^2 + m_2^2}} Z_n(t)$. При стандартных ограничениях $0 \le s \le t \le \Delta < 1$ процесс $Y_n(t)$ сходится к гауссовому процессу $Y(t)$ [4]. При этом

$$E[Y(t)] = 0,\ E[Y(s)\,Y(t)] = \\ = \frac{m_1^2 m_2^2}{k^2\rho m_1^2 + m_2^2}(1-t)\left(\frac{1-(1-s)^{m_1}}{m_1^2\,(1-s)^{m_1-1}} + \frac{\rho k^2\left(1-(1-s)^{\frac{m_2}{k}}\right)}{m_2^2\,(1-s)^{\frac{m_2}{k}-1}}\right) = \\ = (1-t)\frac{k_2\,(1-s)^{\frac{m_2}{k}-m_1} + k_1 - (1-s)^{\frac{m_2}{k}}}{(1-s)^{\frac{m_2}{k}-1}},\ 0 \le s \le t \le \Delta < 1.$$

Для определения приближенного асимптотического распределения статистики (2) предварительно докажем две леммы.

Лемма 1. *Пусть $Y(t)$ — процесс, определенный выше. Существует монотонно возрастающее преобразование $t = \psi(\tau)$ отрезка $[0,1]$ на себя $[0,1] \to [0,1]$, при котором процесс $W(\tau) = Y(\psi(\tau)) \times$ $\times \dfrac{(1-\psi(\tau))^{\frac{m_2}{k}-1}}{k_2 (1-\psi(\tau))^{\frac{m_2}{k}-m_1} + k_1}$ является броуновским мостом.*

◄ Рассмотрим преобразование

$$\tau(t) = \frac{k_2 (1-t)^{\frac{m_2}{k}-m_1} + k_1 - (1-t)^{\frac{m_2}{k}}}{k_2 (1-t)^{\frac{m_2}{k}-m_1} + k_1} : [0,1) \to [0,1).$$

При необходимости доопределим $\tau(1) = 1$. Нетрудно показать, что $\tau'(t) \geq 0$, $\tau(0) = 0$, $\tau(1) = 1$. Тогда существует обратное преобразование $t = \psi(\tau)$. Введем процесс

$$W(\tau) = Y(\psi(\tau)) \frac{(1-\psi(\tau))^{\frac{m_2}{k}-1}}{k_2 (1-\psi(\tau))^{\frac{m_2}{k}-m_1} + k_1} = Y(\psi(\tau)) \lambda(\psi(\tau)).$$

При $0 \leq u \leq v < 1$, $\psi(u) = s$, $\psi(v) = t$ имеем

$$E[W(\tau)] = 0, E[W(u) W(v)] =$$
$$= \lambda(\psi(u)) \lambda(\psi(v)) E[Y(\psi(u)) Y(\psi(v))] =$$
$$= \lambda(s) \lambda(t) E[Y(s) Y(t)] = \frac{(1-s)^{\frac{m_2}{k}-1}}{k_2 (1-s)^{\frac{m_2}{k}-m_1} + k_1} \frac{(1-t)^{\frac{m_2}{k}-1}}{k_2 (1-t)^{\frac{m_2}{k}-m_1} + k_1} \times$$
$$\times (1-t) \frac{k_2 (1-s)^{\frac{m_2}{k}-m_1} + k_1 - (1-s)^{\frac{m_2}{k}}}{(1-s)^{\frac{m_2}{k}-1}} =$$
$$= \frac{(1-t)^{\frac{m_2}{k}}}{k_2 (1-t)^{\frac{m_2}{k}-m_1} + k_1} \frac{k_2 (1-s)^{\frac{m_2}{k}-m_1} + k_1 - (1-s)^{\frac{m_2}{k}}}{k_2 (1-s)^{\frac{m_2}{k}-m_1} + k_1} = u(1-\nu). \blacktriangleright$$

Рассмотрим выборки $(\theta_1^1, \ldots, \theta_1^{n_1})$, $(\theta_2^1, \ldots, \theta_2^{n_2})$. Если полагать их независимыми выборками из совокупностей с функциями распределения $F^1(t) = 1 - (1-t)^{m_1}, F^2(t) = 1 - (1-t)^{\frac{m_2}{k}}$, то справедливо следующее утверждение.

Лемма 2. *Пусть, как было определено выше, $\overset{\frown}{F}{}^1 = \dfrac{d_1(t)}{n_1}$, $\overset{\frown}{F}{}^2 = \dfrac{d_2(t)}{n_2}$ — эмпирические функции распределения выборок $(\theta_1^1, \ldots, \theta_1^{n_1})$, $(\theta_2^1, \ldots, \theta_2^{n_2})$. Тогда*

$$P\left(\sup_t \left| (1- \overset{\frown}{F}{}^1(t))^{\frac{1}{m_1}} - (1-t) \right| \underset{n_1 \to \infty}{\longrightarrow} 0 \right) = 1,$$

$$P\left(\sup_t \left|(1-\overset{\frown}{F}{}^2(t))^{\frac{k}{m_2}}-(1-t)\right| \underset{n_1\to\infty}{\longrightarrow} 0\right)=1,$$

◀ Утверждение леммы очевидно в силу теоремы Гливенко и равномерной непрерывности на отрезке [0,1] степенных функций с положительным показателем степени. ▶

Асимптотическое распределение (2) приближенно может быть получено следующим образом. В функции $\lambda(\psi(\tau))=\lambda(t)$, определенной в лемме 1, множители $(1-t)$ должны быть заменены сходящимися к ним множителями $(1-\overset{\frown}{F}{}^1(t))^{\frac{1}{m_1}}$ или $(1-\overset{\frown}{F}{}^2(t))^{\frac{k}{m_2}}$. Тогда приближенно можно принять, что распределение статистики (2) совпадает с распределением максимума модуля броуновского моста, которое является классической функцией распределения Колмогорова [12]. Рассчитанные значения точных вероятностей, приведенные в таблице, показывают обоснованность этого утверждения. При больших объемах выборок разность значений точных и асимптотических [12] вероятностей составляет менее 0,004.

Оценка параметра модели Кокса. Полученный результат позволяет проверять гипотезы о возможных значениях параметра модели Кокса. Однако во многих случаях более важна оценка данного параметра. Обычно она осуществляется методом максимизации функции частного правдоподобия [10]. В настоящей работе предложена другая оценка, получаемая с помощью минимизации статистики типа Колмогорова – Смирнова $\overset{\frown}{k}=\arg\min T(\tilde{k})$.

Пусть значение параметра Кокса k в степенной гипотезе (1) неизвестно. В качестве оценки предложено значение $\overset{\frown}{k}$, которое минимизирует значение статистики Колмогорова – Смирнова

$$T(\tilde{k})=\frac{m_1m_2\sqrt{\rho n_2}}{\sqrt{\tilde{k}^2\rho m_1^2+m_2^2}}\times$$

$$\times\max_t\frac{\left(k_2\left(1-\overset{\frown}{F}{}^1\right)^{\frac{1}{m_1}}+k_1\left(1-\overset{\frown}{F}{}^2\right)^{\frac{\tilde{k}}{m_2}}\right)^{\frac{m_2}{\tilde{k}}-1}}{k_2\left(k_2\left(1-\overset{\frown}{F}{}^1\right)^{\frac{1}{m_1}}+k_1\left(1-\overset{\frown}{F}{}^2\right)^{\frac{\tilde{k}}{m_2}}\right)^{\frac{m_2}{\tilde{k}}-m_1}+k_1}\times$$

$$\times\left|\overset{\frown}{P}_{\theta_1}(t)-\left(\overset{\frown}{P}_{\theta_2}(t)\right)^{\tilde{k}}\right|,$$

т.е. $\overset{\frown}{k}=\arg\min T(\tilde{k})$, где $\tilde{k}$ — некоторое предполагаемое гипотетическое значение параметра модели Кокса. Исследование точности предложенной оценки коэффициента ускорения проведено методами статистического моделирования.

Алгоритм моделирования. 1. Моделируются $n_1 m_1$ одинаково распределенных случайных величин $(\xi_1^1, \ldots, \xi_{m_1 n_1}^1)$ с функцией распределения $F_0(t)$. В качестве функции $F_0(t)$ рассматривались следующие функции распределения: экспоненциальная (с параметром $\beta = 0{,}001$) и Вейбулла (с параметрами $\beta = 0{,}001$, $p = 1{,}5$). Далее величины $(\xi_1^1, \ldots, \xi_{m_1 n_1}^1)$ также будем называть наработками.

2. Наработки случайным образом разбиваются на n_1 групп по m_1 величин в каждой. Элементы i-й группы обозначаются как $\left(\xi_1^{1,i}, \ldots, \xi_{m_1}^{1,i}\right)$, $i = \overline{1, n_1}$. Определяются $\theta_1^i = \min\left(\xi_1^{1,i}, \ldots, \xi_{m_1}^{1,i}\right)$.

3. Аналогично моделируются $n_2 m_2$ одинаково распределенных случайных величин $(\xi_1^2, \ldots, \xi_{m_2 n_2}^2)$ с функцией распределения $1 - \left(1 - (F_0(t))^k\right)^{\frac{1}{k}}$, где k — некоторое заданное значение параметра Кокса. Наработки $(\xi_1^2, \ldots, \xi_{m_2 n_2}^2)$ случайным образом разбиваются на n_2 групп по m_2 величин в каждой. Элементы i-й группы обозначаются как $\left(\xi_1^{2,i}, \ldots, \xi_{m_2}^{2,i}\right)$, $i = \overline{1, n_2}$. Определяются $\theta_2^i = \min\left(\xi_1^{2,i}, \ldots, \xi_{m_2}^{2,i}\right)$.

4. Для некоторого значения $\tilde{k}$, $1 \le \tilde{k} \le K$, по двум получаемым выборкам $\Theta_1 = (\theta_1^1, \ldots, \theta_1^{n_1})$, $\Theta_2 = (\theta_2^1, \ldots, \theta_2^{n_2})$ вычисляется значение статистики Колмогорова – Смирнова $T(\tilde{k})$.

5. Методом перебора $\tilde{k}$ определяется оценка $\widehat{k}$, минимизирующая значение статистики $T(\tilde{k})$: $\widehat{k} = \arg\min T(\tilde{k})$.

Для определения статистических свойств оценки $\widehat{k}$ пункты 1–5 повторялись 500 раз. По полученным значениям оценок строилась гистограмма этой выборки, а также вычислялись эмпирические среднее $\widehat{k}$ и дисперсия S^2.

Гистограммы полученных оценок $\widehat{k}$ при $m_1 = 2$, $m_2 = 3$,$n_1 = n_2 = 100$ для экспоненциального распределения с параметром $\beta = 0{,}001$ и для распределения Вейбулла с параметрами $\beta = 0{,}001$, $p = 1{,}5$ приведены на рис. 2.

По результатам моделирования для указанных параметров были вычислены математическое ожидание полученной оценки и ее среднеквадратическое отклонение. В случае экспоненциального распределения $M\widehat{k} = 2{,}05$, $\sigma = 0{,}36$, в случае распределения Вейбулла $M\widehat{k} = 2{,}035$, $\sigma = 0{,}37$. Эти результаты свидетельствуют о хороших статистических свойствах предложенной оценки.

Заключение. В работе предложен метод проверки гипотезы о степенной зависимости функции надежности для двух прогрессивно цензурированных выборок. Для проверки данной гипотезы получена статистика типа Колмогорова – Смирнова. Показана сходимость точных

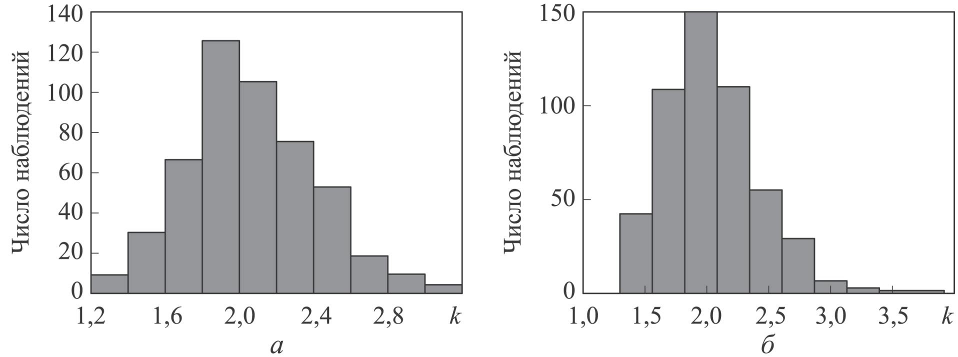

Рис. 2. Гистограммы оценок $\widehat{k}$ для экспоненциального распределения (*а*) и распределения Вейбулла (*б*) при $k = 2$

распределений данной статистики к ее асимптотическим распределениям. Методом Монте-Карло показано, что оценка параметра модели Кокса, получаемая минимизацией предлагаемой статистики, является состоятельной оценкой.

ЛИТЕРАТУРА

1. *Cox D.R.* Regression Models and Life-Tables // Journal of the Royal Statistical Society. Series B (Methodological). 1972. Vol. 34. No. 2. P. 187–220.
2. *Тимонин В.И.*, *Тянникова Н.Д.* Проверка однородности двух цензурированных выборок из наработок изделий, основанная на сравнении оценок Каплана – Мейера их функций надежности // Физические основы приборостроения. 2015. Т. 4. № 1. С. 30–41.
3. *Bagdonavichus V.*, *Kruopis J.*, *Nikulin M.S.* Nonparametric tests for censored data. London: ISTE Ltd, 2011. 233 p.
4. *Balakrishnan N.*, *Cramer E.* The Art of Progressive Censoring. Applications to Reliability and Quality. N.Y.: Springer, 2014. 645 p.
5. *Kaplan E.L.*, *Meier P.* Nonparametric estimation from incomplete observations // J. am. Stat. Assoc. 1958. No. 53. P. 57–81.
6. *Тимонин В.И.*, *Ермолаева М.А.* Оценки Каплана – Мейера в статистиках типа Колмогорова – Смирнова при проверке гипотез в испытаниях с переменной нагрузкой // Электромагнитные волны и электронные системы. 2010. Т. 15. № 7. С. 18–26.
7. *Balakrishnan N.*, *Tripathi R.C.*, *Kannan N.* Some nonparametric precedence type tests based on progressively censored samples and evaluation of power // J. Stat. Plan. Infer. 2010. No. 140. P. 559–573.
8. *Maturi T.A.*, *Coolen-Schrijner P.*, *Coolen F.P.* Nonparametric predictive comparison of lifetime data under progressive censoring // J. Stat. Plan. Infer. 2010. No. 140. P. 515–525.
9. *Тянникова Н.Д.*, *Тимонин В.И.* Метод вычисления точных распределений статистик типа Колмогорова – Смирнова в случае нарушения однородности и независимости анализируемых выборок // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. 2014. № 11. С. 217–227. URL: http://technomag.bmstu.ru/doc/740251.html

10. *Кокс Д.*, *Оукс Д.* Анализ данных типа времени жизни. М.: Финансы и статистика, 1988. 191 с.
11. *Тимонин В.И.* О предельном распределении статистики одного непараметрического критерия // Теория вероятностей и ее применение. 1987. Т. 32. № 4. С. 790–792.
12. *Hajek J.*, *Sidak Z.* Theory of rank tests. London: Academic Press, 2004. 438 p.

REFERENCES

[1] Cox D.R. Regression Models and Life-Tables. *J. of the Royal Statistical Society. Series B* (*Methodological*), 1972, vol. 34, no. 2, pp. 187–220.

[2] Timonin V.I., Tyannikova N.D. The homogeneity testing of two censored samples of times to failure based on a comparison of the Kaplan – Meier estimates of reliability functions. *Fizicheskie osnovy priborostroeniya* [Physical Bases of Instrumentation], 2015, vol. 4, no. 1, pp. 30–41 (in Russ.).

[3] Bagdonavichus V., Kruopis J., Nikulin M.S. Nonparametric tests for censored data. London, ISTE Ltd, 2011. 233 p.

[4] Balakrishnan N., Cramer E. The Art of Progressive Censoring. Applications to Reliability and Quality. N.Y., Springer, 2014. 645 p.

[5] Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations. *J. am. Stat. Assoc.*, 1958, no. 53, pp. 57–81.

[6] Timonin V.I., Ermolaeva M.A. About Kaplan – Meyer Estimators in Statistics Similar to Kolmogorov – Smirnov for Testing the Hypothesis in Variable Load Tests Elektromagnitnye volny i elektronnye sistemy [Electromagnetic Waves and Electronic Systems], 2010, vol. 15, no. 7, pp. 18–26 (in Russ.).

[7] Balakrishnan N., Tripathi R.C., Kannan N. Some nonparametric precedence type tests based on progressively censored samples and evaluation of power. *J. Stat. Plan. Infer.*, 2010, no. 140, pp. 559–573.

[8] Maturi T.A., Coolen-Schrijner P., Coolen F.P. Nonparametric predictive comparison of lifetime data under progressive censoring. *J. Stat. Plan. Infer.*, 2010, no. 140, pp. 515–525.

[9] Timonin V.I., Tyannikova N.D. The Method of Calculating the Exact Distributions of the Kolmogorov–Smirnov Statistics in Case of Violation of Homogeneity and Independence of the Analyzed Samples. "*Jelektr. Nauchno-Tehn. Izd "Nauka i obrazovanie*" [El. Sc.-Tech. Publ. "Science and Education"], 2012, no. 12. Available at: http://technomag.bmstu.ru/doc/740251.html (accessed 14.11.2014) DOI: 10.7463/1114.0740251

[10] Cox D.R., Oakes D. Analysis of Survival Data. London–New York, Chapman & Hall, 1984.

[11] Timonin V.I. On Limiting Behaviour of Some Nonparametric Test. Teoriya veroyatnostey i ee primenenie [Theory of Probability and its Applications], 1987, vol. 32, no. 4, pp. 790–792 (in Russ.).

[12] Hajek J., Sidak Z. Theory of rank tests. London: Academic Press, 2004. 438 p.

Статья поступила в редакцию 03.06.2015

Тимонин Владимир Иванович — д-р физ.-мат. наук, профессор кафедры "Высшая математика" МГТУ им. Н.Э. Баумана.

МГТУ им. Н.Э. Баумана, Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5.

Timonin V.I. — D.Sc. (Phys.-Math.), Professor, Department of Higher Mathematics, Bauman Moscow State Technical University.

Bauman Moscow State Technical University, 2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation.

Тянникова Нина Дмитриевна — аспирантка кафедры “Высшая математика” МГТУ им. Н.Э. Баумана.
МГТУ им. Н.Э. Баумана, Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5.

Tyannikova N.D. — Ph.D. student, Department of Higher Mathematics, Bauman Moscow State Technical University.
Bauman Moscow State Technical University, 2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation.

Просьба ссылаться на эту статью следующим образом:

Тимонин В.И., Тянникова Н.Д. Применение оценок Каплана – Мейера для проверки степенной гипотезы Кокса по двум прогрессивно цензурированным выборкам // Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки. 2015. № 6. С. 68–84.

Please cite this article in English as:

Timonin V.I., Tyannikova N.D. Application of Kaplan – Meier estimates to testing Cox power hypothesis for two progressive lycensored samples. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Estestv. Nauki* [Herald of the Bauman Moscow State Tech. Univ., Nat. Sci.], 2015, no. 6, pp. 68–84.