

Математические методы исследования

Mathematical methods of investigation

DOI: <https://doi.org/10.26896/1028-6861-2024-90-7-84-92>

ИНФОРМАЦИЯ ФИШЕРА, СОДЕРЖАЩАЯСЯ В НЕПОЛНЫХ НАБЛЮДЕНИЯХ

© **Абдурахим Ахмедович Абдушукуров^{1*}, Наргиза Сайдиллаевна Нурмухамедова², Сабитбек Эрисбаев³**

¹ Филиал Московского государственного университета имени М. В. Ломоносова в городе Ташкенте, Узбекистан, 100060, Ташкент, пр. Амира Тимура, д. 22; *e-mail: a_abdushukurov@rambler.ru

² Национальный университет Узбекистана имени Мирзо Улугбека, Узбекистан, 100174, Ташкент, Университетская ул., д. 4; e-mail: rasulova_nargiza@mail.ru

³ Каракалпакский государственный университет имени Бердаха, Узбекистан, 742012, Нукус, ул. Абдилова, д. 1; e-mail: sabitbek@mail.ru

*Статья поступила 2 октября 2023 г. Поступила после доработки 30 октября 2023 г.
Принята к публикации 29 ноября 2023 г.*

Точность статистических оценок для неизвестных параметров распределений зависит не только от объема выборочных данных, но и от способа их получения. Одним из основных требований к сбору экспериментальных данных является их информативность. Задачи математической статистики, особенно параметрического оценивания по цензурированным наблюдениям, обладают специфическими особенностями. Типичными представителями моделей неполных наблюдений на прямой являются модели случайного цензурирования, конкурирующих рисков (однократного, многократного) случайного цензурирования. В работе показано, что не всегда цензурирование приводит к потере информации (Фишера). Если цензурирование информативное, т.е. распределение цензурирующих случайных величин зависит от того же параметра, то можно указать такую модель, где информация из-за цензурирования может и сохраняться. Наоборот, если цензурирование не информативное, то потеря информации неизбежна. Работа носит обзорный характер, основное внимание уделено результатам авторов работы, приведены также данные из современных литературных источников. За критерий качества оценки принята эффективность по Крамеру – Рао, а за критерий информации о неизвестном параметре — информация Фишера.

Ключевые слова: информация Фишера; семейство распределений; интенсивность отказов; цензурирование; конкурирующие риски; неполные наблюдения.

FISHER INFORMATION CONTAINED IN INCOMPLETE OBSERVATIONS

© **Abdurakhim A. Abdushukurov^{1*}, Nargiza S. Nurmukhamedova²,
Sabitbek A. Erisbaev³**

¹ Tashkent Branch of the M. V. Lomonosov Moscow State University, 22, Amir Timur prosp., Tashkent, 100060, Uzbekistan; *e-mail: a_abdushukurov@rambler.ru

² Mirzo Ulugbek National University of Uzbekistan, 4, Universitetskaya ul., Tashkent, 100174, Uzbekistan; e-mail: rasulova_nargiza@mail.ru

³ Berdakh Karakalpak State University, 1, Abdirova ul., Nukus, 742012, Uzbekistan; e-mail: sabitbek@mail.ru

Received October 2, 2023. Revised October 30, 2023. Accepted November 29, 2023.

The accuracy of statistical estimates of unknown distribution parameters depends not only on the bulk of sampling data but also on the method of data acquisition. The information content of experimental data is one of the basic requirements. Problems of mathematical statistics, in particular parametric estimation based on censored observations, have specific features. Typical representatives of models of incomplete observations on a straight line are models of random censoring, competing risks of (single, multiple) random censoring. The purpose of this study is to show that censoring does not always lead to loss of (Fisher) information. It is shown that if censoring is informative, i.e., the distribution of censoring random variables depends on the same parameter, it is possible to specify a model where information can be preserved due to censoring. On the contrary, if the censoring is not informative, then the loss of information is inevitable.

The Cramer – Rao efficiency was taken as a criterion for the quality of the assessment, whereas the Fisher information was taken as the criterion for information about the unknown parameter.

Keywords: Fisher information; distribution family; hazard rate; censorship; competing risks; incomplete observation.

Введение

В математической статистике впервые в 1925 году Р. А. Фишером [1] было введено понятие информации для сравнения свойств статистических оценок параметров. Пусть статистическая модель состоит из тройки $(\mathbf{X}, \mathbf{B}, \mathbf{F}_\theta)$, где $\mathbf{X}$ — выборочное пространство; $\mathbf{B}$ — борелевская σ -алгебра на $\mathbf{X}$, $\{\mathbf{F}_\theta, \theta \in \Theta\}$ — параметрическое семейство функций распределений случайной величины $\mathbf{X}$ (которая может быть и вектором наблюдений), Θ — подмножество вещественной прямой (для простоты изложения). Пусть $f(x, \theta)$ — плотность распределений из семейства $\{\mathbf{F}_\theta\}$. Рассмотрим условия регулярности для семейства $\{\mathbf{F}_\theta\}$:

I) Θ — или вся вещественная прямая, или интервал на вещественной прямой;

II) носитель плотности $N_f = \{x: f(x, \theta) > 0\}$ не зависит от θ ;

III) производная $\partial f(x, \theta)/\partial \theta$ существует и почти всюду конечна для любого $\theta \in \Theta$;

IV) $\int \left| \frac{\partial^i f(x, \theta)}{\partial \theta^i} \right| dx < \infty$, $i = 1, 2$ для любого $\theta \in \Theta$;

V) $M_\theta \left(\frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2 < \infty$ для любого $\theta \in \Theta$.

Условия I – V называются условиями регулярности, а функция

$$I(\theta) = M_\theta \left(\frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2 —$$

информацией Фишера о параметре θ , содержащемся в семействе $\{\mathbf{F}_\theta\}$. В этих условиях, если параметр θ многомерный ($\theta = (\theta_1, \dots, \theta_s)$, $s > 1$), то

$$\frac{\partial \ln f(x, \theta)}{\partial \theta} = \left(\frac{\partial \ln f(x, \theta_1)}{\partial \theta_1}, \dots, \frac{\partial \ln f(x, \theta_s)}{\partial \theta_s} \right)$$

и соответствующая информация $I(\theta)$ будет матрицей. Функция $S(X, \theta) = \partial \ln f(X, \theta)/\partial \theta$ называется вкладом выборки X и является достаточной статистикой семейства $\{\mathbf{F}_\theta\}$. Пусть при $s = 1$ для функции параметра $g(\theta)$, оцениваемой по выборке X , существует несмещенная оценка $\hat{g}(X)$ и выполняется условие VI:

$$\int \left| \hat{g}(x) \frac{\partial f(x, \theta)}{\partial \theta} \right| dx < \infty \text{ для всех } \theta \in \Theta.$$

Основная теорема теории статистического оценивания утверждает, что в условиях регулярности I – VI для дисперсии оценки справедливо неравенство

$$D_\theta \hat{g}(X) \geq \frac{(g'(\theta))^2}{I(\theta)} \quad (1)$$

для всех $\theta \in \Theta$. Равенство в (1) имеет место тогда и только тогда, когда функция плотности распределения оценки $\hat{g}(X)$ является плотностью экспоненциального вида, т.е.

$$f^g(g, \theta) = \exp(\hat{g}(x)\psi_1(\theta) + \psi_2(\theta) + h(g)), \quad (2)$$

где $\psi_1(\theta) \neq 0$.

Таким образом, если в (1) имеет место равенство, то функция информации Фишера обратно пропорциональна дисперсии всех несмещенных оценок $g(\theta)$, т.е. информационное количество $I(\theta)$ пропорционально точности несмещенной оценки $g(\theta)$. В литературе неравенство (1) называется неравенством Крамера – Рао по имени ее авторов (см. обзор работ в монографии Ш. Закса [2]).

Следует отметить, что авторы [3] получали другое представление информации $I(\theta)$ через плотность интенсивности отказов $\mu(x, \theta)/(1 - F(x, \theta))$:

$$I(\theta) = \int \left(\frac{\partial \ln \mu(X, \theta)}{\partial \theta} \right)^2 dF(x, \theta), \theta \in \Theta, \quad (3)$$

где $F(x, \theta)$ — функция распределения, соответствующая плотности $f(x, \theta)$. Они использовали метод сохранения расстояния в пространстве L_2 . Формула (3) более полезна в теории испытания на долговечность. Отметим, что эта формула присутствует и в моделях случайного цензурирования и конкурирующих рисков (см. ниже).

Исследования многих авторов были посвящены задачам эффективности статистических оценок по неполным наблюдениям. Среди них следует упомянуть более поздние работы Пракаса Рао (B. L. S. Prakasa Rao [4 – 7]), в которых улучшены классические неравенства Крамера – Рао, рассмотрены более разнообразные задачи байесовского подхода, последовательного анализа и другие (см. также [8]).

Неравенство типа Крамера – Рао в случае случайно цензурированных наблюдений

Пусть на некотором вероятностном пространстве рассматриваются независимые случайные величины ξ и η с функциями распределения $F(x; \theta)$ и $G(x; \theta)$ соответственно. Через $\xi_1, \dots, \xi_n$ и $\eta_1, \dots, \eta_n$ обозначим независимые реализации случайных величин ξ и η . Пусть $I(A)$ — индикатор события A . Наблюдается выборка $\mathbf{C}^{(n)} = \{(\xi_i, \delta_i), i = 1, \dots, n\}$, где $\xi_i = \min(\xi_i, \eta_i)$ и $\delta_i = I(\xi_i \leq \eta_i)$. Заметим, что число наблюдаемых ξ_i составит $v_n = \delta_1 + \dots + \delta_n$, а $\eta_i = 1 - v_n$. Пусть функции распределения $F(x, \theta)$ и $G(x, \theta)$ абсолютно непрерывны и существуют плотности $f(x, \theta)$ и $g(x, \theta)$. Обозначим $1 - F(x, \theta) = \overline{F}(x, \theta)$ и $1 - G(x, \theta) = \overline{G}(x, \theta)$. Тогда пара (ζ, δ) имеет совместную плотность

$$h(x, y, \theta) = [f(x, \theta)\overline{G}(x, \theta)]^y [g(x, \theta)\overline{F}(x, \theta)]^{1-y} \quad (4)$$

относительно меры $\mu \times \nu$ на $R \times \{0, 1\}$, где μ — лебегова мера на прямой и ν — считающая мера с массами $\{0\}$ и $\{1\}$. Задача состоит в оценивании

$$I(\theta) = I_{(\zeta, \eta)}(\theta) = \int_R \left(\frac{\partial \ln(g(x, \theta)\overline{F}(x, \theta))}{\partial \theta} \right)^2 g(x, \theta)\overline{F}(x, \theta) dx + \int_R \left(\frac{\partial \ln(f(x, \theta)\overline{G}(x, \theta))}{\partial \theta} \right)^2 f(x, \theta)\overline{G}(x, \theta) dx = I_{(\zeta, \eta)}^{(0)}(\theta) + I_{(\zeta, \eta)}^{(1)}(\theta);$$

5) пусть $\hat{\varphi}(\zeta, \eta)$ — несмещенная оценка дифференцируемого параметра $\varphi(\theta)$ и

$$\int_{R \times \{0, 1\}} \left| \hat{\varphi}(x, y) \frac{\partial h(x, y, \theta)}{\partial \theta} \right| dx d\nu(y) < \infty$$

для всех $\theta \in \Theta$.

Теорема 1 [9]. Пусть выполнены условия 1 – 5. Тогда имеет место неравенство

$$D_{\theta} \hat{\varphi}(\zeta, \eta) \geq \frac{(\varphi'(\theta))^2}{I(\theta)} \quad (5)$$

для всех $\theta \in \Theta$. Здесь необходимо рассматривать два случая.

Случай А (неинформативный). Цензурирующее распределение G не зависит от параметра θ . В этом случае информация упрощается и ее обозначим как $I_G(\theta)$. Имеются численные примеры, подтверждающие выполнения неравенства $I_G(\theta) \leq I(\theta)$ (см. далее).

Случай В (информативный). Распределение G функционально выражается через F . Тогда существует монотонная и неотрицательная функ-

F при мешающем распределении G , поэтому нас интересуют только наблюдения ξ_i .

Отметим, что вопросы эффективного оценивания неизвестного параметра при случайном цензурировании справа впервые рассмотрены Абдушукуровым – Кимом [9], в дальнейшем их исследования продолжили другие ученые [10 – 14]. Авторы [9] рассматривали следующие условия регулярности для семейств $f(x, \theta)$ и $g(x, \theta)$:

1) множество

$$\{x: f(x; \theta)\overline{G}(x; \theta) = 0\} \cup \{x: g(x; \theta)\overline{F}(x; \theta) = 0\}$$

не зависит от параметра θ ;

2) плотности f и g дифференцируемы по θ и для всех $\theta \in \Theta$ —

$$\int_R \left| \frac{\partial f(x, \theta)}{\partial \theta} \right| dx < \infty, \quad \int_R \left| \frac{\partial g(x, \theta)}{\partial \theta} \right| dx < \infty;$$

$$3) M_{\theta} \left(\frac{\partial \ln f(\xi, \theta)}{\partial \theta} \right)^2 < \infty, \quad M_{\theta} \left(\frac{\partial \ln g(\eta, \theta)}{\partial \theta} \right)^2 < \infty$$

для всех $\theta \in \Theta$;

4) $0 < I(\theta) < \infty$ для всех $\theta \in \Theta$, где информация Фишера модели определяется формулой

ция $q(t)$, $t \in [0, 1]$, такая, что $G(x, \theta) = q(F(x, \theta))$ или $\overline{G}(x, \theta) = q(\overline{F}(x, \theta))$, где $q(0) = q(1) = 0$ и $x \in R$. Например, если пара (F, G) отвечает модели пропорциональных интенсивностей, то существует положительное, не зависящее от θ число β такое, что

$$\overline{G}(x, \theta) = \overline{F}(x, \theta)^{\beta}, \quad x \in R. \quad (6)$$

Из (6) следует, что

$$\frac{\partial \ln \overline{G}(x, \theta)}{\partial x} = \beta \frac{\partial \ln \overline{F}(x, \theta)}{\partial x}$$

или

$$\frac{g(x, \theta)}{G(x, \theta)} = \beta \frac{f(x, \theta)}{F(x, \theta)},$$

что оправдывает название этой модели. В данном случае

$$I(\theta) = (1 + \beta) \int_R \left(\frac{\partial \ln(f(x, \theta)\overline{G}(x, \theta))}{\partial \theta} \right)^2 f(x, \theta)\overline{G}(x, \theta) dx,$$

$$p = \frac{1}{1 + \beta} = P(\xi \leq \eta).$$

Кроме того, если $\overline{F(x, \theta)} = e^{-\theta x}$, $\overline{G(x, \theta)} = e^{-\beta \theta x}$, $\beta > 0$, $x \geq 0$, то легко увидеть, что в этом случае информация не теряется, т.е. $I(\theta) = \theta^{-2}$.

Пусть $\mu(x, \theta) = f(x, \theta) / \overline{F(x, \theta)}$, $\lambda(x, \theta) = g(x, \theta) / \overline{G(x, \theta)}$ и $\gamma(x, \theta) = \mu(x, \theta) + \lambda(x, \theta)$ — функции плотности интенсивности, соответствующие распределениям F , G и $H = 1 - \overline{FG}$. В [15] установлено следующее утверждение.

Лемма 1 [15]. В условиях регулярности 1 – 5 для всех $\theta \in \Theta$

$$I(\theta) = I_{\mu}^{(1)}(\theta) + I_{\lambda}^{(0)}(\theta), \quad (7)$$

где

$$I_{\mu}^{(1)}(\theta) = \int_R \left(\frac{\partial \ln \mu(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) \overline{G(x, \theta)} dx,$$

$$I_{\lambda}^{(0)}(\theta) = \int_R \left(\frac{\partial \ln \lambda(x, \theta)}{\partial \theta} \right)^2 g(x, \theta) \overline{F(x, \theta)} dx.$$

Формула (7) является обобщением (3) для случайного цензурирования справа.

Рассмотрим более общую модель, являющуюся естественным обобщением предыдущего случая цензурирования — модель конкурирующих рисков (МКР).

Информация Фишера и неравенство типа Крамера – Рао в МКР

Эту модель определим, следуя работам [15 – 19]. Для фиксированного натурального числа k пусть $\{X^{(1)}, X^{(2)}, \dots, X^{(k)}\}$ — совокупность независимых случайных величин с соответствующими функциями распределения $\{F^{(1)}, F^{(2)}, \dots, F^{(k)}\}$, зависящими от общего параметра $\theta \in \Theta \subseteq R$. При этом $F^{(i)}$ абсолютно непрерывны и имеют плотности распределения и интенсивности отказов

$$f^{(i)}(t, \theta) = \frac{\partial F^{(i)}(t, \theta)}{\partial t} \text{ и } \lambda^{(i)}(t, \theta) = \frac{f^{(i)}(t, \theta)}{F^{(i)}(t, \theta)},$$

$$i = 1, \dots, k.$$

В МКР интерес представляют векторы $(X, \delta^{(1)}, \dots, \delta^{(k)})$, где $X = \min(X^{(i)}, 1 \leq i \leq k)$ и индикаторы $\delta^{(i)} = I(X = X^{(i)})$, где $\delta^{(1)} + \dots + \delta^{(k)} = 1$. Пусть $H(t, \theta) = P_{\theta}(X \leq t)$, субраспределения $H^{(i)}(t, \theta) = P_{\theta}(X \leq t, \delta^{(i)} = 1)$, $i = 1, \dots, k$. Теперь видно, что

$$H(t, \theta) = H^{(1)}(t, \theta) + \dots + H^{(k)}(t, \theta) = 1 - \prod_{i=1}^k \overline{F^{(i)}(t, \theta)}$$

и аналогично плотности

$$h(t, \theta) = \frac{\partial H(t, \theta)}{\partial t} = h^{(1)}(t, \theta) + \dots + h^{(k)}(t, \theta),$$

где

$$h^{(i)}(t, \theta) = f^{(i)}(t, \theta) \prod_{\substack{l=1 \\ l \neq i}}^k \overline{F^{(l)}(t, \theta)}.$$

Отсюда

$$H^{(i)}(t, \theta) = \int_{-\infty}^t \prod_{\substack{l=1 \\ l \neq i}}^k \overline{F^{(l)}(u, \theta)} dF^{(i)}(u, \theta),$$

$i = 1, \dots, k$. Определим интегральные интенсивности

$$\Lambda(t, \theta) = \int_{-\infty}^t \frac{dH(u, \theta)}{H(u, \theta)} = \sum_{i=1}^k \Lambda^{(i)}(t, \theta) = \int_{-\infty}^t \lambda(u, \theta) du,$$

где

$$\Lambda^{(i)}(t, \theta) = \int_{-\infty}^t \frac{dH^{(i)}(u, \theta)}{H(u, \theta)} = \int_{-\infty}^t \frac{dF^{(i)}(u, \theta)}{F^{(i)}(u, \theta)} = \int_{-\infty}^t \lambda^{(i)}(u, \theta) du.$$

Определим функцию регрессии индикатора $\delta^{(i)}$ на случайную величину X равенством

$$m^{(i)}(t, \theta) = P_{\theta}(\delta^{(i)} = 1/X = t) = M_{\theta}(\delta^{(i)}/X = t),$$

$$i = 1, \dots, k.$$

Тогда справедливы представления $H^{(i)}(t, \theta) = \int_{-\infty}^t m^{(i)}(u, \theta) d\Lambda(u, \theta)$, а следовательно,

$$\Lambda^{(i)}(t, \theta) = \int_{-\infty}^t m^{(i)}(u, \theta) d\Lambda(u, \theta), \quad i = 1, \dots, k,$$

где $m^{(1)}(t, \theta) + \dots + m^{(k)}(t, \theta) = 1$ для всех $(t, \theta) \in R \times \Theta$. Тогда очевидно

$$\lambda^{(i)}(t, \theta) = m^{(i)}(t, \theta) \lambda(t, \theta), \quad i = 1, \dots, k.$$

Через $I_{(X, \delta^{(1)}, \dots, \delta^{(k)})}(\theta)$, $I_X(\theta)$ и $I_{(X, \delta^{(i)})}(\theta)$ обозначим информации Фишера, содержащиеся в случайных величинах $(X, \delta^{(1)}, \dots, \delta^{(k)})$, $X = \min(X^{(1)}, \dots, X^{(k)})$, и $(X, \delta^{(i)})$, $i = 1, \dots, k$, соответственно. Введем в рассмотрение условия регулярности:

1) множества $N^{(i)} = \{t: h^{(i)}(t, \theta) > 0\}$, $i = 1, \dots, k$

не зависят от параметра θ и $\prod_{i=1}^k N^{(i)} \in \emptyset$;

2) существуют конечные производные

$$\frac{\partial^m h^{(i)}(t, \theta)}{\partial \theta^m}, \quad m = 1, 2; \quad i = 1, \dots, k,$$

для всех $(t, \theta) \in R \times \Theta$;

3) существуют интегралы

$$\int_R \left| \frac{\partial^m h^{(i)}(t, \theta)}{\partial \theta^m} \right| dt, \quad m = 1, 2; \quad i = 1, \dots, k,$$

для всех $\theta \in \Theta$;

4) пусть $I_{(X, \delta^{(i)})}^{(i)}(\theta) \in (0; +\infty)$, $i = 1, \dots, k$, $\theta \in \Theta$.
Согласно (3)

$$I_X(\theta) = \int_R \left(\frac{\partial \ln \lambda(t, \theta)}{\partial \theta} \right)^2 dH(t, \theta).$$

Теорема 2 [15]. Пусть выполнены условия регулярности 1 – 4. Тогда для всех $\theta \in \Theta$:

$$\begin{aligned} \text{А)} \quad I_{(X, \delta^{(1)}, \dots, \delta^{(k)})}(\theta) &= \sum_{i=1}^k I_{(X, \delta^{(i)})}^{(i)}(\theta); \\ \text{В)} \quad I_{(X, \delta^{(1)}, \dots, \delta^{(k)})}(\theta) &= I_X(\theta) + \sum_{i=1}^k I_{(\delta^{(i)}/X)}^{(i)}(\theta), \end{aligned} \quad (8)$$

где

$$I_{(\delta^{(i)}/X)}^{(i)}(\theta) = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln m^{(i)}(t, \theta)}{\partial \theta} \right)^2 m^{(i)}(t, \theta) dH(t, \theta),$$

$$i = 1, \dots, k.$$

Теорема 3 [16]. В условиях теоремы 2 пусть $\hat{\varphi}(X, \delta^{(1)}, \dots, \delta^{(k)})$ — несмещенная оценка для $\varphi(\theta)$, $D_{\theta} \hat{\varphi} \in (0, +\infty)$ и возможно дифференцирование под знаком интеграла равенства $M_{\theta} \hat{\varphi} = \varphi(\theta)$. Тогда справедливо неравенство

$$D_{\theta} \hat{\varphi} \geq \frac{(\varphi'(\theta))^2}{I_{(X, \delta^{(1)}, \dots, \delta^{(k)})}(\theta)} \quad (9)$$

для всех $\theta \in \Theta$.

Пусть $F^{(i)}(t, \theta) = \exp(-\beta_i T(t)g(\theta))$, $0 < \beta_i < \infty$, $i = 1, k$, где β_i не зависит от θ . Тогда $\hat{\varphi}(X) = (\beta_1 + \dots + \beta_k)T(X)$ является несмещенной оценкой $\varphi(\theta) = 1/g(\theta)$, где

$$\begin{aligned} f^{(i)}(t, \theta) &= \prod_{\substack{l=1 \\ l \neq i}}^k (1 - F^{(l)}(t, \theta)) = \\ &= \exp(-\hat{\varphi}(t) \alpha_1(\theta) + \alpha_0(\theta) + v^{(i)}(t)), \quad i = 1, \bar{k}, \\ v^{(i)}(t, y^{(1)}, \dots, y^{(k)}) &= \sum_{i=1}^k y^{(i)} \ln \left(\beta^{(i)} \frac{dT(t)}{dt} \right) \end{aligned}$$

и в неравенстве (9) достигается равенство. Теперь улучшим неравенство (9).

Теорема 4 [20]. Пусть X и Y две случайные величины, заданные на одном вероятностном пространстве. Тогда при существовании у них вторых моментов справедливо неравенство

$$\begin{aligned} [M(XY)]^2 &\leq [MX^2][MY^2] - \\ &- [|MX| \sqrt{DY} - |MY| \sqrt{DX}]^2. \end{aligned} \quad (10)$$

В частности, если $MX = 0$ и $MY^2 > 0$, то согласно (10)

$$MX^2 \geq [MX]^2 + [MXY]^2 / MY^2. \quad (11)$$

Теорема 5. Пусть выполнены условия теоремы 3. Тогда справедливо неравенство

$$\begin{aligned} M_{\theta} [\hat{\varphi}(X, \delta^{(1)}, \dots, \delta^{(k)}) - \varphi(\theta)]^2 &\geq \\ &\geq \left\{ M_{\theta} [\hat{\varphi}(X, \delta^{(1)}, \dots, \delta^{(k)}) - \varphi(\theta)] \right\}^2 + \\ &+ \frac{(\varphi'(\theta))^2}{I_{(X, \delta^{(1)}, \dots, \delta^{(k)})}(\theta)} \end{aligned} \quad (12)$$

для всех $\theta \in \Theta$.

Информация Фишера в порядковых статистиках

Фишеровская информация, имеющаяся в членах вариационного ряда, хорошо изучена многими авторами. Рассмотрены модели по цензурированию I и II рода, их гибриды, рекорды и т.п. Результаты работы в этом направлении отражены в [21, 22]. Ниже приведем данные численных вычислений информации при неинформативном случайном цензурировании справа, а также в модели многократного цензурирования справа [23]. Различные применения фишеровской информации в задачах оценивания представлены в [24].

Информация Фишера при неинформативном цензурировании справа

Пусть ξ — интересующая нас случайная величина, заданная на вероятностном пространстве (Ω, A, P) , со значениями в $X \subseteq R^1 = (-\infty, \infty)$. Через $F(x)$, $x \in R^1$, обозначим функцию распределения случайной величины ξ . Рассмотрим параметрический случай, когда функция распределения F задается с точностью до параметра θ : $F(x, \theta) = P_{\theta}(\xi \leq x)$, $P = \{P_{\theta}, \theta \in \Theta\}$, где $\theta = (\theta_1, \theta_2, \dots, \theta_s)^T \in \Theta$, Θ — открытый интервал в R^s . В случае, когда по выборке $\xi^{(n)} = (\xi_1, \xi_2, \dots, \xi_n)$, являющейся результатом независимых наблюдений над случайной величиной ξ , требуется вычислить информацию Фишера, используют формулу:

$$I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 dF(x, \theta).$$

Однако когда выборка $\xi^{(n)}$ случайно цензурирована справа, при вычислении информации Фишера появляются определенные сложности, так как для некоторых распределений данная информация явно не вычисляется и приходится пользоваться численными методами.

Рассмотрим неинформативное случайное цензурирование. Пусть $G(x)$ — функция распределения случайной величины η , $x \in R^1$. Видно, что случайные величины $\zeta_i = \min(\xi_i, \eta_i)$ незави-

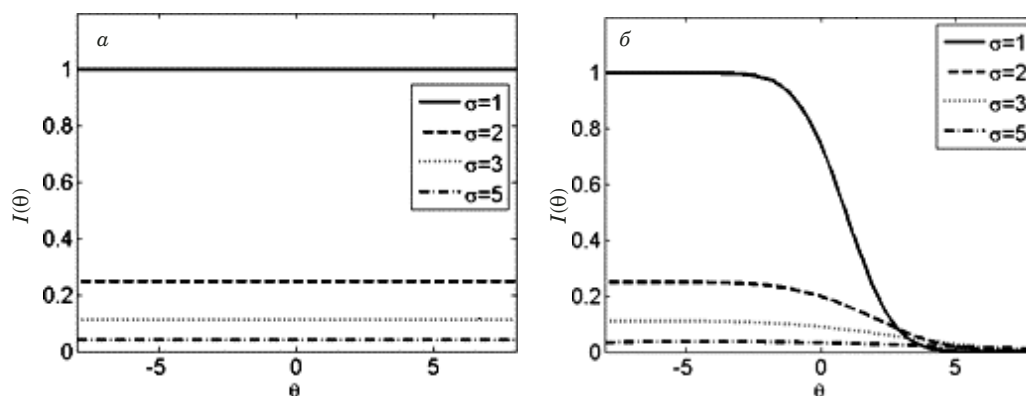


Рис. 1. Информация Фишера при отсутствии цензурирования (а) и при цензурировании справа (б); во втором случае случайная величина η имеет нормальное распределение с параметрами $(0, \sigma^2)$

Fig. 1. Fisher information without censoring (a) and with censoring on the right (b); in the second case, the random variable η obeys a normal distribution with parameters $(0, \sigma^2)$

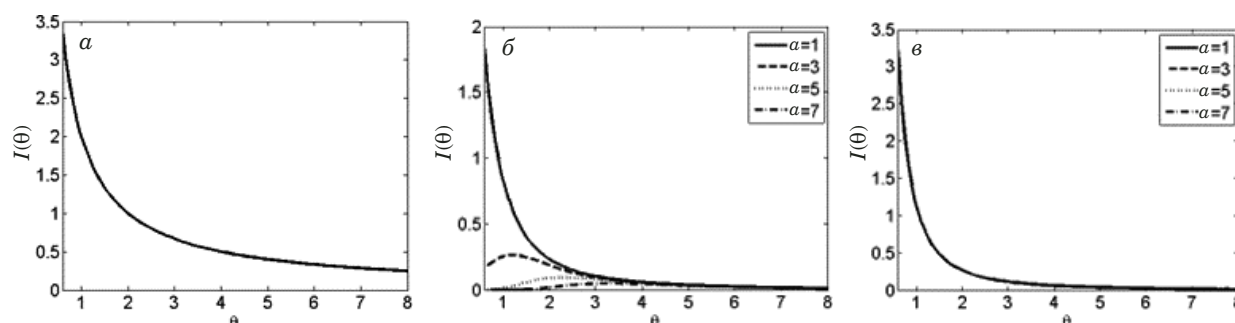


Рис. 2. Информация Фишера при отсутствии цензурирования (а) и при цензурировании — η имеет нормальное распределение с параметрами $(0, 1)$ (б); η имеет нормальное распределение с параметрами $(a, 1)$ (в)

Fig. 2. Fisher information without censoring (a) η has a normal distribution with parameters $(0, 1)$ (b); η has a normal distribution with parameters $(a, 1)$ (c)

симы и одинаково распределены с функцией распределения $H(x, \theta) = 1 - F(x, \theta)G(x)$. Вообще функция распределения G играет роль «мешающего параметра». Информация Фишера определяется следующим образом:

$$I(\theta) = I_G(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln(f(x, \theta))}{\partial \theta} \right)^2 f(x, \theta) \overline{G(x)} dx + \int_{-\infty}^{\infty} \left(\frac{\partial \ln(1 - F(x, \theta))}{\partial \theta} \right)^2 g(x) \overline{F(x, \theta)} dx.$$

Рассмотрим случай, когда случайная величина ξ имеет нормальное распределение с плотностью $f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \theta)^2}{2\sigma^2}\right]$, $x \in (-\infty, \infty)$, $\theta \in (-\infty, \infty)$, $\sigma > 0$, где σ — известный параметр. Пусть η также имеет нормальное распределение с параметрами $(0, 1)$ или $(0, \sigma^2)$. На рис. 1 показано графическое изображение информации Фишера. Видно, что в классических моделях информация Фишера является константой, а при неполных данных она уменьшается с увеличением значения неизвестного параметра.

Теперь рассмотрим случай, когда дисперсия является неизвестной, т.е. плотность распределения ξ имеет вид

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\theta} \exp\left[-\frac{(x - a)^2}{2\theta^2}\right],$$

$x \in (-\infty, \infty)$, $a \in (-\infty, \infty)$, $\theta > 0$. На рис. 2 представлено графическое изображение информации Фишера при отсутствии цензурирования и при цензурировании.

Из табл. 1 видно, что значение информации Фишера больше при отсутствии цензурирования, чем при его наличии.

Аналогичные результаты получены для гамма-распределения. На рис. 3 и 4 показаны графические изображения информации Фишера для полных и цензурированных наблюдений в случае, когда цензурирующая случайная величина имеет показательное распределение с параметрами 1 и a соответственно:

$$1) \text{ гамма-распределение с плотностью } f(x, \theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}, x > 0, \alpha > 0, \theta > 0;$$

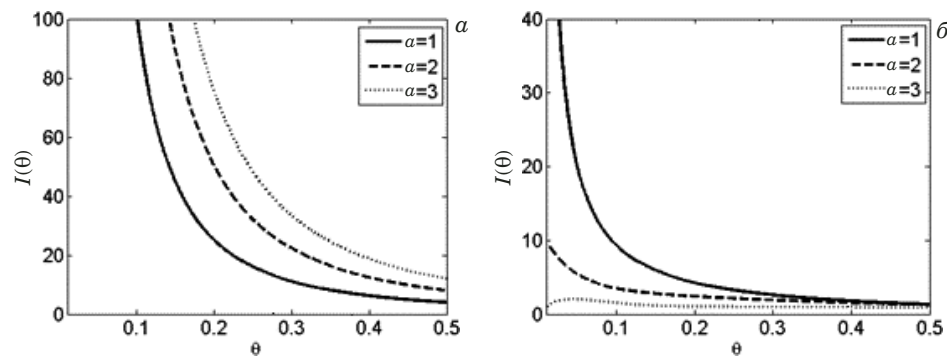


Рис. 3. Информация Фишера в случаях полных данных (а) и цензурирования справа — случайная величина η имеет показательное распределение с параметром 1 (б)

Fig. 3. Fisher information in case of complete data (a) and case of censoring on the right; in the second case, the random variable η has an exponential distribution with parameter 1 (b)

Таблица 1. Значения информации Фишера в моделях нецензурирования и цензурирования

Table 1. Fisher information values in uncensored and censored models

Модель	Параметр	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 5$	$\sigma = 10$
Полная	$I(\theta)$	1	0,25	0,11	0,04	0,01
Цензур.	$\theta = -8,0$	1,0000	0,2500	0,1109	0,0332	0,0024
	$\theta = -6,4$	1,0000	0,2499	0,1108	0,0364	0,0031
	$\theta = -4,8$	1,0000	0,2495	0,1099	0,0374	0,0040
	$\theta = -3,2$	0,9976	0,2462	0,1073	0,0369	0,0048
	$\theta = -1,5$	0,9566	0,2326	0,1009	0,0351	0,0055
	$\theta = 0,1$	0,7215	0,1956	0,0888	0,0324	0,0060
	$\theta = 1,7$	0,2886	0,1314	0,0703	0,0286	0,0062
	$\theta = 3,3$	0,0426	0,0630	0,0479	0,0240	0,0062
	$\theta = 4,9$	0,0019	0,0195	0,0270	0,0190	0,0060
	$\theta = 6,5$	0,0000	0,0037	0,0122	0,0140	0,0056
	$\theta = 8,0$	0,0000	0,0005	0,0048	0,0100	0,0052

2) гамма-распределение с плотностью $f(x, \theta) = \frac{\alpha^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\alpha x}, x > 0, \alpha > 0, \theta > 0$. Из графиков

можно сделать вывод, что информация Фишера для полных и неполных выборок может быть описана неравенством $I_{\text{полная}}(\theta) \geq I_{\text{неполная}}(\theta)$.

Теперь рассмотрим еще одну модель случайного цензурирования справа — изменение информации Фишера в модели многократного случайного цензурирования справа.

Информация Фишера при многократном случайном цензурировании справа

Пусть ξ — неотрицательная случайная величина, означающая продолжительность жизни испытываемого объекта, определена на вероятностном пространстве (Ω, A, P) со значениями в измеримом пространстве $(\mathbf{X}, \mathbf{B})$, где $\mathbf{X} \subseteq R^+$ и $\mathbf{B} = \sigma(\mathbf{X})$. Пусть $F(x, \theta) = P_\theta(\xi \leq x)$ и $\xi_{1n} < \xi_{2n} < \dots < \xi_{nn}$ — вариационный ряд, полученный из

случайных величин $\{\xi_1, \dots, \xi_n\}$. Тогда совместная плотность вектора $\xi^{(n)}$ задается формулой

$$p_n(x^{(m)}, \theta) = \frac{n!}{(n-m)!} \prod_{l=1}^m h^{(i)}(x_{ln}, \theta) [\overline{F}(x_{mn}, \theta)]^{n-m},$$

где $x^{(m)} = (x_{1n}, \dots, x_{mn})$ — реализация вектора $\xi^{(n)}$. Пусть $\beta \in (0, 1]$ такое, что при $n \rightarrow \infty, r/n \rightarrow \beta$ функция плотности интенсивности $h(x, \theta) = f(x, \theta) / \overline{F}(x, \theta)$. В данной схеме информация Фишера определяется следующим образом:

$$I_\beta(\theta) = \int_0^{\lambda_\beta} \left(\frac{\partial \ln h(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx, \tag{13}$$

где λ_β — квантиль уровня β распределения $F(x, \theta)$.

Рассмотрим случай, когда случайная величина ξ имеет нормальное распределение с плот-

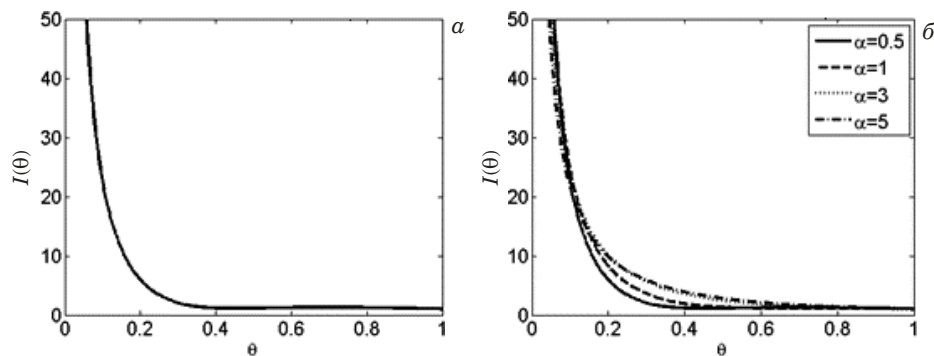


Рис. 4. Информация Фишера в случае полных данных (а) и цензурированных наблюдений — η имеет показательное распределение с параметром α (б)

Fig. 4. Fisher information in case of complete data (a) and censored observations; η has an exponential distribution with parameter α (b)

Таблица 2. Значения информации Фишера в модели многократного случайного цензурирования

Table 2. Fisher information values in progressive random censoring model

Квантиль	$\beta = 0,05$	$\beta = 0,50$	$\beta = 0,95$
$\theta = -8,0$	0,0000	0,0000	0,0000
$\theta = -6,4$	0,0000	0,0000	0,0000
$\theta = -4,8$	0,0000	0,0000	0,0000
$\theta = -3,2$	0,0000	0,0000	0,0000
$\theta = -1,5$	0,0000	0,0000	0,0022
$\theta = 0,1$	0,0000	0,0212	0,1961
$\theta = 1,7$	0,0160	0,6036	0,7786
$\theta = 3,3$	0,2248	0,8123	0,9873
$\theta = 4,9$	0,2308	0,8183	0,9933
$\theta = 6,5$	0,2307	0,8183	0,9933
$\theta = 8,0$	0,2307	0,8183	0,9933

ностью $f(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$, $x \in (-\infty, +\infty)$, $\theta \in (-\infty, +\infty)$. Покажем графически (рис. 5) примерные значения информации Фишера при различных значениях β в формуле (13). Сравнив эти данные с графиком информации Фишера в классической модели (см. рис. 1), можно сделать вывод, что и в последней модели для информации Фишера имеет место соотношение $I_{\text{полная}}(\theta) \geq I_{\text{неполная}}(\theta)$.

В табл. 2 приведены значения информации Фишера в модели многократного случайного цензурирования справа. Видно, что значение информации Фишера больше, когда квантиль стремится к 1.

Заключение

При рассмотрении задач оценивания неизвестных параметров по неполным наблюдениям большую роль играет зависимость цензурирую-

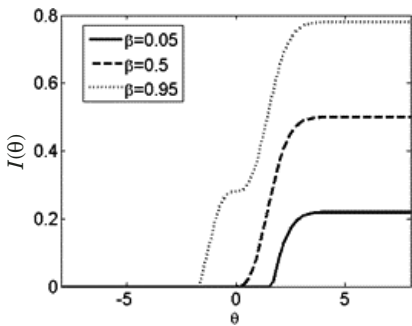


Рис. 5. Информация Фишера в модели многократного случайного цензурирования

Fig. 5. Fisher information in progressive random censoring model

щего распределения от того же параметра. В зависимости от этого различаются информативный и неинформативный случаи. Показано, что в информативном случае встречаются такие модели, в которых цензурирование может приводить и к сохранению информации, в неинформативном случае — наоборот. Численно показано, что неинформативное цензурирование приводит к потере фишеровской информации.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

ЛИТЕРАТУРА

1. **Fisher R. A.** Theory of Statistical Estimation / Proceedings of the Cambridge Philosophical Society. 1925. P. 700 – 725.
2. **Zacks Sh.** The theory of statistical inference. — New York: Wiley, 1971.
3. **Efron B., Zonstone I. M.** Fisher Information in terms of the hazard rate / Ann. Statistics 1990. Vol. 18. N 1. P. 38 – 62.
4. **Prakasa Rao B. L. S.** On Cramer – Rao type integral inequalities / Calcutta Statistical Association Bulletin. 1991. Vol. 40. P. 183 – 205.

5. **Prakasa Rao B. L. S.** Cramer – Rao type integral inequalities for functions of multidimensional parameter / *Sankhya: Ser. A.* 1992. Vol. 54. P. 53.
6. **Prakasa Rao B. L. S.** Improved Cramer – Rao type integral inequalities or Bayesian Cramer – Rao lower-band / *J. Indian Society for Probability and Statistics.* 2018(a). Vol. 19. P. 1 – 7.
7. **Prakasa Rao B. L. S.** Improved sequential Cramer – Rao type integral inequality / *Sequential Analysis.* 2018(b). Vol. 37. P. 59 – 68.
8. **Боровков А. А., Саханенко А. И.** Об оценках для усредненного квадратичного риска / *Probability and Mathematical Statistics.* 1980. Vol. 1. N 2. P. 185 – 195.
9. **Abdushukurov A. A., Kim L. V.** Lower Cramer – Rao and Bhattacharya bounds for randomly censored observation / *J. Soviet. Math.* 1987. Vol. 38. N 5. P. 2171 – 2185.
10. **Prakasa Rao B. L. S.** Remarks on Cramer – Rao type integral inequalities for random censored data / *Analysis of censored data.* 1995. Vol. 27. P. 163 – 175.
11. **Zheng G., Gastwirth J. H.** On the Fisher information in randomly censored data / *Statistics Probability Letters.* 2001. Vol. 52. P. 421 – 426.
12. **Zeng G., Gastwirth J. H.** Fisher information in ordered randomly censored data with application to characterization problems / *Statistica Sinica.* 2003. Vol. 13. P. 507 – 5017.
13. **Prakasa Rao B. L. S.** Improved Cramer – Rao inequality for randomly censored data / *J. Iranian Statistical Society.* 2018 (c). Vol. 17. P. 17 – 26.
14. **Prakasa Rao B. L. S.** Cramer – Rao Inequality Revisited for Randomly Censored Data / *Proceedings of Conf. “Statistics” and its applications* 2019. Tashkent. P. 19 – 28.
15. **Abdushukurov A. A., Erisbaev S. A.** Fisher information decomposition in terms of hazard rate functions under random censoring from the right / *Electronic J. Science and education in Karakalpakstan.* 2020. N 3 – 4. P. 36 – 43.
16. **Абдушукуров А. А., Эрисбаев С. А.** Информация Фишера и неравенство типа Крамера – Рао для модели конкурирующих рисков / *Bulletin of the Institute of Mathematics.* 2022. Vol. 4. N 5. P. 50 – 59.
17. **Abdushukurov A. A., Erisbaev S. A.** On Fisher information function in competing risks model under random censoring from unobservation intervals and efficiency / *Bulletin of the Institute of Mathematics.* 2022. Vol. 5. N 3. P. 13 – 19.
18. **Erisbaev S. A.** Calculation of the Fisher information in Competing risks model in combined gibrad right random censoring / *Bulletin of the Institute of Mathematics.* 2022. Vol. 5. N 6. P. 178 – 181.
19. **Эрисбаев С.** Полупараметрическое оценивание условной функции распределения в регрессионной модели Кокса при частично-информативном случайном цензурировании справа / *Bulletin of the Institute of Mathematics.* 2023. Vol. 6. N. 3. P. 114 – 120.
20. **Walker S. C.** A self-improvement to the Cauchy – Swartz inequality / *Statist. Probab. Lett.* 2017. Vol. 122. P. 86 – 90.
21. **Zheng G., Balakrishnan N., Parks.** Fisher information in ordered data: A review / *Statistics and its Interface.* 2009. Vol. 2. P. 101 – 113.
22. **Dorn M. F.** Evaluating Fisher Information in Order Statistics / *A research paper.* — University of Chicago, 2011. — 52 p.
23. **Нурмухамедова Н. С., Юсупов Ж. Р.** Информация Фишера в моделях случайного цензурирования справа / *Modern problems of applied mathematics and information technology AL-KHORASMY-2016.* P. 222 – 226.
24. **Абдушукуров А. А., Нурмухамедова Н. С.** Локальная асимптотическая нормальность статистических экспериментов и ее роль в теории оценивания и проверки гипотез / *Заводская лаборатория. Диагностика материалов.* 2016. Т. 82. № 3. С. 74 – 79.
2. **Zacks Sh.** The theory of statistical inference. — New York: Wiley, 1971.
3. **Efron B., Zonstone I. M.** Fisher Information in terms of the hazard rate / *Ann. Statistics* 1990. Vol. 18. N 1. P. 38 – 62.
4. **Prakasa Rao B. L. S.** On Cramer – Rao type integral inequalities / *Calcutta Statistical Association Bulletin.* 1991. Vol. 40. P. 183 – 205.
5. **Prakasa Rao B. L. S.** Cramer – Rao type integral inequalities for functions of multidimensional parameter / *Sankhya: Ser. A.* 1992. Vol. 54. P. 53.
6. **Prakasa Rao B. L. S.** Improved Cramer – Rao type integral inequalities or Bayesian Cramer – Rao lower-band / *J. Indian Society for Probability and Statistics.* 2018(a). Vol. 19. P. 1 – 7.
7. **Prakasa Rao B. L. S.** Improved sequential Cramer – Rao type integral inequality / *Sequential Analysis.* 2018(b). Vol. 37. P. 59 – 68.
8. **Borovkov A. A., Sakhanenko A. I.** On estimates of the average quadratic risk / *Probability and Mathematical Statistics.* 1980. Vol. 1. N 2. P. 185 – 195.
9. **Abdushukurov A. A., Kim L. V.** Lower Cramer – Rao and Bhattacharya bounds for randomly censored observation / *J. Soviet. Math.* 1987. Vol. 38. N 5. P. 2171 – 2185.
10. **Prakasa Rao B. L. S.** Remarks on Cramer – Rao type integral inequalities for random censored data / *Analysis of censored data.* 1995. Vol. 27. P. 163 – 175.
11. **Zheng G., Gastwirth J. H.** On the Fisher information in randomly censored data / *Statistics Probability Letters.* 2001. Vol. 52. P. 421 – 426.
12. **Zeng G., Gastwirth J. H.** Fisher information in ordered randomly censored data with application to characterization problems / *Statistica Sinica.* 2003. Vol. 13. P. 507 – 5017.
13. **Prakasa Rao B. L. S.** Improved Cramer – Rao inequality for randomly censored data / *J. Iranian Statistical Society.* 2018 (c). Vol. 17. P. 17 – 26.
14. **Prakasa Rao B. L. S.** Cramer – Rao Inequality Revisited for Randomly Censored Data / *Proceedings of Conf. “Statistics” and its applications* 2019. Tashkent. P. 19 – 28.
15. **Abdushukurov A. A., Erisbaev S. A.** Fisher information decomposition in terms of hazard rate functions under random censoring from the right / *Electronic J. Science and education in Karakalpakstan.* 2020. N 3 – 4. P. 36 – 43.
16. **Abdushukurov A. A., Erisbaev S. A.** Fisher information and Cramer – Rao inequality for Competing Risks Model / *Bulletin of the Institute of Mathematics.* 2022. Vol. 4. N 5. P. 50 – 59.
17. **Abdushukurov A. A., Erisbaev S. A.** On Fisher information function in competing risks model under random censoring from unobservation intervals and efficiency / *Bulletin of the Institute of Mathematics.* 2022. Vol. 5. N 3. P. 13 – 19.
18. **Erisbaev S. A.** Calculation of the Fisher information in Competing risks model in combined gibrad right random censoring / *Bulletin of the Institute of Mathematics.* 2022. Vol. 5. N 6. P. 178 – 181.
19. **Erisbaev S.** Semiparametric estimator of the conditional distribution function in the Cox regression model with partially informative random censoring on the right / *Bulletin of the Institute of Mathematics.* 2023. Vol. 6. N 3. P. 114 – 120.
20. **Walker S. C.** A self-improvement to the Cauchy – Swartz inequality / *Statist. Probab. Lett.* 2017. Vol. 122. P. 86 – 90.
21. **Zheng G., Balakrishnan N., Parks.** Fisher information in ordered data: A review / *Statistics and its Interface.* 2009. Vol. 2. P. 101 – 113.
22. **Dorn M. F.** Evaluating Fisher Information in Order Statistics / *A research paper.* — University of Chicago, 2011. — 52 p.
23. **Nurmukhamedova N. S., Yusupov J. R.** Fisher information in models of right random censoring / *Modern problems of applied mathematics and information technology AL-KHORASMY-2016.* P. 222 – 226.
24. **Abdushukurov A. A., Nurmukhamedova N. S.** Local asymptotic normality of statistical experiments and its role in estimation theory and hypothesis testing / *Industr. lab. Mater. diagn.* 2016. Vol. 82. N 3. P. 74 – 79.

REFERENCES