

УДК 517.24

ИНФОРМАЦИЯ ФИШЕРА ДЛЯ СЛУЧАЙНО ЦЕНЗУРИРОВАННЫХ НАБЛЮДЕНИЙ

Нурмухамедова Н. С. *

Аннотация

Информация Фишера является фундаментальным понятием в теории статистического вывода и играет важную роль в оценке параметров статистических моделей. Она представляет собой меру информативности наблюдаемых данных относительно неизвестных параметров модели. В условиях случайного цензурирования, когда часть наблюдений может быть неполной или подвергаться цензурированию, вычисление информации Фишера представляет собой сложную задачу. В данной работе исследовано информация Фишера в модели случайного цензурирования справа.

Ключевые слова: случайное цензурирование, информация Фишера, функция интенсивности, модель пропорциональных интенсивностей.

Пусть X и Y независимые случайные величины (с.в.) и определим $Z = \min(X, Y)$ - наблюдаемое с.в. и $\delta = I(X \leq Y)$. В случае случайного цензурирования с справа мы наблюдаем пары (Z, δ) . Это означает, что если событие произошло до момента цензурирования ($X \leq Y$), мы знаем точное время события X . Если же событие не произошло до момента цензурирования ($X > Y$), мы знаем только, что $X > Y$, то есть время события превышает время цензурирования. В n независимых экспериментах мы получаем выборку:

$$C^{(n)} = \{(Z_i, \delta_i), \quad i = 1, 2, \dots, n\}.$$

Информация Фишера случайно цензурированной выборки была изучена многими учеными. Абдушукуров и Ким [1], а также Пракаса Рао [2] исследовали информацию Фишера, когда распределения как наблюдаемой с.в. X , так и цензурирующей с.в. Y зависят от неизвестного параметра θ . Ган [3] представил выражения для информации Фишера через функции интенсивности, что позволяет анализировать потерю информации из-за цензурирования и изменение информации при наличии цензурированных данных. Эфрон [4] исследовал связь между информацией Фишера и функциями интенсивности, что упрощает вычисления и понимание влияния цензурирования. В данной статье рассмотрим, как вычисляется информация Фишера для случайно цензурированных наблюдений: когда цензурирующая случайная величина не зависит от неизвестного параметра, а также когда зависит от неизвестного параметра.

1. Цензурирующая случайная величина независит от неизвестного параметра. Пусть X с.в. с функцией распределения (ф.р.) $F(x; \theta)$ и плотностью $f(x; \theta)$, где $\theta \in \Theta \subseteq \mathbb{R}^s$ неизвестный параметр. Функция интенсивности для случайной величины X определяется как

$$\lambda_X(x; \theta) = \frac{f(x; \theta)}{\bar{F}(x; \theta)} = \frac{f(x; \theta)}{1 - F(x; \theta)},$$

*Нурмухамедова Н. С. – Национальный университет Узбекистана, gasulova_nargiza@mail.ru

где $\bar{F}(x; \theta) = 1 - F(x; \theta)$. Функция плотности интенсивности $\lambda_X(x; \theta)$ показывает, как изменяется вероятность наступления наблюдаемого события в данный момент времени.

Информация Фишера это мера точности оценки параметра θ на основе наблюдений. Информация Фишера о неизвестном параметре определяется следующим образом:

$$I_X(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \right].$$

Здесь $\frac{\partial}{\partial \theta} \log f(X; \theta)$ функция вклада, показывающая чувствительность логарифма плотности к изменению параметра θ . При выполнении определённых регулярных условий, Эфрон [4] показал, что информацию Фишера можно выразить через функцию интенсивности

$$I_X(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta} \log \lambda_X(x; \theta) \right)^2 f(x; \theta) dx. \quad (1)$$

Это равенство удобно для анализа информации Фишера в контексте цензурированных данных.

Пусть X и Y независимые с.в. с ф.р. $F(x; \theta)$ и $G(x)$ соответственно. Тогда совместная плотность вероятности наблюдения (Z, δ) при цензурировании справа имеет следующий вид:

$$L(z, \delta) = [f(z; \theta) \bar{G}(z)]^\delta [g(z) \bar{F}(z; \theta)]^{1-\delta}, \quad (2)$$

где $g(z)$ плотность распределения с.в. Y . Если функция распределения цензурирующей величины $G(x)$ не зависит от параметра θ , то производная по θ от $\bar{G}(x)$ равна нулю. Тогда выражение для плотности упрощается $L(z, \delta) = f^\delta(z; \theta) [1 - F(z; \theta)]^{1-\delta}$. Информация Фишера относительно параметра θ в наблюдении (Z, δ) определяется как:

$$I^{(Z, \delta)}(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log L(Z, \delta) \right)^2 \right] = E_\theta \left\{ \left(\delta \frac{\partial}{\partial \theta} \log f(Z; \theta) + (1 - \delta) \frac{\partial}{\partial \theta} \log(1 - F(Z; \theta)) \right)^2 \right\}. \quad (3)$$

Цензурирование существенно снижает информацию Фишера о параметре θ , что затрудняет точное оценивание параметра. Это демонстрирует необходимость учитывать цензурирование при статистическом анализе данных.

Пример. Пусть $X \sim \text{Exp}(\lambda)$ имеет экспоненциальное распределение с неизвестным параметром λ . Цензурирующая с.в. Y имеет известное распределение с функцией распределения $G(y)$. Информация Фишера в случае полной выборки $I_X(\lambda) = \frac{1}{\lambda^2}$. Рассмотрим случайное цензурирование справа, тогда информации Фишера имеет следующий вид

$$I^{(Z, \delta)}(\lambda) = E_\lambda \left[\left(\delta \left(\frac{1}{\lambda} - Z \right) + (1 - \delta) Z \lambda \right)^2 \right]. \quad (4)$$

Чтобы решить (4) нужно знать точное распределение с.в. Y . Если с.в. Y также имеет показательное распределение с параметром 1, то (4) имеет следующий вид:

$$I^{(Z, \delta)}(\lambda) = \frac{2\lambda^2}{\lambda + 1} - \frac{1}{\lambda^3 + \lambda^2} - \frac{2}{(\lambda + 1)^3} + \frac{1}{\lambda^2}.$$

Этот результат представляет собой информацию Фишера с учётом как нецензурированных, так и цензурированных данных. В случае случайного цензурирования справа информация Фишера уменьшается по сравнению с информацией Фишера в случае полной выборки.

Данный пример показывает, что цензурирование приводит к снижению информации Фишера по сравнению с полными данными. Однако цензурирование не всегда приводит к потере информации. В случае, когда цензурирующая случайная величина зависит от неизвестного параметра, а её ф.р. выражается через ф.р. с.в. X , информация Фишера может не изменяться. Рассмотрим этот случай.

2. Цензурирующая случайная величина зависит от неизвестного параметра. Рассмотрим случай, когда ф.р. цензурирующей с.в. Y , обозначаемая как $G(y; \theta)$, зависит от того же неизвестного параметра θ , что и ф.р. наблюдаемой величины X , $F(x; \theta)$. В таких случаях оценка информации Фишера становится более сложной задачей, так как необходимо учитывать зависимость обеих с.в. от параметра θ .

Пусть X и Y независимые с.в. с ф.р. $F(x; \theta)$ и $G(y; \theta)$ соответственно, где $\theta \in \Theta \subseteq \mathbb{R}^s$ - неизвестный параметр. Определим наблюдаемые величины $Z = \min(X, Y)$, $\delta = I(X \leq Y)$, где $I(A)$ - индикатор события A . Совместная плотность вероятности случайно правосторонне цензурированного наблюдения (Z, δ) определяется следующим образом:

$$L(z, \delta) = [f(z; \theta) \overline{G}(z; \theta)]^\delta [g(z; \theta) \overline{F}(z; \theta)]^{1-\delta}, \quad (5)$$

где $\overline{F}(x; \theta) = 1 - F(x; \theta)$ и $\overline{G}(x; \theta) = 1 - G(x; \theta)$. Информация Фишера относительно параметра θ в наблюдениях (Z, δ) определяется как:

$$I^{(Z, \delta)}(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log L(Z, \delta) \right)^2 \right]. \quad (6)$$

Используя выражение для $\log L(Z, \delta)$, получаем

$$\log L(Z, \delta) = \delta \log [f(Z; \theta) \overline{G}(Z; \theta)] + (1 - \delta) \log [g(Z; \theta) \overline{F}(Z; \theta)].$$

Учитывая это выражение для информации Фишера, получаем

$$I^{(Z, \delta)}(\theta) = \int_{-\infty}^{\infty} \left\{ \delta \left(\frac{\partial}{\partial \theta} \log f(z; \theta) + \frac{\partial}{\partial \theta} \log \overline{G}(z; \theta) \right) + (1 - \delta) \left(\frac{\partial}{\partial \theta} \log g(z; \theta) + \frac{\partial}{\partial \theta} \log \overline{F}(z; \theta) \right) \right\}^2 L(z, \delta) dz. \quad (7)$$

Рассмотрим следующие условия регулярности:

(У1) Носитель плотности $f(x; \theta)$ не зависит от θ : $N_f = \{x : f(x; \theta) > 0\}$ постоянно по θ .

(У2) Смешанные производные $\frac{\partial^2 F(x; \theta)}{\partial x \partial \theta}$ и $\frac{\partial^2 G(x; \theta)}{\partial x \partial \theta}$ существуют и равны $\frac{\partial f(x; \theta)}{\partial \theta}$ и $\frac{\partial g(x; \theta)}{\partial \theta}$ соответственно.

(У3) Производные $\frac{\partial f(x; \theta)}{\partial \theta}$ и $\frac{\partial g(x; \theta)}{\partial \theta}$ существуют и ограничены почти всюду для всех $\theta \in \Theta$.

(У4) Информации Фишера $I^X(\theta)$ и $I^Y(\theta)$ конечны и положительны: $0 < I^X(\theta), I^Y(\theta) < \infty$.

(У5) $\theta \in \Theta \subseteq \mathbb{R}$.

Теорема 1. Пусть выполнены условия регулярности (У1)-(У5). Тогда информация Фишера для (Z, δ) может быть выражена следующим образом:

$$\begin{aligned} I^{(Z, \delta)}(\theta) &= I^X(\theta) - \int_{-\infty}^{\infty} \left(\frac{\partial \log \lambda_X(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) G(x; \theta) dx + \\ &+ I^Y(\theta) - \int_{-\infty}^{\infty} \left(\frac{\partial \log \lambda_X(x; \theta)}{\partial \theta} \right)^2 g(y; \theta) F(y; \theta) dy = \\ &= \int_{-\infty}^{\infty} \left(\frac{\partial \log \lambda_X(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) \bar{G}(x; \theta) dx + \int_{-\infty}^{\infty} \left(\frac{\partial \log \lambda_Y(y; \theta)}{\partial \theta} \right)^2 g(y; \theta) \bar{F}(y; \theta) dy. \end{aligned} \quad (8)$$

Теперь рассмотрим случай, когда цензурирующая с.в. зависит от неизвестного параметра специальным образом: модель пропорциональных интенсивностей (МПИ). В МПИ между распределениями наблюдаемой и цензурирующей с.в. выполняется следующее соотношение:

$$\bar{G}(x; \theta) = [\bar{F}(x; \theta)]^\beta, \quad (9)$$

где $\beta > 0$ - константа, характеризующая степень цензурирования. Используя функции интенсивности с.в. X и Y выражение (9) можно написать следующим образом

$$\lambda_Y(x; \theta) = \beta \lambda_X(x; \theta).$$

Таким образом, МПИ подразумевает, что функция интенсивности для Y пропорциональна функции интенсивности для X с коэффициентом пропорциональности β . Имеет место

Теорема 2. В МПИ информация Фишера для (Z, δ) совпадает с информацией Фишера для X

$$I^{(Z, \delta)}(\theta) = I^X(\theta).$$

В МПИ цензурирование происходит таким образом, что информация о параметре θ не теряется, несмотря на наличие цензурирования. Это связано с тем, что цензурирующая величина Y несёт такую же информацию о θ , как и наблюдаемая величина X .

Пример. В этом примере мы покажем, что в МПИ информация Фишера для цензурированных наблюдений (Z, δ) совпадает с информацией Фишера для нецензурированных наблюдений X , то есть $I^{(Z, \delta)}(\theta) = I^X(\theta)$. Для этого рассмотрим случай, когда время до события X и время цензурирования Y распределены экспоненциально. Пусть $f_X(x; \theta) = \theta e^{-\theta x}$, $x \geq 0$, $f_Y(y; \theta) = \beta \theta e^{-\beta \theta y}$, $y \geq 0$. Проверим, что выполняется условие МПИ:

$$\bar{F}_Y(x; \theta) = [\bar{F}_X(x; \theta)]^\beta.$$

Действительно,

$$[\bar{F}_X(x; \theta)]^\beta = [e^{-\theta x}]^\beta = e^{-\beta \theta x} = \bar{F}_Y(x; \theta).$$

Вычислим информацию Фишера для (Z, δ) :

$$I^{(Z, \delta)}(\theta) = I^X(\theta) - \int_0^\infty \left(\frac{\partial}{\partial \theta} \log \lambda_X(x; \theta) \right)^2 f_X(x; \theta) \bar{F}_Y(x; \theta) dx +$$

$$I^Y(\theta) - \int_0^\infty \left(\frac{\partial}{\partial \theta} \log \lambda_Y(y; \theta) \right)^2 f_Y(y; \theta) \bar{F}_X(y; \theta) dy.$$

В МПИ $\lambda_Y(y; \theta) = \beta \lambda_X(y; \theta)$, и обе функции интенсивности имеют одинаковые логарифмические производные $\frac{\partial}{\partial \theta} \log \lambda_X(x; \theta) = \frac{1}{\theta}$, $\frac{\partial}{\partial \theta} \log \lambda_Y(y; \theta) = \frac{\partial}{\partial \theta} \log(\beta \theta) = \frac{1}{\theta}$. Учитывая это вычислим интегралы

$$\int_0^\infty \left(\frac{1}{\theta} \right)^2 f_X(x; \theta) \bar{F}_Y(x; \theta) dx = \frac{1}{\theta^2(1 + \beta)},$$

$$\int_0^\infty \left(\frac{1}{\theta} \right)^2 f_Y(y; \theta) \bar{F}_X(y; \theta) dy = \frac{\beta}{\theta^2(1 + \beta)}.$$

Подставляем все в выражение для $I^{(Z, \delta)}(\theta)$

$$I^{(Z, \delta)}(\theta) = I^X(\theta) - \frac{1}{\theta^2(1 + \beta)} + I^Y(\theta) - \frac{\beta}{\theta^2(1 + \beta)}.$$

Учитывая, что $I^X(\theta) = \frac{1}{\theta^2}$ и $I^Y(\theta) = \frac{1}{\theta^2}$, получаем

$$I^{(Z, \delta)}(\theta) = \frac{1}{\theta^2} - \frac{1}{\theta^2(1 + \beta)} + \frac{1}{\theta^2} - \frac{\beta}{\theta^2(1 + \beta)} = \frac{1}{\theta^2}.$$

Таким образом в МПИ

$$I^{(Z, \delta)}(\theta) = \frac{1}{\theta^2} = I^X(\theta).$$

ЛИТЕРАТУРА

1. Abdushukurov, I.A., Kim, L.V. Lower Cramer-Rao and Bhattacharya bounds for randomly censored observations J. Soviet Math. 1987. P. 2171–2185.
2. Prakasa Rao, B. L. S. Identifiability in Stochastic Models: Characterization of Probability Distributions. 1992. Academic Press.
3. Gang, A. L. Fisher Information in Randomly Censored Data. Journal of Statistical Planning and Inference, 2000. 90(2), P. 333–345.
4. Efron, B., Hinkley, D. V. Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. Biometrika, 1978. 65(3), P. 457–483.

REZYUME

Fisher informatsiyasi statistik xulosa qilish nazariyasining asosiy tushunchalaridan biri bo'lib, statistik modellarning parametrlarini baholashda muhim rol o'ynaydi. U noma'lum parametrga nisbatan kuzatilmaning informativligini o'lchash mezoni hisoblanadi. Tasodifiy senzurlanish modelida, ya'ni kuzatilmalarning bir qismi to'liq bo'lmagan hollarda, Fisher informatsiyasini hisoblash murakkab hisoblanadi. Ushbu maqolada o'ng tomondan tasodifiy senzurlanish modelida Fisher informatsiyasi o'rganilgan.

Kalit so'zlar: tasodifiy senzurlanish, Fisher informatsiyasi, intensivlik funksiyasi, proporsional intensivliklar modeli.

RESUME

Fisher information is a fundamental concept in the theory of statistical inference and plays an important role in parameter estimation of statistical models. It represents a measure of the informativeness of observed data with respect to the unknown parameters of the model. In the context of random censoring, where part of the observations may be incomplete or censored, calculating Fisher information becomes a challenging task. This paper investigates Fisher information in the model of right-censored data.

Key words: random censoring, Fisher information, hazard function, proportional hazards model.