Теория вероятностей и математическая статистика—2
Теоретический и задачный минимумы
ФЭН НИУ ВШЭ

Винер Даниил @danya_vin

Версия от 12 мая 2025 г.

Содержание

1 Теоретический минимум

1.1 Сформулируйте неравенство Крамера - Рао для несмещённых оценок

Пусть $\widehat{\theta}$ — несмещенная оценка параметра θ, а также выполняются все условия гладкости и регулярности, тогда для несмещённых оценок верно:

$$\mathbb{D}\left[\widehat{\theta}\right] \geqslant \frac{1}{I(\theta)}$$

1.2 Дайте определение функции правдоподобия и логарифмической функция правдоподобия

Определение. Пусть задана случайная выборка $X = (X_1, \ldots, X_n)$, компоненты которой имеют функцию распределения $F(x;\theta)$, зависящую от неизвестного параметра $\theta \in \Theta$

- Для абсолютно непрерывных величин: $\mathcal{L}\left(x_1, \ldots, x_n; \theta\right) = f\left(x_1, \ldots, x_n; \theta\right) = \prod\limits_{i=1}^{n} f(x_i; \theta)$
- Для дискретных величин: $\mathcal{L}\left(x_1, \ldots, x_n; \theta\right) = \mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n; \theta\right) = \prod\limits_{i=1}^{n} \mathbb{P}\left(x_i; \theta\right)$

Определение. Логарифмической функцией правдоподобия называется функция

$$l\left(x_1, \ldots, x_n; \theta\right) := \ln \mathcal{L}\left(x_1, \ldots, x_n; \theta\right)$$

1.3 Дайте определение информации Фишера о параметре θ, содержащейся в одном наблюдении

Имеется выборка с неизвестным параметром — $X_1, \ldots, X_n \sim F(x;\theta)$

Определение. Информацией Фишера называется

$$I(\theta; X) = \mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right],$$

где $\mathcal{L}$ — функция правдоподобия

Примечание. Определние применимо для регулярного случая, то есть область значений X не зависит от θ

Определение. Равносильное определение информации Фишера:

$$I(\theta; X) = -\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right]$$

1.4 Дайте определение оценки метода моментов параметра θ с использованием первого момента, если $\mathrm{E}\left(X_i\right) = g(\theta)$ и существует обратная функция g^{-1}

Оценка ММ параметра θ определяется как:

$$\hat{\theta}_{MM} = g^{-1}\left(\bar{X}\right),$$

где $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ — выборочное среднее, $g(\theta) = \mathbb{E}(X_i)$, а g^{-1} — обратная функция к g, существующая по условию[1]

[1] стр.23 учебника Черновой — тоже самое, только в общем виде

1.5 Дайте определение оценки метода максимального правдоподобия параметра θ

Определение. Оценкой $\hat{\theta}_{ML}$ неизвестного параметра $\theta \in \Theta$ по ММП называется точка глобального максимума функции правдоподобия по переменной $\theta \in \Theta$ при фиксированных значениях переменных $x_1, \ldots, x_n$, т.е.

$$\mathcal{L}\left(x_1, \ldots, x_n; \hat{\theta}_{ML}\right) = \max_{\theta \in \Theta} \mathcal{L}\left(x_1, \ldots, x_n; \theta\right)$$

1.6 Укажите закон распределения выборочного среднего, величины $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, величины $\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}}$, величины $\frac{\hat{\sigma}^2(n-1)}{\sigma^2}$

Пусть $X_1, \ldots, X_n$ — независимые нормальные случайные величины с параметрами μ и σ^2, $\bar{X} := \frac{X_1+\ldots+X_n}{n}$ - выборочное среднее, а $\widehat{\sigma^2} := \frac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right)^2$ исправленная выборочная дисперсия. Тогда

- $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1)$
- $\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$
- $\frac{\hat{\sigma}^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$

1.7 Укажите формулу доверительного интервала с уровнем доверия $(1-\alpha)$ для μ при известной дисперсии, для μ при неизвестной дисперсии, для σ^2

Дана выборка $X_1, \ldots X_n \sim N\left(\mu, \sigma^2\right)$

- Если известна σ^2:
$$\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right),$$
где $z_{\alpha/2}$ — квантиль стандартного нормального распределения
- Если σ^2 неизвестна:
$$\left(\bar{X} - t_{\alpha/2,\, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}},\ \bar{X} + t_{\alpha/2,\, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right),$$
где $t_{\alpha/2,\, n-1}$ — квантиль распределения Стьюдента с $n-1$ степенями свободы
- Для σ^2:
$$\left(\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}},\ \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\alpha/2}}\right),$$
где $\chi^2_{\alpha/2}$, $\chi^2_{1-\alpha/2}$ — квантили хи-квадрат распределения с $n-1$ степенями свободы.

1.8 Дайте определение ошибки первого и второго рода, критической области

Определение. Есть выборка $X_1, \ldots, X_n$, а множество значений $\mathcal{X} \in \mathbb{R}^n$

$$\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$$
$$\mathcal{X}_0 \cap \mathcal{X}_1 = 0$$

$\mathcal{X}_1$ — критическая область, где H_0 отвергается, а в $\mathcal{X}_0$ — не отвергается

Определение. Ошибка первого рода — вероятность отвергнуть H_0, когда она на самом деле верна:

$$\mathbb{P}(X \in \mathcal{X}_1 \mid H_0 \text{ верна})$$

Определение. Ошибка второго рода — вероятность не отвергнуть H_0, когда на самом деле верна H_1:

$$\mathbb{P}(X \in \mathcal{X}_0 \mid H_1 \text{ верна})$$

Определение. Говорят, что произошла ошибка i-го рода критерия δ, если критерий отверг верную гипотезу H_i. Вероятностью ошибки i-го рода критерия $\boldsymbol{\delta}$ называется число

$$\alpha_i(\boldsymbol{\delta}) = \mathrm{P}_{H_i}\left(\boldsymbol{\delta}(\vec{X}) \neq H_i\right)$$

1.9 Укажите формулу доверительного интервала с уровнем доверия $(1-\alpha)$ для вероятности успеха, построенного по случайной выборке большого размера из распределения Бернулли $\mathrm{Bin}(1,p)$

При больших n почти всегда имеет место интервал:

$$\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Тогда, для выборки из распределения Бернулли $\mathrm{Bin}(1,p)$:

$$\left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right),$$

где

- $\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ — выборочная доля успехов,
- $z_{\alpha/2}$ — квантиль стандартного нормального распределения

2 Задачный минимум