Экзаменационные вопросы по теории вероятностей

Андрей Ильин

май 2025

Билет 8. Многопараметрические модели. Матрица информации Фишера. Пример для нормального распределения

---

1. Что такое многопараметрические модели?

Многопараметрическая модель — это статистическая модель, в которой распределение данных зависит от нескольких параметров. Тогда вектор параметров обозначается как:

$$\theta = (\theta_1, \theta_2, \dots, \theta_k)$$

Это естественно возникает во многих задачах:

- В нормальном распределении: два параметра — среднее μ и дисперсия σ^2
- В линейной регрессии: $\beta_0, \beta_1, \dots, \beta_p$

В таких моделях мы ищем совместную оценку всех параметров, и важно понимать, как информация о данных распределяется по параметрам и между ними.
—

2. Матрица информации Фишера (Fisher Information Matrix)

Это обобщение скалярной информации Фишера на многомерный случай.

Если $L(\theta)$ — функция правдоподобия, а $\ell(\theta) = \log L(\theta)$ — логарифмическая функция правдоподобия, то матрица Фишера определяется как:

$$\boxed{\mathcal{I}(\theta) = \mathbb{E}\left[\nabla_\theta \ell(\theta) \cdot \nabla_\theta \ell(\theta)^T\right] = -\mathbb{E}\left[\nabla_\theta^2 \ell(\theta)\right]}$$

Где:

- $\nabla_\theta \ell(\theta)$ — градиент (вектор первых производных)
- $\nabla_\theta^2 \ell(\theta)$ — гессиан (матрица вторых производных)

Интуитивно:

- Матрица Фишера описывает информацию, которую данные содержат о каждом параметре.
- Диагональные элементы отражают "точность" оценок каждого параметра.
- Вне диагонали — зависимости между параметрами (если они есть).

Чем больше элемент $\mathcal{I}_{ii}$, тем меньше асимптотическая дисперсия оценки параметра θ_i.
—

3. Пример: $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, оба параметра неизвестны

1. Запишем лог-правдоподобие:

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

2. Вычислим производные:

Градиент по μ и σ^2:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu), \quad \frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (X_i - \mu)^2$$

3. Вычисляем матожидания вторых производных и получаем:

$$\mathcal{I}(\mu, \sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

Интерпретация:

- Элементы на диагонали:

$$\frac{n}{\sigma^2} \quad \text{— информация о } \mu$$
$$\frac{n}{2\sigma^4} \quad \text{— информация о } \sigma^2$$

- Нет корреляции между параметрами (обнуляется вне диагонали) — потому что лог-правдоподобие в нормальном распределении с неизвестными μ и σ^2 устроено так, что его частные производные по этим переменным не зависят друг от друга.

Следовательно, оценки $\hat{\mu}$ и $\hat{\sigma}^2$ асимптотически независимы.