Больше определений Богу определений

Оценки и их свойства

Чертоги разума

Определение. *Вероятностно-статистической моделью* называют тройку $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, где $\mathcal{F}$ — σ-алгебра на $\mathcal{X}$, а $\mathcal{P} = \{\mathsf{P}_\alpha : \alpha \in \Lambda\}$ — семейство вероятностных мер на $(\mathcal{X}, \mathcal{F})$.

Основной пример. $\mathcal{X} = \left(\mathbb{R}^k\right)^n$, $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{F} = \mathcal{B}(\mathcal{X})$, а распределения из $\mathcal{P}$ индуцированы некоторым случайным элементом $\mathbf{X} = (X_i \in \mathbb{R}^k : i = 1, \ldots, n)$ и имеют вид $\mathsf{P}_\alpha = \mathsf{P}^{\otimes n}$, то есть $X_1, X_2, \ldots$ независимы в совокупности и имеют распределение P. Такой случайный элемент называют *выборкой.*

Замечание. Впредь под $\mathcal{P}$ будем подразумевать семейство из P, а не $\mathsf{P}^{\otimes n}$.

Зачастую распределение P можно охарактеризовать каким-то конечномерным параметром, то есть $\mathcal{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, где $\Theta \subset \mathbb{R}^m$.

Примеры.

- $\{\mathrm{Bern}(p) : p \in (0; 1)\}$;
- $\{\mathrm{Pois}(\lambda) : \lambda \in (0; +\infty)\}$;
- $\{\mathcal{N}(a, \sigma^2) : a \in \mathbb{R}, \sigma > 0\}$;
- *модель сдвига* — $\{\theta + \mathsf{P}_0 : \theta \in \mathbb{R}\}$;
- *модель масштаба* — $\{\theta \cdot \mathsf{P}_0 : \theta \in \mathbb{R}_+\}$;
- *модель сдвига-масштаба* — $\{a \cdot \mathsf{P}_0 + b : (a, b) \in \mathbb{R}_+ \times \mathbb{R}\}$.

Определение. *Статистикой* называют композицию выборки $\mathbf{X}$ и измеримого отображения $S : \mathcal{X} \to \mathbb{R}^l$, обозначают как $S(\mathbf{X})$. Статистика со значениями в Θ называют *оценкой.* Обычно их обозначают как $\widehat{\theta}(\mathbf{X})$, $\widetilde{\theta}(\mathbf{X})$, $\theta^*(\mathbf{X})$ и т.д.

- Пусть $S : (x_1, \ldots, x_n) \mapsto \frac{g(x_1) + \ldots + g(x_n)}{n}$, тогда $S(\mathbf{X}) = \frac{g(X_1) + \ldots + g(X_n)}{n} =: \overline{g(\mathbf{X})}$ — *выборочная характеристика.* В частности, выборочную характеристику для $g(x) = x$ называют *выборочным средним*, а для $g(x) = x^k$ — *выборочным k-ым моментом*;
- $s^2(\mathbf{X}) := \overline{X^2} - \overline{X}^2$ — *выборочная дисперсия.*

Цель. Найти такую оценку, которая бы «лучше всего» приближала истинное значение θ или какой-то функции $\tau(\theta)$. Например, в нормальной модели чаще всего оценивают σ^2, а не σ.

Свойства оценок.

- несмещённость — $\forall \theta \in \Theta : \mathsf{E}_\theta \widehat{\theta}(X_1, \ldots, X_n) = \theta$;
- состоятельность — $\forall \theta \in \Theta : \widehat{\theta}(X_1, \ldots, X_n) \xrightarrow{\mathsf{P}_\theta} \theta$ при $n \to \infty$;
- сильная состоятельность — $\forall \theta \in \Theta : \widehat{\theta}(X_1, \ldots, X_n) \xrightarrow{\mathsf{P}_\theta\text{-п.н.}} \theta$ при $n \to \infty$;
- асимптотическая нормальность —

$$\forall \theta \in \Theta : \sqrt{n}\left(\widehat{\theta}(X_1, \ldots, X_n) - \theta\right) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2(\theta)), \quad n \to \infty.$$

Задачи

Напоминание из теорвера (дельта-метод).

Пусть $\Theta \subset \mathbb{R}^m$, $\widehat{\boldsymbol{\theta}}$ — асимптотически нормальная оценка параметра $\boldsymbol{\theta}$ с асимптотической ковариационной матрицей $\Sigma(\boldsymbol{\theta})$, а $\tau\colon \mathbb{R}^m \to \mathbb{R}^k$ дифференцируема на Θ. Тогда $\tau\big(\widehat{\boldsymbol{\theta}}\big)$ асимптотически нормальная оценка $\tau(\boldsymbol{\theta})$ с асимптотической ковариационной матрицей $d_{\boldsymbol{\theta}}\tau \cdot \Sigma(\boldsymbol{\theta}) \cdot d_{\boldsymbol{\theta}}\tau^T$, то есть

$$\sqrt{n}\left(\tau\big(\widehat{\boldsymbol{\theta}}\big) - \tau(\boldsymbol{\theta})\right) \xrightarrow{d_{\boldsymbol{\theta}}} \mathcal{N}(\mathbf{0}, d_{\boldsymbol{\theta}}\tau \cdot \Sigma(\boldsymbol{\theta}) \cdot d_{\boldsymbol{\theta}}\tau^T).$$

1. (а) Докажите, что асимптотически нормальная оценка является также и состоятельной.

(б) Предложите какую-нибудь модель и состоятельную оценку параметра в ней, которая не является асимптотически нормальной. *Замечание.* Считаем константу нормально распределённой с дисперсией 0.

2. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения с параметром σ^2, где $\mathsf{D}_{\sigma^2} X_1 = \sigma^2$ и $\mathsf{E}_{\sigma^2} X_1^4 < \infty$. Докажите, что $s^2(\mathbf{X})$ является сильно состоятельной и асимп. нормальной оценкой для σ^2. Является ли она несмещённой оценкой для σ^2?

3. Пусть выборка $\mathbf{X} = (X_1, \ldots, X_n)$ пришла из распределения со средним μ, дисперсией σ^2 и конечным четвёртым моментом. При каком условии предельное распределение случайного вектора

$$\sqrt{n}\left(\begin{pmatrix} \overline{\mathbf{X}} \\ s^2(\mathbf{X}) \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}\right)$$

имеет независимые компоненты?

4. Пусть $\mathbf{X} = (X_1, \ldots, X_n)$ — выборка из $\mathrm{U}(0, \theta)$. Для какой функции от параметра θ оценка $e^{-\overline{\mathbf{X}^2}}$ будет состоятельной? Является ли она при этом несмещённой? А асимптотически нормальной?

5. Пусть $\mathbf{X} = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{N}(\theta, 1)$, где $\theta > 0$. Покажите, что $\overline{\mathbf{X}}^{-1}$ является асимптотически нормальной оценкой $1/\theta$. Какая у неё асимптотическая дисперсия? А какая обычная?

6. Постройте несмещённую асимптотически нормальную оценку для параметра $e^{-\theta}$ с помощью выборки $\mathbf{X} = (X_1, \ldots, X_n)$ из распределения $\mathrm{Pois}(\theta)$, где $\theta > 0$, и найдите её асимптотическую дисперсию.

7. Оценка $\widehat{\theta}_n(\mathbf{X})$ параметра θ называется *асимптотически несмещённой*, если

$$\forall \theta\colon \mathsf{E}_\theta \widehat{\theta}_n \to \theta, \quad n \to \infty.$$

Докажите, что асимптотически несмещённая оценка со стремящейся к нулю дисперсией является состоятельной.

8. Пусть $G(n, p)$ — случайный граф в модели Эрдеша-Реньи (каждое ребро берётся независимо от других с вероятностью p). Найдите какую-нибудь сильно состоятельную оценку параметра p как функцию от числа треугольников в графе.