Заняття 2.

Точкові оцінки невідомих параметрів

Метод максимальної вірогідності. Емпірична функція розподілу. Порядкові статистики. Гістограма. Ядерні оцінки щільності.

Останній термін здачі завдання: 17.03.2017.

Оцінка за завдання обчислюється за формулою
$\max(3 * \text{Кількість набраних балів},\ 100)\%$.

1. (4 бали) За значеннями вибірки

   1.1, 2.1, 1.3, 0.9, 2.7, 1.4, 1.6, 1.3, 1.4, 2.4

   знайти а) вибіркове середнє, б) вибіркову дисперсію, в) варіаційний ряд. Побудуйте гістограму з трьома класами $[0,1), [1,2)$ та $[2,3]$.

2. (3 бали) Припустимо, що $X_1, ..., X_n$ – вибірка з рівномірного розподілу $U(0,1)$. Нехай $[a,b] \subset [0,1]$. Покладемо $Y_k = 1$, якщо $X_k \in [a,b]$, та $Y_k = 0$ в супротивному випадку.

   а) Знайти математичне сподівання $\mathrm{E}Y$.

   б) Нехай $m \in \mathbb{N}$. Знайти середню кількість інтервалів з сукупності $[0, \frac{1}{m}), [\frac{1}{m}, \frac{2}{m}), \dots, [\frac{m-1}{m}, 1]$ в які не потрапило жодного зі спостережень $X_1, ..., X_n$. Яким має бути m, щоб це число не перевищувало 0.02, якщо $n = 1000$?

3. (1 бал) Нехай $\{Y_n\}$ – послідовність випадкових величин така, що

$$\lim_{n\to\infty} \mathrm{E}Y_n = a, \quad \lim_{n\to\infty} DY_n = 0.$$

   Довести, що

$$\lim_{n\to\infty} \mathrm{E}(Y_n - a)^2 = 0.$$

   Зауважте, що з цього випливає збіжність послідовності $\{Y_n\}$ до a за ймовірністю при $n \to \infty$.

4. (4 бали) а) Знайти оцінку максимальної вірогідності θ^* невідомого параметру $\theta > 0$ рівномірного розподілу $U(\theta, 2\theta)$.

   б) Знайти $\mathrm{E}\theta^*, D\theta^*$.

   в) Чи буде ця оцінка консистентною?

5. (1 бал) Знайти оцінку максимальної вірогідності a^*, b^* невідомих параметрів a, b рівномірного розподілу $U(a, b)$.

6. (4 бали) Знайти оцінку максимальної вірогідності параметру $\alpha > 0$ для розподілу, що має густину а) $p(x|\alpha) = \alpha x^{\alpha-1},\ x \in [0,1]$, б) $p(x|\alpha) = \frac{\alpha}{x^{\alpha+1}},\ x \geq 1$.

   Чи будуть ці оцінки консистентними?

7. (1 бал) Нехай $\{X_n\}$ – послідовність незалежних однаково розподілених випадкових величин з щільністю $p(x)$. Припустимо, що $p(x) = 0, x < a$ та існує $\delta > 0$ таке, що $p(x) > 0, x \in [a, a+\delta]$. Довести, що $\min_{k=1,\ldots,n} X_k$ прямує до a за ймовірністю при $n \to \infty$.

8. (4 бали) Знайти оцінку максимальної вірогідності параметрів $\alpha > 0$ та x_0 для розподілу, що має густину а) $p(x|\alpha, x_0) = \alpha e^{-\alpha(x-x_0)},\ x \geq x_0$, б) $p(x|\alpha, x_0) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}},\ x \geq x_0$.

   Чи будуть ці оцінки консистентними?

9. (6 балів) Знайти оцінку максимальної вірогідності для невідомого параметру а) біноміального розподілу $Bi(m, p)$ (m відоме, p невідоме), б) розподілу Пуассона $Pois(\lambda)$, в) геометричного розподілу $G(p)$.

   Чи будуть ці оцінки консистентними?

10. (4 бали) Знайти оцінку максимальної вірогідності α_n^* невідомого параметру для розподілу з задачі 2 домашнього завдання №1.

    Знайти $\mathrm{E}\alpha_n^*,\ D\alpha_n^*$. Чи буде ця оцінка консистентною?

11. (2 бали) Припустимо, що $X_1, \ldots, X_n$ – вибірка з розподілу, що має щільність $p(x|x_0, \gamma) = \frac{1}{\pi}\frac{\gamma}{(x-x_0)^2+\gamma^2}$ (розподіл Коші), Знайдіть консистентну оцінку для невідомих параметрів x_0, γ.

Підказка. Знайдіть медіану та нижній квартиль розподілу.

12. (1 бал) За значеннями вибірки 1, 2, 1, 5, 7, 8, 16, -3, -2, яку взято з розподілу з щільністю $p(x|a) = 0.5e^{-|x-a|}$, $x \in \mathbb{R}$, знайти оцінку максимальної вірогідності невідомого параметру a.

13. (2 бали) Припустимо, що $X_1, ..., X_n$ – вибірка з показникового розподілу $Exp(\alpha)$, $Y_k = 0.1[10X_k]$ – округлення X_k вниз до першого десяткового знаку після коми (наприклад, якщо $X_k = 0.1923$, то $Y_k = 0.1$). Знайдіть консистентну оцінку для α.

14. (3 бали) Змоделюйте n спостережень із показникового розподілу $Exp(1)$.

    а) Побудуйте гістограми та порівняйте їх з щільністю розподілу для різної кількості класів $m = 5, 10, 20, 30, 50, 70, 100$.

    б) Побудуйте графіки відхилення оцінки максимальної вірогідності α_n^* від істинного параметру в залежності від n. Як ви думаєте, який порядок має швидкість спадання $O(1/n), O(1/n^2), O(1/2^n), O(1/\sqrt{n}), O(1/\ln n)$, і т.п.?

    В цій задачі вимагається не строге математичне обґрунтування, а емпіричний аналіз графіку.

15. (4 бали) Нехай $X_1, ..., X_n$ – вибірка з рівномірного розподілу $U(0, 1)$, $F(x)$ – функція розподілу X_k, $\hat{F}_n(x)$ – емпірична функція розподілу, $D_n = \sup_x |\hat{F}_n(x) - F(x)|$.

    За допомогою моделювання знайдіть приблизно:

    а) математичне сподівання, дисперсію та графік щільності D_5;

    б) таке m, що $\mathrm{P}(D_m > 0.1) \approx 0.05$.

16. (7 балів) Змоделюйте n спостережень із а) рівномірного розподілу $U(0, 1)$, б) нормального розподілу $N(0, 1)$.

    Побудуйте графіки ядерних оцінок $\hat{p}_n(x) = \frac{1}{nh_n} \sum_{k=1}^{n} K(\frac{x-X_k}{h_n})$ для щільностей розподілу та порівняйте їх з істинними значеннями, якщо

$n = 10, 100, 1000$. Візьміть а) $K(x) = \mathbf{1}_{|x| \le 0.5}$, б) $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Значення h_n оберіть самостійно.

Обчисліть (чисельно) $a_n = \max_x |p(x) - \hat{p}_n(x)|$, $b_n = \int_{-\infty}^{\infty} |p(x) - \hat{p}_n(x)| dx$.

Чи можна зробити $a_n < 0.2$ для рівномірного розподілу, для якогонебудь n і h_n, якщо $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$?

17. (1 бал) Нехай U_1, U_2 – незалежні випадкові величини, що мають рівномірний розподіл $U(0,1)$. Доведіть, що $(X_1, X_2) = (\sqrt{-2 \ln U_1} \cos(2\pi U_2), \sqrt{-2 \ln U_1} \sin(2\pi U_2))$ – незалежні випадкові величини, що мають нормальний розподіл $N(0,1)$.