Лабораторная №2

Черновая версия описания задачи (будет дополняться и уточняться)

Задание

1. Сгенерировать выборки для разных распределений и посчитать доверительные интервалы для них. Распределения: нормальное, бернулли, пуассона, экспоненциальное.
2. Построить ядерную оценку для `N(3, 5)` с прямоугольным и гауссовским ядром (за h принять h_{opt}). То же самое сделать для `uniform[3, 5]`

Дополнительная информация

Часть 1. Генерация выборок и доверительные интервалы

Доверительным интервалом параметра θ с уровнем доверия $1 - \alpha$ называется

$$P\{\hat{\theta}_{1,n} \leq \theta \leq \hat{\theta}_{2,n}\} = 1 - \alpha$$

$\alpha = 0.05$

Нормальное распределение

Нормальное распределение $N(M, \sigma^2), \{x_1, \dots, x_n\}$ имеет функцию распределения

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-M)^2}{2\sigma^2}} dt$$

Доверительный интервал для M

$$\overline{x_n} - \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \leq M \leq \overline{x_n} + \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}$$

Выборку генерировать размера $n = 200$
За k принять $k = [log_2 200] + 1 \approx 8$

Доверительный интервал для σ^2

$$\frac{n-1}{\chi^2_{1-\frac{\alpha}{2},n-1}} S_n^2 \leq \sigma^2 \leq \frac{n-1}{\chi^2_{\frac{\alpha}{2},n-1}} S_n^2$$

Распределение Бернулли

Распределение Бернулли $B(0.3; 200)$

Доверительный интервал

$$\frac{m}{n} - \mu_{\frac{\alpha}{2}} \frac{\sqrt{m(1-\frac{m}{n})}}{n} \leq P \leq \frac{m}{n} + \mu_{\frac{\alpha}{2}} \frac{\sqrt{m(1-\frac{m}{n})}}{n}$$

Для выборки размера n= 50

$$B(0.3; 50)$$

Можно посчитать отдельно оценки $P_{нижс}$ и $P_{верх}$, такие что $P_{нижс} \leq P \leq P_{верх}$
[Доверительный интервал по формуле Уилсона](#):

$$\hat{P}_{нижс} = \frac{\frac{m}{n} + \frac{\mu^2_{\frac{\alpha}{2}}}{2n} - \mu_{\frac{\alpha}{2}} \sqrt{\frac{m}{n^2}(1-\frac{m}{n}) + \frac{\mu^2_{\frac{\alpha}{2}}}{4n^2}}}{1 + \frac{\mu^2_{\frac{\alpha}{2}}}{n}}$$

$$\hat{P}_{верх} = \frac{\frac{m}{n} + \frac{\mu^2_{\frac{\alpha}{2}}}{2n} + \mu_{\frac{\alpha}{2}} \sqrt{\frac{m}{n^2}(1-\frac{m}{n}) + \frac{\mu^2_{\frac{\alpha}{2}}}{4n^2}}}{1 + \frac{\mu^2_{\frac{\alpha}{2}}}{n}}$$

Распределение Пуассона

Распределение Пуассона $P(\theta)$ имеет функцию распределения распределения:

$$P(\theta) = \frac{\theta^k}{k!} e^{-\theta}$$

И среднее и дисперсия равны θ

Доверительный интервал:

$$\overline{x_n} - \frac{\sqrt{\overline{x_n}}}{\sqrt{n}} \mu_{\frac{\alpha}{2}} \leq \theta \leq \overline{x_n} + \frac{\sqrt{\overline{x_n}}}{\sqrt{n}} \mu_{\frac{\alpha}{2}}$$

Выборку генерировать размера $n = 200$

Экспоненциальное распределение

Экспоненциальное распределение имеет имеет следующую плотность распределения

$$\lambda e^{-\lambda x}, x > 0$$

$$0, x \leq 0$$

Где $\lambda = \frac{1}{\nu}$, ν - экспоненциальный параметр scale в `scipy.stats.expon.rvs(scale, n)`.
Доверительный интервал

$$\frac{1}{\overline{x_n}} + \frac{\mu_{\frac{\alpha}{2}}}{\sqrt{n\overline{x_n}}} \leq \nu \leq \frac{1}{\overline{x_n}} - \frac{\mu_{\frac{\alpha}{2}}}{\sqrt{n\overline{x_n}}}$$

$$\frac{1}{\frac{1}{\overline{x}_n} - \frac{\mu_{\frac{\alpha}{2}}}{\sqrt{n\overline{x}_n}}} \leq \lambda \leq \frac{1}{\frac{1}{\overline{x}_n} + \frac{\mu_{\frac{\alpha}{2}}}{\sqrt{n\overline{x}_n}}}$$

Часть 2. Ядерные оценки

Ядро

Исходная выборка "портится" другим распределением
Функция $k(x)$ называется ядром, если

$$k(x) \geq 0$$

$$\int_{-\infty}^{\infty} k(x)dx = 1$$

$$k(-x) = k(x)$$

Ядра бывают разными: полукруглыми, треугольными, прямоугольными (равномерными) и др.

Нормальная ядерная функция

$$k(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

Прямоугольная ядерная функция

k(x) задаётся системой

$$\frac{1}{2}, |x| \leq 1$$

$$0, |x| > 1$$

Ядерный оценщик плотности

Ядерный оценщик плотности равен

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} k_h(x - x_j)$$

где

$$k_h(x) = \frac{1}{h} k(\frac{x}{h})$$

h называется шириной окна

С учётом k_h:

$$\hat{f}(x) = \frac{1}{hn} \sum_{j=1}^{n} k(\frac{x - x_j}{h})$$

Где k - дельтоорбразное `что-то там`, оно стремится к дельта функции при $h \to 0$

При $n \to \infty$ и $h \to 0$, h можно принять за $h \sim \frac{1}{\sqrt[5]{n}}$
h оптимизированная под распределение гаусса:

$$h_{onm} = \frac{1,05 S_n^2}{\sqrt[5]{n}}$$

где S_n - исправленная дисперсия

$$S^2_{n \, исправленная} = \frac{n}{n-1} \sum_{i=1}^{k} (x_i - \overline{x_n})^2 n_i$$

Оценка для нормального и равномерного распределений

Построить ядерную оценку для `N(3;5)` с h_{onm}, за $k(x)$ взять гауссовское и прямоугольное ядра

Построить ядерную оценку для `uniform[3;5]` с h_{onm}, за $k(x)$ взять гауссовское и прямоугольное ядра

Построить графики исходных плотностей и ядерных оценок

> Материалы по матстату
>
> - [Теория вероятностей и математическая статистика : учебное пособие Е. А. Трофимова, Н. В. Кисляк, Д. В. Гилёв]
> - [mathprofi: Статистические оценки параметров генеральной совокупности. Доверительный интервал и доверительная вероятность]