Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

Расчётно-графическая работа №2

По дисциплине «Математическая статистика»

Михайлов Дмитрий Андреевич
P3206
368530
Медведев Владислав Александрович
P3206
368508

Санкт-Петербург
2025 год

Содержание

Задача №1

Условие задачи.

Для выполнения первого упражнения будут весьма полезны знания об определении распределения хи-квадрат, Стьюдента и Фишера. Плюс теорема Фишера для выборок из нормального закона.

Предъявите доверительный интервал уровня $1-\alpha$ для указанного параметра при данных предположениях (**с математическими обоснованиями**). Сгенерируйте 2 выборки объёма 25 и посчитайте доверительный интервал. Повторить 1000 раз. Посчитайте, сколько раз 95-процентный доверительный интервал покрывает реальное значение параметра. То же самое сделайте для объёма выборки 10000. Как изменился результат? Как объяснить? Что изменяется при росте объёмов выборки?

Задача представлена в 4 вариантах. Везде даны две независимые выборки X_1, X_2 из нормальных распределений $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ объёмов n_1, n_2 соответственно. Сначала указывается оцениваемая функция, потом данные об остальных параметрах, затем параметры эксперимента и подсказки.

4. $\tau = \sigma_1^2/\sigma_2^2$ μ_1, μ_2; известны; $\mu_1 = 0, \mu_2 = 0, \sigma_1^2 = 2, \sigma_2^2 = 1$; воспользуйтесь функцией $\frac{n_2 \sum\limits_{i=1}^{n}(X_{1,i}-\mu_1)^2}{n_1 \sum\limits_{j=1}^{m}(X_{2,i}-\mu_2)^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$

Решение.

Исходя из условия, что $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, можно понять, что оцениваемый параметр: $\tau = \frac{\sigma_1^2}{\sigma_2^2} = 2$. При этом $\mu_1 = 0, \mu_2 = 0$. Воспользуемся функцией и найдём статистику для оценки τ:

$$T = \frac{n_2 \sum_{i=1}^{n_1} X_{1,i}^2}{n_1 \sum_{j=1}^{n_2} X_{2,j}^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

Так как $\sum_{i=1}^{n_1} X_{1,i}^2 \sim \sigma_1^2 \cdot \chi^2(n_1)$, $\sum_{j=1}^{n_2} X_{2,j}^2 \sim \sigma_2^2 \cdot \chi^2(n_2)$, в конечном итоге получаем, что статистика для оценки τ:

$$T = \frac{n_2 \cdot \sigma_1^2 \cdot \chi^2(n_1)}{n_1 \cdot \sigma_2^2 \cdot \chi^2(n_2)} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

Упростив выражение, получаем:

$$T = \frac{n_2 \chi^2(n_1)}{n_1 \chi^2(n_2)}$$

Полученная формула сводится к функции распределения Фишера со степенями свободы n_1, n_2:

$$F(n_1, n_2) \sim \frac{n_2 \chi^2(n_1)}{n_1 \chi^2(n_2)}$$

После этого распишем доверительный интервал уровня $1-\alpha$ для функции распределения:

$$P\left(F_{\alpha/2}(n_1, n_2) \leq \frac{n_2 \chi^2(n_1)}{n_1 \chi^2(n_2)} \leq F_{1-\alpha/2}(n_1, n_2)\right) = 1 - \alpha$$

Подставим выведенную формулу для статистики и получим интервал для параметра τ:

$$\left[\frac{\sum X_{2,j}^2/n_2}{\sum X_{1,i}^2/n_1} \cdot F_{\alpha/2}(n_2, n_1), \frac{\sum X_{2,j}^2/n_2}{\sum X_{1,i}^2/n_1} \cdot F_{1-\alpha/2}(n_2, n_1)\right].$$

После генерации выборок малого и большого объёма вычислим $\hat{\tau}$ с истинностью 2 и построим доверительный интервал с $\alpha = 0.05$. Далее приведён код для этого эксперимента.

Задача №2

Условие задачи.

В данном упражнении для построения асимптотического доверительного интервала весьма полезным будет применение ЦПТ к самой выборке или некоторому преобразованию выборки (например, к X^2).

Постройте асимптотический доверительный интервал уровня $1-\alpha$ для указанного параметра. Проведите эксперимент по схеме, аналогичной первой задаче.

Задача представлена в 5 вариантах. Сначала указывается класс распределений (однопараметрический), затем параметры эксперимента и подсказки.

3. $U[-\theta, \theta]$; θ^2; $\theta = 5$

Решение.

Рассмотрим случайную величину $X \sim U[-\theta, \theta]$, где $\theta > 0$. Это симметричное равномерное распределение, у которого:

1. $\mathbb{E}[X] = 0$
2. $\mathbb{D}[X] = \frac{\theta^2}{3}$

Переходя к оцениванию параметра, из формулы дисперсии следует:

$$\theta^2 = 3 \cdot \mathbb{D}[X]$$

Поскольку в реальности у нас есть только выборка $X_1, X_2, \ldots, X_n$, оценим дисперсию с помощью **выборочной дисперсии**:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Тогда оценка параметра:

$$\widehat{\theta^2} = 3 \cdot S^2$$

$$\widehat{\theta^2} = 3 \cdot \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Переходя к оцениванию асимптотического доверительного интервала, Выборочная дисперсия S^2 — **асимптотически нормальная** оценка, следовательно, с увеличением размера выборки можно применить **центральную предельную теорему** к $\widehat{\theta^2}$:

$$\sqrt{n} \cdot \left(\widehat{\theta^2} - \theta^2\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),$$

где $\sigma^2 = \mathbb{D}[\widehat{\theta^2}] = \mathbb{D}[3S^2] = 9 \cdot \mathbb{D}[S^2]$

Найдем $\mathbb{D}[S^2]$:

$$\mathbb{D}[S^2] = \frac{1}{n}\left(\mu_4 - \sigma^4\right),$$

где $\mu_4 = \mathbb{E}\left[(X - \mathbb{E}[X])^4\right] = \frac{\theta^4}{5}$, $\sigma^2 = \mathbb{D}[X] = \frac{\theta^2}{3} \longrightarrow \sigma^4 = \frac{\theta^4}{9}$

Подставим:

$$\mathbb{D}[S^2] = \frac{1}{n}\left(\frac{\theta^4}{5} - \frac{\theta^4}{9}\right) = \frac{\theta^4}{n}\left(\frac{4}{45}\right)$$

Значит:

$$\mathbb{D}[\widehat{\theta^2}] = 9 \cdot \mathbb{D}[S^2] = 9 \cdot \frac{\theta^4}{n} \cdot \frac{4}{45} = \frac{4\theta^4}{5n}$$

Переходя к применению центральной предельной теоремы (ЦПТ), для оценки параметра $\theta^2 = 3 \cdot \mathbb{D}[X]$, мы используем выборочную дисперсию S^2, а значит, рассматриваем оценку:

$$\widehat{\theta^2} = 3S^2$$

$$\sqrt{n} \cdot (\widehat{\theta^2} - \theta^2) = \sqrt{n}(3S^2 - 3\sigma^2) \xrightarrow{d} \mathcal{N}(0, 9 \cdot \mathbb{D}[S^2])$$

$$\sqrt{n} \cdot \left(\widehat{\theta^2} - \theta^2\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{4\theta^4}{5}\right)$$

Из этой сходимости вытекает доверительный интервал, если $\sqrt{n}(\widehat{\theta^2} - \theta^2)$ приближённо распределена нормально, то мы можем записать:

$$\frac{\widehat{\theta^2} - \theta^2}{\sqrt{9 \cdot \mathbb{D}[S^2]/n}} \sim \mathcal{N}(0, 1) \quad \text{(приближённо)}$$

$$P\left(-z_{1-\alpha/2} < \frac{\widehat{\theta^2} - \theta^2}{\sqrt{9 \cdot \mathbb{D}[S^2]/n}} < z_{1-\alpha/2}\right) \approx 1 - \alpha$$

$$P\left(\widehat{\theta^2} - z_{1-\alpha/2} \cdot \frac{\sqrt{9 \cdot \mathbb{D}[S^2]}}{\sqrt{n}} < \theta^2 < \widehat{\theta^2} + z_{1-\alpha/2} \cdot \frac{\sqrt{9 \cdot \mathbb{D}[S^2]}}{\sqrt{n}}\right) \approx 1 - \alpha$$

$$\theta^2 \in \left[\widehat{\theta^2} \pm z_{1-\alpha/2} \cdot \frac{\sqrt{9 \cdot \mathbb{D}[S^2]}}{\sqrt{n}}\right]$$

Далее написан код на языке Python.

Выводы

1. При очень маленьком объеме выборки (n=10) интервалы значительно шире, что сопровождается большим разбросом оценок. Уровень доверия составил 78.40%, что заметно ниже теоретического 95%. Это подчёркивает, как малый размер выборки снижает точность оценок.
2. При большом объеме выборки (n=100000) интервалы становятся существенно уже, а уровень доверия 94.20%, близкий к теоретическому значению 95%, демонстрирует, как с ростом выборки качество оценок улучшается и доверительные интервалы сужаются.
3. Результаты подтверждают, что с увеличением размера выборки качество статистических оценок значительно улучшается, и интервалы становятся уже, что согласуется с предсказаниями центральной предельной теоремы.

Приложения

Задача №1

Ссылка на исходник с кодом программы, решающей эту задачу на языке Python. [1]

Задача №2

Ссылка на исходник с кодом программы, решающей эту задачу на языке Python. [2]

Список использованных источников

[1] Задача №1. *URL*: Исходник с кодом, решающий задачу №1.

[2] Задача №2. *URL*: Исходник с кодом, решающий задачу №2.