Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №8
по дисциплине
«Математическая статистика»

Выполнил студент:
Самутичев Евгений Романович
группа: 3630102/70201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

Список таблиц

1 Постановка задачи

Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону $N(0, 1)$, для параметров положения и масштаба построить асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия и классические интервальные оценки на основе статистик χ^2 и Стьюдента. В качестве доверительной вероятности взять $\gamma = 0.95$.

2 Теория

2.1 Интервальное оценивание

Интервальной оценкой (или *доверительным интервалом*) числовой характеристики или параметра распределения θ генеральной совокупности с доверительной вероятностью γ называется интервал (θ_1, θ_2), границы которого являются случайными функциями: $\theta_1 = \theta_1(x_1, ..., x_n), \theta_2 = \theta_2(x_1, ..., x_n)$, который накрывает θ с вероятностью γ:

$$\mathbf{P}(\theta_1 < \theta < \theta_2) = \gamma \tag{1}$$

Часто вместо доверительной вероятности γ рассматривается *уровень значимости* $\alpha = 1 - \gamma$. Важной характеристикой данной интервальной оценки является половина длины доверительного интервала, она называется *точностью* интервального оценивания

$$\Delta = \frac{\theta_2 - \theta_1}{2} \tag{2}$$

Рассмотрим общий метод построения интервальных оценок [1, стр. 456- – 457]. Пусть известна статистика $Y(\widehat{\theta}, \theta)$, содержащая оцениваемый параметр θ и его точечную оценку $\widehat{\theta}$ со следующими свойствами:

- Функция распределения $F_Y(x)$ известна и не зависит от θ
- Функция $Y(\widehat{\theta}, \theta)$ непрерывна и строго монотонна (для определенности - строго возрастает) по θ

которые мы будем проверять при построении интервальных оценок нормального распределения. Зададим уровень значимости α и будем строить доверительный интервал так чтобы $(-\infty, \alpha_1), (\alpha_2, +\infty)$ накрывали θ с вероятностью $\frac{\alpha}{2}$.

Пусть $y_{\alpha/2}, y_{1-\alpha/2}$ – квантили распределения Y соотв. порядков, тогда

$$\begin{aligned}\mathbf{P}\left(y_{\alpha/2} < Y(\widehat{\theta}, \theta) < y_{1-\alpha/2}\right) = F_Y\left(y_{1-\alpha/2}\right) - F_Y\left(y_{\alpha/2}\right) = \\ = 1 - \alpha/2 - \alpha/2 = 1 - \alpha = \gamma\end{aligned} \tag{3}$$

Т.к. $Y(\widehat{\theta}, \theta)$ – строго возрастает по θ, то у неё есть обратная функция $Y^{-1}(y)$ относительно θ и она также строго возрастает, а значит:

$$\begin{gathered} y_{\alpha/2} < Y(\widehat{\theta}, \theta) < y_{1-\alpha/2} \\ Y^{-1}(y_{\alpha/2}) < \theta < Y^{-1}(y_{1-\alpha/2}) \end{gathered} \tag{4}$$

итого $\theta_1 = Y^{-1}(y_{\alpha/2})$ и $\theta_2 = Y^{-1}(y_{1-\alpha/2})$ – мы построили границы интервала. Применим это для построения интервальных оценок нормального распределения по выборке $(x_1, ..., x_n)$.

2.2 Классические оценки

2.2.1 Для математического ожидания m

Доказано что случайная величина $T = \sqrt{n-1} \cdot \frac{\bar{x}-m}{s}$ называемая статистикой Стьюдента, распределена по закону Стьюдента с $n-1$ степенями свободы, применяя с некоторыми деталями [1, стр. 457 – 458] выкладки, получаем оценки границ интервала:

$$\begin{gathered} m_1 = \bar{x} - \frac{xt_{1-\alpha/2}(n-1)}{\sqrt{n-1}} \\ m_2 = \bar{x} + \frac{xt_{1-\alpha/2}(n-1)}{\sqrt{n-1}} \end{gathered} \tag{5}$$

, где $t_{1-\alpha/2}(n-1)$ – квантиль порядка $1-\alpha/2$ распределения Стьюдента с $n-1$ степенями свободы.

2.2.2 Для среднего квадратичного отклонения σ

Доказано что случайная величина ns^2/σ^2 распределена по закону χ^2 с $n-1$ степенями свободы. Применяя общий метод построения интервальных оценок получаем оценки границ интервала:

$$\begin{aligned} \sigma_1 &= \frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} \\ \sigma_2 &= \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}} \end{aligned} \tag{6}$$

, где $\chi^2_{1-\alpha/2}(n-1), \chi^2_{\alpha/2}(n-1)$ - квантили соотв. порядков χ^2-распределения с $n-1$ степенями свободы.

2.3 Асимптотически нормальные оценки

2.3.1 Для математического ожидания m

В силу центральной предельной теоремы центрированная и нормированная случайная величина $\sqrt{n}(\bar{x}-m)/\sigma$ распределена приблизительно нормально с параметрами 0 и 1. Исходя из этого [1, стр. 460] получаем оценку:

$$\begin{aligned} m_1 &= \bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} \\ m_2 &= \bar{x} + \frac{su_{1-\alpha/2}}{\sqrt{n}} \end{aligned} \tag{7}$$

, где $u_{1-\alpha/2}$ - квантиль нормального распределения $N(0,1)$ порядка $1-\alpha/2$

2.3.2 Для среднего квадратичного отклонения σ

Аналогично, в силу центральной предельной теоремы центрированная и нормированная случайная величина $(s^2-\mathbf{M}s^2)/\sqrt{\mathbf{D}s^2}$ при большом объеме выборки n распределена приблизительно нормально с параметрами 0 и 1. Исходя из этого [1, стр. 461] получаем оценку:

$$\begin{aligned} \sigma_1 &= s\left(1+u_{1-\alpha/2}\sqrt{(e+2)/n}\right)^{-1/2} \\ \sigma_2 &= s\left(1-u_{1-\alpha/2}\sqrt{(e+2)/n}\right)^{-1/2} \end{aligned} \tag{8}$$

, где e - выборочный эксцесс, определяемый как

$$e = \frac{m_4}{s^4} - 3 \tag{9}$$

, где m_4 - четвертый выборочный центральный момент, определяемый как

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4 \tag{10}$$

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **SciPy** - модуль **stats** для генерации данных, вычисления квантилей различных распределений для построения интервальных оценок

Исходный код работы приведен в приложении.

4 Результаты

4.1 Классические оценки

	m(5)	σ(6)
$n = 20$	$-0.6 < m < 0.27$	$0.71 < \sigma < 1.36$
$n = 100$	$-0.04 < m < 0.34$	$0.84 < \sigma < 1.12$

4.2 Асимптотически нормальные оценки

	m(7)	σ(8)
$n = 20$	$-0.56 < m < 0.23$	$0.71 < \sigma < 1.46$
$n = 100$	$-0.03 < m < 0.34$	$0.84 < \sigma < 1.13$

5 Обсуждение

Полученные интервальные оценки говорят о том что с вероятностью 0.95 значения $m = 0$ и $\sigma = 1$ лежат в соответствующих интервалах. По постановке эксперимента, интервалы действительно накрывают истинные значения параметров. Следует заметить что при большом объеме n выборки - асимптотические оценки практически совпадают с классическими.

6 Приложения

1. Исходный код лабораторной `https://github.com/zhenyatos/statlabs/tree/master/Lab8`

Список литературы

[1] **Вероятностные разделы математики.** Учебник для бакалавров технических направлений. // Под ред. Максимова Ю.Д. - СПб «Иван Федоров», 2001. - 592 с., илл