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 группы 3630102/70401

Мельникова Анна Николаевна

Проверил
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020

Содержание

Список таблиц

Список иллюстраций

1 Постановка задачи

Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону $N(x, 0, 1)$, для параметров положения и масштаба построить:

- Асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия;
- Классические интервальные оценки на основе статистик χ^2 и Стьюдента.

В качестве параметра надёжности взять $\gamma = 0.95$.

2 Теория

2.1 Доверительные интервалы для параметров нормального распределения

2.1.1 Доверительный интервал для математического ожидания нормального распределения

Дана выборка $(x_1, x_2, ..., x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочное среднее $\bar{x}$ и выборочное среднее квадратическое отклонение s. Параметры m и σ нормального распределения неизвестны.
Доказано, что случайная величина

$$T = \sqrt{n-1} \cdot \frac{\bar{x} - m}{s}, \tag{1}$$

называемая статистикой Стьюдента, распределена по закону Стьюдента с $n-1$ степенями свободы. Пусть $f_T(x)$ — плотность вероятности распределения случайной величины (1). Тогда

$$P\left(-x < \sqrt{n-1} \cdot \frac{\bar{x} - m}{s} < x\right) = P\left(-x < \sqrt{n-1} \cdot \frac{m - \bar{x}}{s} < x\right) =$$
$$= \int_{-x}^{x} f_T(t)dt = 2\int_0^x f_T(t)dt = 2\left(\int_{-\infty}^{x} f_T(t)dt - \frac{1}{2}\right) = 2F_T(x) - 1$$

Здесь $F_T(x)$ — функция распределения Стьюдента с $n-1$ степенями свободы.
Полагаем $2F_T(x) - 1 = 1 - \alpha$, где α — выбранный уровень значимости. Тогда $F_T(x) = 1 - \alpha/2$. Пусть $t_{1-\alpha/2}(n-1)$ — квантиль распределения Стьюдента с $n-1$ степенями свободы и порядка $1 - \alpha/2$. Из предыдущих равенств мы получаем

$$\begin{gathered} P\left(\bar{x} - \frac{sx}{\sqrt{n-1}} < m < \bar{x} + \frac{sx}{\sqrt{n-1}}\right) = 2F_T(x) - 1 = 1 - \alpha, \\ P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \bar{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1 - \alpha. \end{gathered} \tag{2}$$

Выражение (2) и даёт доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$ [1, с. 457-458].

2.1.2 Доверительный интервал для среднего квадратического отклонения нормального распределения

Дана выборка $(x_1, x_2, ..., x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочную дисперсию s^2. Параметры m и σ нормального распределения неизвестны.
Доказано, что случайная величина ns^2/σ^2 распределена по закону χ^2 с $n-1$ степенями свободы.
Задаёмся уровнем значимости α и находим квантили $\chi^2_{\alpha/2}(n-1)$ и $\chi^2_{1-\alpha/2}(n-1)$.
Это значит, что

$$P\left(\chi^2(n-1) < \chi^2_{\alpha/2}(n-1)\right) = \alpha/2,$$
$$P\left(\chi^2(n-1) < \chi^2_{1-\alpha/2}(n-1)\right) = 1-\alpha/2.$$

Тогда

$$\begin{aligned} &P\left(\chi^2_{\alpha/2}(n-1) < \chi^2(n-1) < \chi^2_{1-\alpha/2}(n-1)\right) = \\ &\qquad P\left(\chi^2(n-1) < \chi^2_{1-\alpha/2}(n-1)\right) - P\left(\chi^2(n-1) < \chi^2_{\alpha/2}(n-1)\right) = \\ &\qquad\qquad = 1-\alpha/2-\alpha/2 = 1-\alpha. \end{aligned}$$

Отсюда

$$\begin{aligned} P\left(\chi^2_{\alpha/2}(n-1) < \frac{ns^2}{\sigma^2} < \chi^2_{1-\alpha/2}(n-1)\right) &= P\left(\frac{1}{\chi^2_{1-\alpha/2}(n-1)} < \frac{\sigma^2}{ns^2} < \frac{1}{\chi^2_{\alpha/2}(n-1)}\right) = \\ &= P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1-\alpha. \end{aligned}$$

Окончательно

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1-\alpha. \tag{3}$$

Выражение (3) и даёт доверительный интервал для σ с доверительной вероятностью (параметром надёжности) $\gamma = 1-\alpha$ [1, с. 458-459].

2.2 Доверительные интервалы для математического ожидания и среднего квадратического отклонения произвольного распределения при большом объёме выборки. Асимптотический подход

При большом объёме выборки для построения доверительных интервалов может быть использован асимптотический метод на основе центральной предельной теоремы.

2.2.1 Доверительный интервал для математического ожидания произвольной генеральной совокупности при большом объёме выборки

Выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ при большом объёме выборки является суммой большого числа взаимно независимых одинаково распределённых случайных величин. Предполагаем, что исследуемое генеральное распределение имеет конечные математическое ожидание m и дисперсию σ^2.
Тогда в силу центральной предельной теоремы (ЦПТ) центрированная и нормированная случайная величина $(\bar{x} - M\bar{x})/\sqrt{D\bar{x}} = \sqrt{n}\cdot(\bar{x} - m)/\sigma$ распределена приблизительно нормально с параметрами 0 и 1.
Пусть функция Лапласа –

$$\Phi(x) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-t^2/2}dt. \tag{4}$$

Тогда

$$P\left(-x < \sqrt{n}\frac{\bar{x} - m}{\sigma} < x\right) = P\left(-x < \sqrt{n}\frac{m - \bar{x}}{\sigma} < x\right) \approx$$
$$\approx \Phi(x) - \Phi(-x) = \Phi(x) - [1 - \Phi(x)] = 2\Phi(x) - 1.$$

Отсюда

$$P\left(\bar{x} - \frac{\sigma x}{\sqrt{n}} < m < \bar{x} - \frac{\sigma x}{\sqrt{n}}\right) \approx 2\Phi(x) - 1. \tag{5}$$

Полагаем $2\Phi(x) - 1 = \gamma = 1 - \alpha$; тогда $\Phi(x) = 1 - \alpha/2$. Пусть $u_{1-\alpha/2}$ — квантиль нормального распределения $N(0, 1)$ порядка $1 - \alpha/2$. Заменяя в равенстве (5) σ на s, запишем его в виде

$$P\left(\bar{x} - \frac{s u_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} - \frac{s u_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma. \tag{6}$$

Выражение (6) и даёт доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$ [1, с. 460].

2.2.2 Доверительный интервал для среднего квадратического отклонения произвольной генеральной совокупности при большом объёме выборки

Выборочная дисперсия $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$ при большом объёме выборки является суммой большого числа практически взаимно независимых случайных величин (имеется одна связь $\sum_{i=1}^{n} x_i = n\bar{x}$, которой при большом n можно пренебречь). Предполагаем, что исследуемая генеральная совокупность имеет конечные первые четыре момента.
В силу центральной предельной теоремы центрированная и нормированная случайная величина $(s^2 - Ms^2)/\sqrt{Ds^2}$ при большом объёме выборки n распределена приблизительно нормально с параметрами 0 и 1. Пусть $\Phi(x)$ — функция Лапласа (4). Тогда

$$P\left(-x < \frac{s^2 - Ms^2}{\sqrt{Ds^2}} < x\right) \approx \Phi(x) - \Phi(-x) = \Phi(x) - [1 - \Phi(x)] = 2\Phi(x) - 1$$

Положим $2\Phi(x) - 1 = \gamma = 1 - \alpha$. Тогда $\Phi(x) = 1 - \alpha/2$. Пусть $u_{1-\alpha/2}$ — корень этого уравнения — квантиль нормального распределения $N(0, 1)$ порядка $1 - \alpha/2$.

Известно, что $Ms^2 = \sigma^2 - \frac{\sigma^2}{n} \approx \sigma^2$ и $Ds^2 = \frac{\mu_4 - \mu_2^2}{n} + o(\frac{1}{n}) \approx \frac{\mu_4 - \mu_2^2}{n}$. Здесь μ_k — центральный момент k-го порядка генерального распределения; $\mu_2 = \sigma^2; \mu_4 = M[(x - Mx)^4]$; $o(\frac{1}{n})$— бесконечно малая высшего порядка, чем $1/n$, при $n \to \infty$.
Итак,

$$Ds^2 \approx \frac{\mu_4 - \mu_2^2}{n}.$$

Отсюда

$$Ds^2 \approx \frac{\sigma^4}{n}(\frac{\mu_4}{\sigma^4} - 1) = \frac{\sigma^4}{n}((\frac{\mu_4}{\sigma^4} - 3) + 2) = \frac{\sigma^4}{n}(E + 2) \approx \frac{\sigma^4}{n}(e + 2),$$

где $E = \frac{\mu_4}{\sigma^4} - 3$ — эксцесс генерального распределения, $e = \frac{m_4}{s^4} - 3$ — выборочный эксцесс; $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$ — четвёртый выборочный центральный момент.
Далее,

$$\sqrt{Ds^2} \approx \frac{\sigma^2}{\sqrt{n}}\sqrt{e + 2}$$

Преобразуем неравенства, стоящие под знаком вероятности в формуле
$P\left(-x < \frac{s^2 - Ms^2}{\sqrt{Ds^2}} < x\right) = \gamma$:

$$\begin{aligned} -\sigma^2 U < s^2 - \sigma^2 < \sigma^2 U; \\ \sigma^2(1 - U) < s^2 < \sigma^2(1 + U); \\ 1/[\sigma^2(1 + U)] < 1/s^2 < 1/[\sigma^2(1 - U)]; \\ s^2/(1 + U) < \sigma^2 < s^2/(1 - U); \\ s(1 + U)^{-1/2} < \sigma < s(1 - U)^{-1/2}, \end{aligned} \tag{7}$$

где $U = u_{1-\alpha/2}\sqrt{(e + 2)/n}$ или
$s(1 + u_{1-\alpha/2}\sqrt{(e + 2)/n})^{-1/2} < \sigma < s(1 - u_{1-\alpha/2}\sqrt{(e + 2)/n})^{-1/2}$.
Разлагая функции в биномиальный ряд и оставляя первые два члена, получим

$$s(1 - 0.5U) < \sigma < s(1 + 0.5U)$$

или

$$s(1 - 0.5u_{1-\alpha/2}\sqrt{(e + 2)/n}) < \sigma < s(1 + 0.5u_{1-\alpha/2}\sqrt{(e + 2)/n}). \tag{8}$$

Формулы (7) или (8) дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$ [1, с. 461-462].
Замечание. Вычисления по формуле (7) дают более надёжный результат, так как в ней меньше грубых приближений.

3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm (для визуализации использовался пакет matplotlib, для вычислений - numpy). Исходный код лабораторной работы и LaTeX-файлы отчета приведёны в приложении в виде ссылки на репозиторий GitHub.

4 Результаты и описание выполненной работы

4.1 Доверительные интервалы для параметров нормального распределения

n = 20	m	σ
	-0.35 < m < 0.42	0.63 < σ < 1.2
n = 100	m	σ
	-0.12 < m < 0.3	0.92 < σ < 1.22

Таблица 1: Доверительные интервалы для параметров нормального распределения

4.2 Доверительные интервалы для параметров нормального распределения. Асимптотический подход

n = 20	m	σ
	-0.32 < m < 0.38	0.65 < σ < 1.14
n = 100	m	σ
	-0.12 < m < 0.29	0.93 < σ < 1.23

Таблица 2: Доверительные интервалы для параметров нормального распределения. Асимптотический подход

5 Обсуждение

По сгенерированной согласно стандартному нормальному закону $N(x, 0, 1)$ выборке мы нашли интервалы, в которые с вероятностью 0.95 попадут параметры закона, описывающего выборку.
То есть, имея выборку и зная про нее лишь то, что закон, по которому она сгенерирована, нормальный, смогли найти интервалы, в которых с заданной вероятностью лежат параметры неизвестного нам распределения. Видим, что известные нам в этом эксперименте $m = 0, \sigma = 1$ действительно лежат в соответствующих найденных доверительных интервалах.
Заметим так же, что асимптотические оценки при увеличении мощности выборки приюлижаются к классическим.

6 Приложения

Код программы - GitHub URL: https://github.com/kaustika/Statistics2020

Список литературы

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