

Санкт-Петербургский Политехнический университет
Петра Великого
Институт Прикладной Математики и Механики
Кафедра «Прикладная Математика и Информатика»

Отчет
По лабораторной работе № 8
По Дисциплине «Математическая статистика»

Выполнил:
Студент Селянкин Федор
Группа 3630102/70301
Проверил:
к.ф. – м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

Содержание	2
Постановка задачи	3
Теория.....	3
Доверительные интервалы для параметров нормального распределения.....	3
Доверительный интервал для математического ожидания m нормального распределения	3
Доверительный интервал для среднего квадратического отклонения σ нормального распределения	3
Доверительные интервалы для математического ожидания m и среднего квадратического отклонения σ произвольного распределения при большом объёме выборки. Асимптотический подход	4
Доверительный интервал для математического ожидания m произвольной генеральной совокупности при большом объёме выборки	4
Доверительный интервал для среднего квадратического отклонения σ произвольной генеральной совокупности при большом объёме выборки	5
Реализация.....	6
Результаты.....	6
Доверительные интервалы для параметров нормального распределения.....	6
Доверительные интервалы для параметров произвольного распределения. Асимптотический подход	6
Обсуждения	6
Литература	7
Список Таблиц:	
Таблица 1 Доверительные интервалы для параметров нормального распределения	6
Таблица 2 Доверительные интервалы для параметров нормального распределения. Асимптотический подход	6

Постановка задачи

Задание 8. Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону $N(x, 0, 1)$ для параметров положения 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 надёжности взять $\gamma = 0.95$.

Теория

Доверительные интервалы для параметров нормального распределения

Доверительный интервал для математического ожидания m нормального распределения. Дана выборка (x_1, x_2, \dots, x_n) объема n из нормальной генеральной совокупности. На её основе строим выборочное среднее \bar{x} и выборочное среднее квадратичное отклонение s . Параметры m и σ нормального распределения неизвестны.

Доказано, что случайная величина

$$T = \sqrt{n-1} * \frac{\bar{x} - m}{s}$$

(1)

называется статистикой Стьюдента, распределена по закону Стьюдента с $n - 1$ степенями свободы. Пусть $f_T(x)$ – плотность вероятности этого распределения. Тогда

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

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

(2)

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

что и даёт доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$.

Доверительный интервал для среднего квадратического отклонения σ нормального распределения

Дана выборка (x_1, x_2, \dots, x_n) объема n из нормальной генеральной совокупности. На её основе строим выборочное среднее \bar{x} и 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, например, с помощью средств языка программирования python (функции `chi2.ppf` из пакета `scipy.stats`) 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

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

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

Окончательно

(3)

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

что и даёт доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$.

Доверительные интервалы для математического ожидания m и среднего квадратического отклонения σ 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.

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

Выборочное среднее $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$ 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 центрированная и нормальная случайная величина $(\bar{x} - M\bar{x})/\sqrt{D\bar{x}} = \sqrt{n}(\bar{x} - m)/\sigma$ распределена приблизительно нормально с параметрами 0 и 1. Пусть

(4)

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

-функция Лапласа. Тогда

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

Отсюда

(5)

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

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

(6)

$$P\left(\bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} + \frac{su_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma$$

что и даёт доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$.

Доверительный интервал для среднего квадратического отклонения σ 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 $2\Phi(x) - 1 = \gamma - \alpha$. 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 $Ds^2 = \frac{\mu_4 - \mu_2^2}{n} + o\left(\frac{1}{n}\right) \approx \frac{\mu_4 - \mu_2^2}{n}$. Здесь μ_k – центральный момент k -го порядка генерального распределения; $\mu_2 = \sigma^2$; $\mu_4 = M[(x - Mx)^4]$; $o\left(\frac{1}{n}\right)$ – бесконечно малая высшего порядка, чем $1/n$, при $n \rightarrow \infty$. Итак, $Ds^2 \approx \frac{\mu_4 - \sigma^2}{n}$. Отсюда

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

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

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

Преобразуем неравенства, стоящие под знаком вероятности в формуле

$$P\left(-x < \frac{s^2 - Ms^2}{\sqrt{Ds^2}} < x\right) = \gamma:$$

$$-\sigma^2 U < s^2 - \sigma^2 < \sigma^2 U;$$

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

$$s(1+U)^{-1/2} < \sigma < s(1-U)^{-1/2},$$

где $U = u_{1-\alpha/2}\sqrt{(e+2)/n}$ или

$$s\left(1 + u_{1-\alpha/2}\sqrt{(e+2)/n}\right)^{-1/2} < \sigma < s\left(1 - u_{1-\alpha/2}\sqrt{(e+2)/n}\right)^{-1/2}.$$

Разлагая функции в биномиальный ряд и оставляя первые два члена, получим

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

(8)

Или

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

Формулы (7) и (8) дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$.

Замечание. Вычисления по формуле (7) дают более надёжный результат, так как в ней меньше грубых приближений.

Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm, с использованием дополнительных библиотек для отображения и расчетов. Исходный код лабораторной выложен на веб-сервисе GitHub [1].

Результаты

Доверительные интервалы для параметров нормального распределения

$n = 20$	m	σ
	$-0.49 < m < 0.31$	$0.66 < \sigma < 1.26$
$n = 100$	m	σ
	$-0.14 < m < 0.29$	$0.95 < \sigma < 1.25$

Таблица 1 Доверительные интервалы для параметров нормального распределения

Доверительные интервалы для параметров произвольного распределения.

Асимптотический подход.

$n = 20$	m	σ
	$-0.53 < m < 0.35$	$0.69 < \sigma < 1.17$
$n = 100$	m	σ
	$-0.12 < m < 0.27$	$0.96 < \sigma < 1.23$

Таблица 2 Доверительные интервалы для параметров нормального распределения.
Асимптотический подход

Обсуждения

Доверительные интервалы для параметров распределения

Генеральные характеристики накрываются построенными доверительными интервалами.

Доверительные интервалы, полученные по большей выборке, являются соответственно более точными, т.е. меньшими по длине.

Асимптотический подход на большем объеме выборки даёт результат точнее.

Литература

1. Ссылка на репозиторий GitHub
<https://github.com/SelyankinFyodor/math-statistics>
2. Вероятностные разделы математики. Учебник для бакалавров технических направлений.
//Под ред. Максимова Ю.Д. – Спб.: .Иван Федоров., 2001. – 592 с.Максимов Ю.Д.
Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей : учеб. пособие Ю.Д. Максимов; под ред. В.И. Антонова. — Спб. : Изд-во Политехн. ун-та, 2009. — 395 с. (Математика в политехническом университете).
3. Максимов Ю.Д. Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей: учеб. пособие / Ю. Д. Максимов; под ред. В. И. Антонова. -СПб. : Изд-во Политехн. Ун-та, 2009. -395 с.(Математика в политехническом университете).