Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Кафедра прикладной математики и информатики

Математическая статистика

Отчет по лабораторной работе №5

Выполнил студент гр. 5030102/20202 Тишковец С.Е.

Преподаватель Баженов А.Н.

Санкт-Петербург

2025

Оглавление

1. Постановка задачи

Сгенерировать выборку объёмом 100 элементов для нормального распределения N(x,0,1). По сгенерированной выборке оценить параметры μ и σ нормального закона методом максимального правдоподобия. В качестве основной гипотезы H_0 будем считать, что сгенерированное распределение имеет вид $N(x, \hat{\mu}, \hat{\sigma})$.

Проверить основную гипотезу, используя критерий согласия χ^2. В качестве уровня значимости взять α = 0.05. Привести таблицу вычислений χ^2.

Исследовать точность (чувствительность) критерия χ^2 - сгенерировать выборки равномерного распределения и равномерного распределения объема 20 элементов. Проверить их на нормальность.

2. Теоретическая информация

2.1. Метод максимального правдоподобия

Одним из универсальных методов оценивания является метод максимального правдоподобия, предложенный Р.Фишером (1921).

Пусть $x_1, \dots, x_n$ — случайная выборка из генеральной совокупности с плотностью вероятности $f(x, \theta)$; $L(x_1, \dots, x_n, \theta)$ — функция правдоподобия (ФП), представляющая собой совместную плотность вероятности независимых с.в. $x_1, \dots, x_n$ и рассматриваемая как функция неизвестного параметра θ:

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta)$$

Оценкой максимального правдоподобия (о.м.п) будем называть такое значение $\widehat{\theta_{\text{мп}}}$ из множества допустимых значений параметра θ, для которого ФП принимает наибольшее значение при заданных $x_1, \dots, x_n$:

$$\widehat{\theta_{\text{мп}}} = \arg\max_{\theta} L(x_1, \dots, x_n, \theta)$$

Если ФП дважды дифференцируема, то её стационарные значения даются корнями уравнения

$$\frac{\partial L(x_1, \dots, x_n, \theta)}{\partial \theta} = 0$$

Достаточным условием того, чтобы некоторое стационарное значение $\tilde{\theta}$ было локальным максимумом, является неравенство

$$\frac{\partial^2 L}{\partial \theta^2}\left(x_1, \dots, x_n, \tilde{\theta}\right) < 0$$

Определив точки локальных максимумов ФП (если их несколько), находят наибольший, который и даёт решение задачи.

Часто проще искать максимум логарифма ФП, так как он имеет максимум в одной точке с ФП:

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta}, \text{если} L > 0$$

и соответственно решать уравнение

$$\frac{\partial \ln L}{\partial \theta} = 0$$

которое называют *уравнением правдоподобия*.

В задаче оценивания векторного параметра $\theta = (\theta_1, \dots, \theta_m)$ аналогично находится максимум ФП нескольких аргументов:

$$\widehat{\theta_{\text{мп}}} = \arg \max_{\theta_1, \theta_2 \dots \theta_m} L\,(x_1, x_2, \dots x_n, \theta_1, \theta_2, \dots \theta_m)$$

и в случае дифференцируемости ФП выписывается система уравнений правдоподобия

$$\frac{\partial L}{\partial \theta_k} = 0 \ \text{ или } \ \frac{\partial \ln L}{\partial \theta_k} = 0, k = 1, \dots m$$

2.2. Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

Исчерпывающей характеристикой изучаемой случайной величины является её закон распределения. Поэтому естественно стремление исследователей построить этот закон приближённо на основе статистических данных. Сначала выдвигается гипотеза о виде закона распределения. После того как выбран вид закона, возникает задача оценивания его параметров и проверки (тестирования) закона в целом.

Для проверки гипотезы о законе распределения применяются критерии согласия. Таких критериев существует много. Мы рассмотрим наиболее обоснованный и наиболее часто используемый в практике — критерий χ^2 (хи-квадрат), введённый К. Пирсоном (1900 г.) для случая, когда параметры распределения известны. Этот критерий был существенно уточнён Р.

Фишером (1924 г.), когда параметры распределения оцениваются по выборке, используемой для проверки.

Мы ограничимся рассмотрением случая одномерного распределения.

Итак, выдвинута гипотеза H_0 о генеральном законе распределения с функцией распределения F(x)

Рассматриваем случай, когда гипотетическая функция распределения $F(x)$ не содержит неизвестных параметров.

Разобьём генеральную совокупность, т.е. множество значений изучаемой случайной величины X на k непересекающихся подмножеств $\Delta_1, \Delta_2, \dots, \Delta_k$.

Пусть $p_i = P(X \in \Delta_i), i = 1, \dots, k$

Если генеральная совокупность — вся вещественная ось, то подмножества $\Delta_i = (a_{i-1}, a_i]$ — полуоткрытые промежутки ($i = 2, \dots, k-1$). Крайние промежутки будут полубесконечными: $\Delta_1 = (-\infty, a_1]$,

$\Delta_k = (a_{k-1}, +\infty)$. В этом случае $p_i = F(a_i) - F(a_{i-1}); a_0 = -\infty, a_k = +\infty (i = 1, \dots, k)$

Отметим, что $\sum_{i=1}^{k} p_i = 1$. Будем предполагать, что все $p_i > 0 (i = 1, \dots, k)$

Пусть, далее, $n_1, n_2, \dots, n_k$ — частоты попадания выборочных элементов в подмножества $\Delta_1, \Delta_2, \dots, \Delta_k$. соответственно.

В случае справедливости гипотезы H_0 относительные частоты $\frac{n_i}{n}$ при большом n должны быть близки к вероятностям $p_i (i = 1, \dots, k)$, поэтому за меру отклонения выборочного распределения от гипотетического с функцией F(x) естественно выбрать величину

$$Z = \sum_{i=1}^{k} c_i \left(\frac{n_i}{n} - p_i\right)^2$$

где c_i — какие-нибудь положительные числа (веса). К. Пирсоном в качестве весов выбраны числа $c_i = \frac{n}{p_i} (i = 1, \dots, k)$. Тогда получается статистика критерия хи-квадрат К. Пирсона

$$\chi^2 = \sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

которая обозначена тем же символом, что и закон распределения хи-квадрат.

Правило проверки гипотезы о законе распределения по методу:

1. Выбираем уровень значимости α
2. Находим квантиль ${\chi_{1-\alpha}}^2(k-1)$ распределения хи-квадрат с k-1 степенями свободы порядка $1-\alpha$.
3. С помощью гипотетической функции распределения F(x) вычисляем вероятности $p_i = P(X \in \Delta_i), i = 1, \dots, k$.
4. Вычисляем выборочное значение статистики критерия χ^2:

$$\chi_B{}^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

5. Сравниваем ${\chi_B}^2$ и квантиль ${\chi_{1-\alpha}}^2(k-1)$

- Если ${\chi_B}^2 < {\chi_{1-\alpha}}^2(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.
- Если ${\chi_B}^2 \geq {\chi_{1-\alpha}}^2(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

3. Результаты исследования

Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат.

Метод максимального правдоподобия: $\hat{\mu} \approx 0.01, \hat{\sigma} \approx 1.02$

Критерий согласия: χ^2

Кол-во промежутков: 8

Уровень значимости: $\alpha = 0.05$

Квантиль ${\chi_{1-\alpha}}^2(k-1) = {\chi_{0.95}}^2(7) \approx 14.07$

i	limits	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty;\ -1.1]$	10	0,13	13,57	-3,57	0,93
2	(-1,1; -0,73]	8	0,10	9,60	-1,60	0,26
3	(-0,73; -0,37]	17	0,13	12,53	4,47	1,59
4	(-0,37; 0,0]	18	0,14	14,31	3,69	0,95
5	(0,0; 0,37]	13	0,14	14,31	-1,31	0,11
6	(0,37; 0,73]	9	0,13	12,53	-3,53	0,99
7	(0,73; 1,1]	10	0,10	9,60	0,40	0,01
8	$(1.1; \infty)$	15	0,13	13,57	1,43	0,15
$\sum$		100	1,00	100,00	0	7,06

Таблица 1. Вычисление χ_B^2 при проверке гипотезы H_0 о законе распределения $U(x, \hat{\mu}, \hat{\sigma})$

Сравнивая $\chi_B^2 = 7{,}06$ и $\chi_{0.95}^2(7) \approx 14.07$, видим, что $\chi_B^2 < \chi_{0.95}^2(7)$

i	limits	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty;\ -1.1]$	1	0,14	2,714	-1,714	1,08
2	(-1,1; -0,37]	5	0,22	4,426	0,574	0,07
3	(-0,37; -0,37]	6	0,29	5,722	0,278	0,01
4	(0,37; 1,1]	4	0,22	4,426	-0,426	0,04
5	(1.1; ∞)	4	0,14	2,714	1,286	0,61
Σ		20	1,00	20	-0,002	3,68

Таблица 1. Вычисление χ_B^2 при проверке гипотезы H_0 о законе распределения $L(x, \hat{\mu}, \hat{\sigma})$, n=20

Сравнивая $\chi_B^2 = 3{,}68$ и $\chi_{0.95}^2(7) \approx 9.49$, видим, что $\chi_B^2 < \chi_{0.95}^2(4)$

i	limits	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty;\ -1.1]$	3	0,14	2,71	0,286	0,03
2	(-1,1; -0,37]	4	0,22	4,43	-0,426	0,04
3	(-0,37; -0,37]	5	0,29	5,72	-0,722	0,09
4	(0,37; 1,1]	4	0,22	4,43	-0,426	0,04
5	(1.1; ∞)	4	0,14	2,71	1,286	0,61
Σ		20	1,00	20	-0,002	7,44

Таблица 1. Вычисление χ_B^2 при проверке гипотезы H_0 о законе распределения $U(x, \hat{\mu}, \hat{\sigma})$, n=20

Сравнивая $\chi_B^2 = 7{,}44$ и $\chi_{0.95}^2(7) \approx 9.49$, видим, что $\chi_B^2 < \chi_{0.95}^2(4)$

4. Выводы

По полученным результатам можно сделать вывод о том, что гипотеза H_0 о нормальном законе распределения $U(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$ согласуется с выборкой для нормального распределения N(x, 0, 1).

Также можно заметить, что для выборок, сгенерированных по равномерному закону и закону Лапласа, гипотеза H_0 оказалась принята.