Санкт-Петербургский Политехнический Университет
им. Петра Великого

Институт прикладной математики и механики

Кафедра "Прикладная математика"

Отчёт
Курсовая работа
по дисциплине
"Математическая статистика"

Выполнил студент:
Шарапов Сергей Андреевич
группа: 3630102/70401

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

Санкт-Петербург
2020 год

Содержание

1 Список таблиц

2 Постановка задачи

Провести исследования о распознавании распределения Лапласа как нормального. Генерировать выборки по закону распределения Лапласа, методом максимального правдоподобия оценивать параметры нормального распределения(μ и σ), предполагая, что полученная выборка может подчиняться нормальному закону. В итоге требуется использовать критерий согласия Пирсона для определения сходства распределений Лапласа и нормального распределения. При заданных параметрах изучить зависимость от мощности выборки $n = 50, 100, 200, 1000$. При мощности выборки $n = 100$ варьировать параметр распределения Лапласа: $\alpha = 0.5, 1, 2, 4$. В качестве уровня значимости взять $\alpha = 0,05$.

3 Теория

3.1 Метод максимального правдоподобия

Метод максимального правдоподобия – метод оценивания неизвестного параметра путём максимимзации функции правдоподобия.

$$\hat{\theta}_{\text{МП}} = argmax\mathbf{L}(x_1, x_2, \ldots, x_n, \theta) \tag{1}$$

Где $\mathbf{L}$ это функция правдоподобия, которая представляет собой совместную плотность вероятности независимых случайных величин $X_1, x_2, \ldots, x_n$ и является функцией неизвестного параметра θ

$$\mathbf{L} = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta) \tag{2}$$

Оценкой максимального правдоподобия будем называть такое значение $\hat{\theta}_{\text{МП}}$ из множества допустимых значений параметра θ, для которого функция правдоподобия принимает максимальное значение при заданных $x_1, x_2, \ldots, x_n$.

Тогда при оценивании математического ожидания m и дисперсии σ^2 нормального распределения $N(m, \sigma)$ получим:

$$\ln(\mathbf{L}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2 \tag{3}$$

Отсюда находятся выражения для оценок m и σ^2:

$$\begin{cases} m = \bar{x} \\ \sigma^2 = s^2 \end{cases} \tag{4}$$

3.2 Критерий согласия Пирсона

Разобьём генеральную совокупность на k непересекающихся подмножеств $\Delta_1, \Delta_2, \ldots, \Delta_k,\ \Delta_i = (a_i, a_{i+1}]$, $p_i = P(X \in \Delta_i),\ i = 1, 2, \ldots, k$ – вероятность того, что точка попала в iый промежуток.

Так как генеральная совокупность это $\mathbb{R}$, то крайние промежутки будут бесконечными: $\Delta_1 = (-\infty, a_1],\ \Delta_k = (a_k, \infty),\ p_i = F(a_i) - F(a_{i-1})$

n_i – частота попадания выборочных элементов в $\Delta_i,\ i = 1, 2, \ldots, k$.

В случае справедливости гипотезы H_0 относительно частоты $\frac{n_i}{n}$ при больших n должны быть близки к p_i, значит в качестве меры имеет смысл взять:

$$Z = \sum_{i=1}^{k}\frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 \tag{5}$$

Тогда

$$\chi_B^2 = \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \tag{6}$$

Для выполнения гипотезы H_0 должны выполняться следующие условия:

$$\chi_B^2 < \chi_{1-\alpha}^2(k-1) \tag{7}$$

где $\chi_{1-\alpha}^2(k-1)$ – квантиль распределения χ^2 с $k-1$ степенями свободы порядка $1-\alpha$, где α заданный уровень значимости.

4 Реализация

Работа была выполнена на языке *Python*3.7 Для генерации выборок и обработки функции распределения использовалась библиотека *scipy.stats*.

5 Результаты

5.1 Исследование зависимости распознования распределения Лапласа как нормального от величины выборки

Дано распределение Лапласа:

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{2\sqrt{2}} e^{-\frac{|x|}{\sqrt{2}}} \tag{8}$$

Размер выборки $n = 50$:

$$\begin{aligned} \hat{m}_{\text{МП}} &= -0.0692 \\ \hat{\sigma}^2_{\text{МП}} &= 1.0152 \end{aligned} \tag{9}$$

Таблица 1: Таблица вычислений $\chi^2, n = 50$

i	Δ_i	n_i	p_i	$\frac{(n_i-np_i)^2}{np_i}$
0	$(-\infty, -1.0]$	8	0.1796	0.1070
1	$(-1.0, -0.5]$	7	0.1561	0.0826
2	$(-0.5, 0.0]$	15	0.1915	3.0717
3	$(0.0, 0.5]$	6	0.1853	1.1513
4	$(0.5, 1.0]$	7	0.1414	0.0007
5	$(1.0, \infty)$	7	0.1461	0.0128

$$\chi_B^2 = 4.4261$$

Размер выборки $n = 100$:

$$\begin{aligned} \hat{m}_{\text{МП}} &= 0.0746 \\ \hat{\sigma}^2_{\text{МП}} &= 0.8623 \end{aligned} \tag{10}$$

Таблица 2: Таблица вычислений χ^2, n = 100

i	Δ_i	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
0	$(-\infty, -1.0]$	9	0.1063	0.2510
1	$(-1.0, -0.5]$	9	0.1462	2.1633
2	$(-0.5, 0.0]$	26	0.2129	1.0398
3	$(0.0, 0.5]$	33	0.2236	5.0643
4	$(0.5, 1.0]$	9	0.1693	3.7141
5	$(1.0, \infty)$	14	0.1416	0.0018

$$\chi_B^2 = 12.2342$$

Размер выборки $n = 200$:

$$\begin{aligned} \hat{m}_{\text{МП}} &= -0.0052 \\ \hat{\sigma}^2_{\text{МП}} &= 1.0047 \end{aligned} \tag{11}$$

Таблица 3: Таблица вычислений χ^2, n = 200

i	Δ_i	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
0	$(-\infty, -1.0]$	26	0.1611	1.1979
1	$(-1.0, -0.75]$	10	0.0682	0.9715
2	$(-0.75, -0.5]$	10	0.0819	2.4898
3	$(-0.5, -0.25]$	23	0.0926	1.0879
4	$(-0.25, 0.0]$	30	0.0983	5.4340
5	$(0.0, 0.25]$	33	0.0982	9.0920
6	$(0.25, 0.5]$	16	0.0922	0.3231
7	$(0.5, 0.75]$	17	0.0814	0.0316
8	$(0.75, 1.0]$	10	0.0676	0.9155
9	$(1.0, \infty)$	25	0.1585	1.4183

$$\chi_B^2 = 22.9616$$

Размер выборки $n = 1000$:

$$\begin{aligned} \hat{m}_{\text{МП}} &= -0.0210 \\ \hat{\sigma}^2_{\text{МП}} &= 0.9983 \end{aligned} \tag{12}$$

Таблица 4: Таблица вычислений χ^2, n = 1000

i	Δ_i	n_i	p_i	$\frac{(n_i-np_i)^2}{np_i}$
0	$(-\infty,-1.0]$	129	0.1634	7.2424
1	$(-1.0,-0.8571]$	28	0.0378	2.5226
2	$(-0.8571,-0.7143]$	35	0.0426	1.3424
3	$(-0.7143,-0.5714]$	39	0.0470	1.3606
4	$(-0.5714,-0.4286]$	45	0.0508	0.6724
5	$(-0.4286,-0.2857]$	66	0.0539	2.7162
6	$(-0.2857,-0.1429]$	74	0.0560	5.8002
7	$(-0.1429,0.0]$	96	0.0570	26.7487
8	$(0.0,0.1429]$	99	0.0568	31.3657
9	$(0.1429,0.2857]$	69	0.0555	3.2962
10	$(0.2857,0.4286]$	55	0.0531	0.0684
11	$(0.4286,0.5714]$	41	0.0498	1.5507
12	$(0.5714,0.7143]$	47	0.0457	0.0347
13	$(0.7143,0.8571]$	29	0.0412	3.5985
14	$(0.8571,1.0]$	30	0.0363	1.0966
15	$(1.0,\infty)$	118	0.1532	8.0947

$$\chi_B^2 = 97.5109$$

5.2 Исследование зависимости распознования распределения Лапласа как нормального от параметра α распределения

Дано распределение Лапласа:

$$L(x,0,\alpha) = \frac{\alpha}{2}e^{-\alpha|x|} \tag{13}$$

Размер выборки $n = 100$.
$\alpha = \frac{1}{2}$:

$$\begin{aligned} \hat{m}_{\text{МП}} &= 0.1507 \\ \hat{\sigma}^2_{\text{МП}} &= 0.5960 \end{aligned} \tag{14}$$

Таблица 5: Таблица вычислений $\chi^2, \alpha = \frac{1}{2}$

i	Δ_i	n_i	p_i	$\frac{(n_i-np_i)^2}{np_i}$
0	$(-\infty,-1.0]$	3	0.0268	0.0391
1	$(-1.0,-0.5]$	4	0.1107	4.5154
2	$(-0.5,0.0]$	35	0.2627	2.9011
3	$(0.0,0.5]$	34	0.3209	0.1140
4	$(0.5,1.0]$	17	0.2019	0.5026
5	$(1.0,\infty)$	7	0.0771	0.0655

$$\chi_B^2 = 8.1377$$

$\alpha = 1$:

$$\hat{m}_{\text{МП}} = 0.0157$$
$$\overset{\wedge}{\sigma}^2_{\text{МП}} = 1.2034 \tag{15}$$

Таблица 6: Таблица вычислений $\chi^2, \alpha = 1$

i	Δ_i	n_i	p_i	$\frac{(n_i-np_i)^2}{np_i}$
0	$(-\infty, -1.0]$	15	0.1993	1.2205
1	$(-1.0, -0.5]$	10	0.1348	0.8983
2	$(-0.5, 0.0]$	23	0.1607	2.9928
3	$(0.0, 0.5]$	24	0.1615	3.8129
4	$(0.5, 1.0]$	9	0.1370	1.6114
5	$(1.0, \infty)$	19	0.2067	0.1352

$$\chi_B^2 = 10.6710$$

$\alpha = 2$:

$$\hat{m}_{\text{МП}} = 0.2995$$
$$\overset{\wedge}{\sigma}^2_{\text{МП}} = 2.5786 \tag{16}$$

Таблица 7: Таблица вычислений $\chi^2, \alpha = 2$

i	Δ_i	n_i	p_i	$\frac{(n_i-np_i)^2}{np_i}$
0	$(-\infty, -1.0]$	26	0.3071	0.7235
1	$(-1.0, -0.5]$	4	0.0711	1.3613
2	$(-0.5, 0.0]$	6	0.0755	0.3185
3	$(0.0, 0.5]$	17	0.0772	11.1471
4	$(0.5, 1.0]$	11	0.0761	1.5135
5	$(1.0, \infty)$	36	0.3929	0.2763

$$\chi_B^2 = 15.3401$$

$\alpha = 4$:

$$\hat{m}_{\text{МП}} = -0.6772$$
$$\overset{\wedge}{\sigma}^2_{\text{МП}} = 5.4524 \tag{17}$$

Таблица 8: Таблица вычислений $\chi^2, \alpha = 4$

i	Δ_i	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
0	$(-\infty, -1.0]$	39	0.4764	1.5668
1	$(-1.0, -0.5]$	8	0.0366	5.1584
2	$(-0.5, 0.0]$	7	0.0365	3.0855
3	$(0.0, 0.5]$	8	0.0360	5.3594
4	$(0.5, 1.0]$	4	0.0353	0.0614
5	$(1.0, \infty)$	34	0.3792	0.4050

$$\chi^2_B = 15.6365$$

Табличные значения квантилей: $\chi^2_{0.95}(5) = 11.0705, \chi^2_{0.95}(15) = 24.9958$.

6 Выводы

Для выборки $n = 50$ распределения Лапласа полученно значение критерия Пирсона: $\chi^2_B = 4.4261 < \chi^2_{0.95}(5) = 11.0705$. Это означает, что из полученной выборки мы не можем опровергнуть гипотезу о нормальности исходного распределения на уровне значимости $\alpha = 0.05$. При данных параметрах и размере выборки распределение Лапласа распознается как нормальное, что не правда. При увеличении выборки до 100, 200 или 1000 элементов значение критерия Пирсона возрастает, превышая соответствующие значения $\chi^2_{1-\alpha}$, и уже позволяет отвергнуть потенциальную нулевую гипотезу о нормальном распределении. В результате увеличения выборки в интервалы попадает больше точек, что сильно влияет на получаемое значение критерия.

Исследуя распределения Лапласа с параметрами $\alpha = 1/2$ и 1, получилось, что они сходны с соотвествующими нормальными распределениями. Однако, с увеличением α росло и χ^2_B, поэтому при $\alpha = 2$ и 4 значение критерия Пирсона стало больше $\chi^2_{0.95}(5) = 11.0705$. Получается с увеличением параметра масштаба распределения Лапласа, оно становиться менее распознаваемым как нормальное.

7 Литература

Модуль numpy

Модуль matplotlib

Модуль scipy

Шевляков Г. Л. Лекции по математической статистике, 2019.

8 Приложения

Код курсового проекта