Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра прикладной математики

ОТЧЕТ

Тема: ***Точечные оценки характеристик распределения***

Направление: 01.03.02 Прикладная математика и информатика

Выполнил студент гр. 33631/4 Камалетдинова Ю.

Преподаватель Баженов А.

Санкт-Петербург
2019

Содержание

Постановка задачи **2**

Описание алгоритма **2**

Реализация **4**

Результат **6**

Вывод **7**

Постановка задачи

Рассматривается один из методов точечной оценки характеристик распределения. Требуется сгенерировать выборку объемом $n = 100$ элементов для нормального распределения $N(x; 0, 1)$ и оценить по ней параметры μ и σ нормального распределения методом максимального правдоподобия. В качестве основной гипотезы H_0 будем считать, что сгенерированное распределение имеет вид $N(x; \hat{\mu}, \hat{\sigma})$. Проверить гипотезу, используя критерий согласия χ^2. В качестве уровня значимости взять $\alpha = 0.05$. Привести таблицу вычислений χ^2.

Описание алгоритма

Метод максимального правдоподобия

Пусть $x_1, \dots, x_n$ — случайная выборка из распределения с плотностью вероятности $f(x; \theta)$. Функцией правдоподобия (ФП) назовем совместную плотность вероятности независимых случайных величин $x_1, \dots, x_n$, рассматриваемую как функцию неизвестного параметра θ

$$L(x_1, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) \tag{1}$$

Оценкой максимального правдоподобия $\hat{\theta}_{\text{МП}}$ будем называть такое значение , для которого из множества допустимых значений параметра θ ФП имеет наибольшее значение при заданных $x_1, \dots, x_n$

$$\hat{\theta}_{\text{МП}} = arg\ max\ L(x_1, \dots, x_n; \theta) \tag{2}$$

Если функция правдоподобия дважды дифференцируема, ее стационарные значения задаются корнями уравнения

$$\frac{\partial L(x_1, \ldots, x_n; \theta)}{\partial \theta} = 0 \tag{3}$$

Запишем условие локального максимума $\overline{\theta}$

$$\frac{\partial^2 L}{\partial \theta^2}(x_1, \ldots, x_n; \overline{\theta}) < 0 \tag{4}$$

Наибольший локальный максимум будет являться решением задачи (2).

Мы будем искать максимум логарифма функции правдоподобия в виду того, что он имеет максимум в одной точке с функцией правдоподобия

$$\frac{\partial \ln L}{\partial \theta}, \text{ если } L > 0, \tag{5}$$

и будем решать уравнение правдоподобия

$$\frac{\partial \ln L}{\partial \theta} = 0 \tag{6}$$

Для проверки гипотезы о характере распределения воспользуемся критерием χ^2 для случая, когда параметры распределения известны. Пусть H_0 — гипотеза о генеральном законе распределения, H_1 — гипотеза о справедливости одного из конкурирующих законов распределений. Разобьем генеральную совокупность на k непересекающихся подмножеств $\Delta_1, \ldots, \Delta_k$ при условиях

$$p_i = P(X \in \Delta_i),\ i = \overline{1, k}\ ;\ \ \sum_{i=1}^{k} p_i = 1 \tag{7}$$

Положим n_i — частота попадания выборочного элемента в подмножество Δ_i. За меру отклонения выборочного распределения от гипотетического примем величину

$$Z = \sum_{i=1}^{k} c_i \left(\frac{n_i}{n} - p_i \right)^2, \tag{8}$$

где $\frac{n_i}{n}$ — относительные частоты, c_i — некие положительные числа (веса). В качестве весов К. Пирсоном были взяты числа $c_i = \frac{n}{p_i}$. Получаем статистику критерия хи-квадрат К. Пирсона

$$\chi^2 = \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{n_i}{p} - p_i \right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \tag{9}$$

По теореме К. Пирсона из пособия [1] статистика критерия χ^2 асимптотически распределена по закону χ^2 с $k-1$ степенями свободы

Формулы (1) — (9) и определения взяты из источника [1]

Реализация

Для выполнения поставленной задачи будем пользоваться библиотеками для языка Python: *numpy, scipy* – расчеты, законы распределения вероятностей; *matplotlib, seaborn* – визуализация результатов. Ход работы:

- Генерируем выборку из распределения $N(x; 0, 1)$ объемом $n = 100$
- Запишем выражение для логарифма функции правдоподобия:

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{10}$$

- Получим два уравнения правдоподобия:

$$\begin{cases} \dfrac{\partial \ln L}{\partial \mu} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \hat{\mu}) = \dfrac{n}{\sigma^2} (\overline{x} - \hat{m}) = 0 \\ \dfrac{\partial \ln L}{\partial (\sigma^2)} = -\dfrac{n}{\hat{\sigma}^2} + \dfrac{1}{2(\hat{\sigma}^2)^2} \displaystyle\sum_{i=1}^{n} (x_i - \hat{\mu})^2 = \dfrac{n}{2(\hat{\sigma}^2)^2} \left[\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (x_i - \hat{\mu})^2 - \hat{\sigma}^2 \right] = 0 \end{cases} \tag{11}$$

- Из уравнений (11) получили, что выборочное среднее $\overline{x}$ — оценка максимума правдоподобия математического ожидания: $\hat{\mu}_{\text{МП}} = \overline{x}$, а выборочная дисперсия

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}^2)$ — оценка максимума правдоподобия генеральной дисперсии $\hat{\sigma}^2_{\text{МП}} = s^2$

Опишем порядок работы для проверки гипотезы о нормальности распределения по методу хи-квадрат

- Найдем квантиль $\chi^2_{1-\alpha}(k-1)$, где $\alpha = 0.05$ — уровень значимости, $k = 1 + 3.3 \ln n = 16$ — число подмножеств разбиения генеральной совокупности, вычисляемое по формуле Старджесса из источника [1]
- Вычисляем выборочное значение статистики хи-квадрат по формуле

$$\chi^2_B = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \tag{12}$$

- Сравниваем χ^2_B и $\chi^2_{1-\alpha}(k-1)$
  - Если $\chi^2_B < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на этапе проверки принимается
  - Если $\chi^2_B \geq \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений и процедура проверки повторяется

Результат

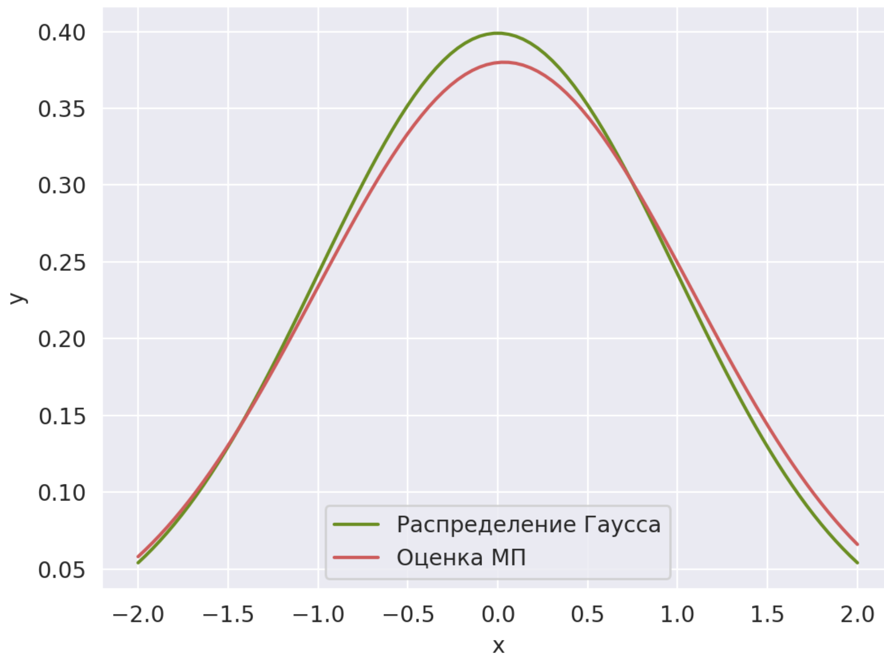

Рис. 1: Графики стандартного нормального распределения и распределения с параметрами, полученными методом МП

$\chi^2_{1-\alpha}(k-1) = 26,296$

$k=16$	a_{i-1} ; a_i	n_i	$\hat{p}_i$	$n_i - n\hat{p}_i$	$(n_i - n\hat{p}_i)^2/(n\hat{p}_i)$
1	-inf ; -3.50	0	0.0004	-0.04	0.04
2	-3.50 ; -2.96	0	0.0018	-0.18	0.18
3	-2.96 ; -2.42	1	0.0074	0.26	0.09
4	-2.42 ; -1.88	2	0.0241	-0.41	0.07
5	-1.88 ; -1.35	3	0.0604	-3.04	1.53
6	-1.35 ; -0.81	14	0.1169	2.31	0.46
7	-0.81 ; -0.27	22	0.1749	4.51	1.16
8	-0.27 ; 0.27	20	0.2023	-0.23	0.00
9	0.27 ; 0.81	15	0.1809	-3.09	0.53
10	0.81 ; 1.35	11	0.1251	-1.51	0.18
11	1.35 ; 1.88	7	0.0668	0.32	0.01
12	1.88 ; 2.42	4	0.0276	1.24	0.56
13	2.42 ; 2.96	1	0.0088	0.12	0.02
14	2.96 ; 3.50	0	0.0022	-0.22	0.22
15	3.50 ; inf	0	0.0005	-0.05	0.05
$\sum$	————	100	1	0	$\chi^2_B = 5.09$

Таблица 1: Таблица вычислений χ^2_B при проверке гипотезы о нормальности распределения

$\chi^2_B = 5.09 < 26.296 \approx \chi^2_{1-\alpha}(k-1)$ — гипотеза принимается

Вывод

В случае нормального распределения оценка параметров распределения методом максимума правдоподобия эффективна, состоятельна, асимптотически нормальна. В ходе проверки гипотезы было получено значение $\chi^2_B = 5.09$, являющееся очень малым, и соответствующий ему уровень значимости равен $\alpha = 0.99$, что говорит об очень хорошем согласии гипотезы H_0 и полученных данных.

Список литературы

[1] *Амосова Н.Н., Куклин Б.А., Макарова С.Б., Максимов Ю.Д., Митрофанова Н.М., Полищук В.И., Шевляков Г.Л.* Вероятностные разделы математики. Учебник для бакалавров технических направлений. — СПб.: Иван Федоров, 2001. — 592 с.: илл. — ISBN 5-81940-050-X.