

Студент: Киреев Константин
 Группа: 8383
 Вариант: 8
 Дата: 27 марта 2021 г.

Статистический анализ

Индивидуальное домашнее задание №2

Часть 1

В результате эксперимента получены данные.

0	1	1	1	1	0	0	0	2	0	1	0	0	1	1	1	2	0	0	0	0	4	0	1	0
0	0	0	0	0	1	1	1	2	3	1	0	1	0	0	2	0	0	0	1	1	1	0	1	2

$\alpha_1 = 0.10$	$a = 0.00$	$b = 1.37$	$\lambda_0 = 0.70$	$\lambda_1 = 1.40$
-------------------	------------	------------	--------------------	--------------------

Задача 1. Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.

Решение.

Вариационный ряд:

0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3 4

Построим таблицу частот для выборки.

x_j	0	1	2	3	4
m_j	25	18	5	1	1
p_j^*	$1/2$	$9/25$	$1/10$	$1/50$	$1/50$

Построим эмпирическую функцию распределения по полученным данным:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0.5, & 0 < x \leq 1 \\ 0.86, & 1 < x \leq 2 \\ 0.96, & 2 < x \leq 3 \\ 0.98, & 3 < x \leq 4 \\ 1, & x > 4 \end{cases}$$

□

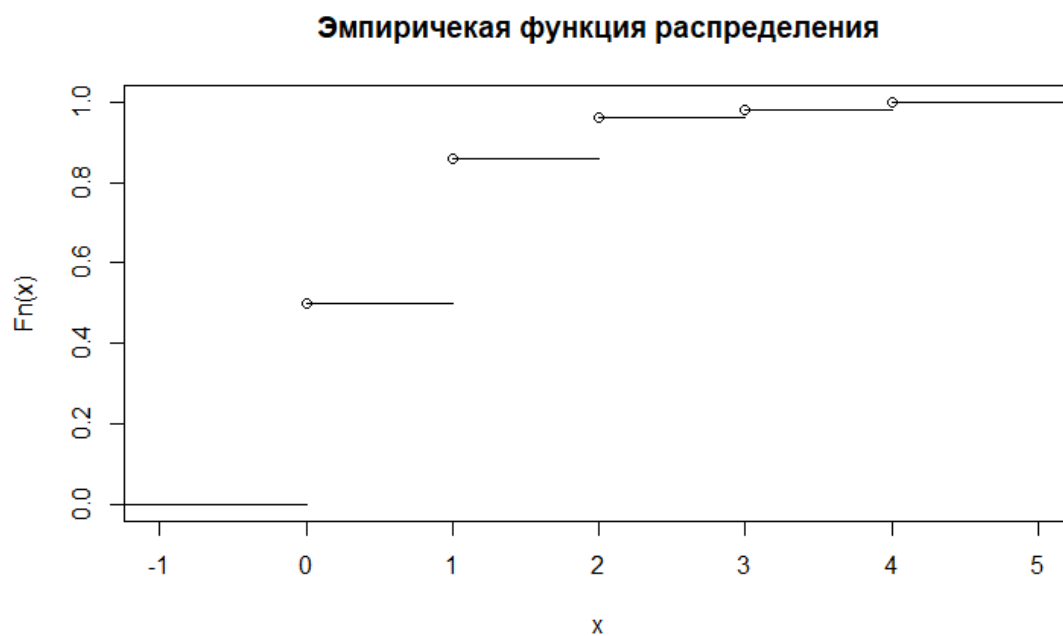


Рис. 1. Эмпирическая функция распределения

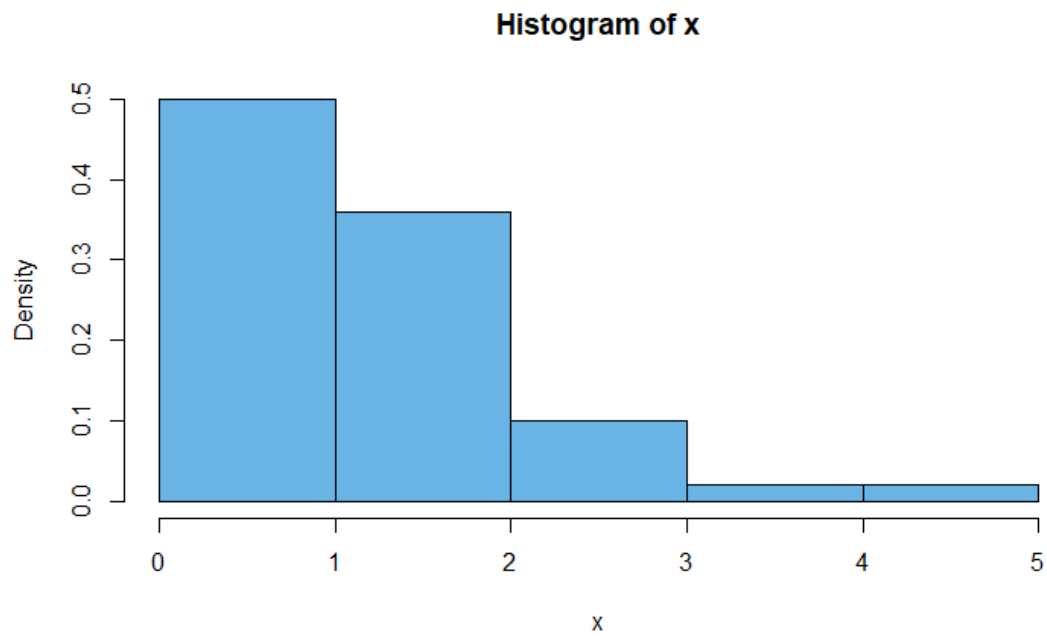


Рис. 2. Гистограмма частот

Задача 2. Вычислить выборочные аналоги следующих характеристик:

- математического ожидания
- дисперсии
- медианы
- асимметрии
- эксцесса
- вероятности $P(X \in [a, b])$

Решение.

- Математическое ожидание:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^n x_i m_i = \frac{1}{50} \cdot (18 + 10 + 7) = \frac{35}{50} = 0.7$$

- Дисперсия:

$$D_B = \bar{x^2} - \bar{x}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{50} \cdot (18 + 20 + 9 + 16) - 0.49 = \frac{63}{50} - 0.49 = 1.26 - 0.49 = 0.77$$

- Медиана:

$$Me = 0.5$$

- Асимметрия:

$$As = \frac{\mu_3^*}{\sigma^3} = 1.464547$$

- Эксцесс:

$$Ex = \frac{\mu_4^*}{\sigma^4} - 3 = 2.633496$$

- Вероятность:

$$P(x \in [a, b]) = P(x \in [0.00, 1.37]) = F(1.37) - F(0.00) = 0.86$$

□

Задача 3. В предположении, что исходные наблюдения являются выборкой из распределения Пуассона, построить оценку максимального правдоподобия параметра λ , а также оценку λ по методу моментов. Найти смещение оценок

Решение.

Распределение Пуассона: $P_\lambda = (x = k) = \frac{\lambda^k}{k!} \cdot \exp(-\lambda)$

- Метод максимального правдоподобия

$$l(x, \lambda) = \frac{\lambda^{\sum_1^n x_i} \cdot \exp(-\lambda n)}{\prod_1^n x_i!} \Rightarrow ll(\bar{x}, \lambda) = \sum_1^n x_i \ln \lambda - n \ln \lambda - \sum_1^n \ln x_i! \Rightarrow \frac{\partial ll(\bar{x}, \lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_1^n x_i - n = 0 \Rightarrow$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n} \sum_1^n x_i = \bar{x}_B = 0.7$$

- Метод моментов

$$P(x, \theta)$$

$$M_1^* = \bar{x}_B; \mathbb{E}X = \bar{x}_B;$$

$$\mathbb{E}X = \int_{\mathbb{R}} xp(x, \theta) dx = \varphi(\theta)$$

$$M_1 = \mathbb{E}X = \lambda; M_1^* = \bar{x}_B \Rightarrow \underline{\hat{\lambda} = \bar{x}_B = 0.7}$$

Чтобы найти смещение оценки, найдем:

$$\mathbb{E}\hat{\lambda} = \frac{1}{n} \sum_1^n \mathbb{E}x_i = \frac{n\lambda}{n} = \lambda \Rightarrow \text{оценки несмещенные}$$

□

Задача 4. Построить асимптотический доверительный интервал уровня значимости α_1 для параметра λ на базе оценки максимального правдоподобия.

Решение.

$$\hat{\lambda} = \bar{x}_B = 0.7; \alpha_1 = 0.10; \gamma = 1 - \alpha_1 = 0.90;$$

$$\frac{\sqrt{n}(\bar{x} - \lambda)}{\sqrt{\lambda}} \xrightarrow{n \rightarrow \infty} N(0,1)$$

$$t_\gamma : \phi(t_\gamma) = 1 - \frac{\alpha_1}{2} \Rightarrow t_\gamma = 1.645$$

$$P(-t_\gamma \leq \frac{\sqrt{n}(\bar{x}_B - \lambda)}{\sqrt{\lambda}} \leq t_\gamma) \longrightarrow 1 - \alpha$$

$$n(\bar{x}_B - \lambda)^2 = t_\gamma^2 \lambda$$

$$\lambda^2 - 2\lambda(\bar{x}_B + \frac{t_\gamma^2}{2n}) + \bar{x}_B = 0$$

$$D = \frac{t_\gamma^2}{n}(\bar{x}_B + \frac{t_\gamma^2}{4n})$$

$$\lambda_{1,2} = \bar{x}_B + \frac{t_\gamma^2}{2n} \pm t_\gamma \sqrt{\frac{1}{n}(\bar{x}_B + \frac{t_\gamma^2}{4n})} = 0.7 + 0.027 \pm 0.1965 \Rightarrow [0.5305; 0.9235]$$

□

Задача 5. Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением Пуассона с параметром λ_0 . Проверить гипотезу на уровне значимости α_1 . Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Решение.

$$\hat{\lambda}_0 = 0.70; \alpha_1 = 0.10;$$

m_i — эмпирические частоты;

m'_i — выравнивающие частоты; $m'_i = n \cdot p_i$

Простая гипотеза H_0 имеет вид:

$$H_0 : p(x) = \frac{\lambda_0^x}{x!} \exp(-\lambda_0)$$

Тогда искомая вероятность примет вид:

$$p_k = P(X = k) = \frac{0.7^k}{k!} \exp(-0.7)$$

Построим таблицу оценки методом χ^2 .

x_i	0	1	2	3	4	Σ
m_i	25	18	5	1	1	50
p_i	0.497	0.348	0.122	0.029	0.005	1
m'_i	24.8292652	17.3804856	6.0831700	1.4194063	0.2876729	50
$m_i - m'_i$	0.1707348	0.6195144	-1.0831700	-0.4194063	0.7123271	0
$\frac{(m_i - m'_i)^2}{m'_i}$	0.001174033	0.022082125	0.192869374	0.123926224	1.763843457	$\chi^2_{\text{набл}}$

$$\chi_{\text{набл}}^2 = \sum_1^k \frac{(m_i - m'_i)^2}{m'_i} = 2.103895$$

$$l = k - r - 1 = 5 - 1 - 1 = 3$$

$$\chi_{\text{кр}}^2 = \chi_{\alpha;l}^2 = \chi_3^2 \text{ на уровне значимости } 0.1 = 6.251$$

$\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2 \Rightarrow$ гипотеза H_0 принимается, выборка принадлежит распределению Пуассона. Наибольшее значение уровня значимости, при котором еще нет оснований отвергнуть данную гипотезу = 0.5511251 □

Задача 6. Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением Пуассона. Проверить гипотезу на уровне значимости α_1 . Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Решение.

Сложная гипотеза H_0 имеет вид:

$$H_0 : x_1, \dots, x_n \sim P_{ois}(\lambda)$$

$$\sum_1^k \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)} \longrightarrow \chi_{k-r-1}^2$$

Метод минимизации хи-квадрат:

$$\underset{\lambda}{\operatorname{argmin}} \sum_1^r \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)}$$

Задача реализована в R следующим скриптом:

```
P <- -function(a){
p < -0
p[1] < -ppois(0, a)
p[2] < -ppois(1, a) - sum(p)
p[3] < -ppois(2, a) - sum(p)
p[4] < -ppois(3, a) - sum(p)
```

$$p[5] < -1 - \text{sum}(p)$$

$$p\}$$

$$X2 < -\text{function}(a)\{g < -n \cdot P(a); f < -(nu - g)^2/g; \text{sum}(f)\}$$

$$nu < -c(25, 17, 6, 1, 1)$$

$$XM < -nlm(X2, 0.70)$$

В результате вычислений получим, что $\chi^2_{\text{набл}} = 1.861649 < \chi^2_{\text{крит}} = 4.60517$ Таким образом, гипотеза принимается.

Наибольшее значение уровня значимости, при котором еще нет оснований отвергнуть данную гипотезу = 0.3942285 □

Задача 7. Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром $\lambda = \lambda_0 = 0.70$ при альтернативе пуассоновости с параметром $\lambda = \lambda_1 = 1.40$. Проверить гипотезу на уровне значимости α_1 . Что получится, если поменять местами основную и альтернативную гипотезы?

Решение.

Сформулируем гипотезы.

$$H_0 : \lambda = \lambda_0 = 0.70$$

$$H_1 : \lambda = \lambda_1 = 1.40$$

По лемме Неймана-Пирсона:

$$\phi(\bar{x}) = \begin{cases} 0, & \text{if } l(\bar{x}, \lambda_0, \lambda_1) < C \\ p, & \text{if } l(\bar{x}, \lambda_0, \lambda_1) = C \\ 1, & \text{if } l(\bar{x}, \lambda_0, \lambda_1) > C \end{cases}$$

$$l(\bar{x}, 0.7, 1.4) = \frac{L(\bar{x}, 1.4)}{L(\bar{x}, 0.7)} = 2^{\sum x_i} \cdot \exp(n * (\lambda_0 - \lambda_1)) = 2^{\sum x_i} \cdot \exp(-0.7n)$$

$$ll(\bar{x}, \lambda_0, \lambda_1) = - \sum x_i \cdot \ln 2 - 0.7n < \ln C$$

$$\sum x_i > \frac{-\ln C - 0.7n}{\ln 2}$$

$$\hat{C} = \frac{-\ln C - 0.7n}{\ln 2}$$

Критерий принимает вид:

$$\phi(\bar{x}) = \begin{cases} 0, \text{ if } \sum x_i > \hat{C} \\ p, \text{ if } \sum x_i = \hat{C} \\ 1, \text{ if } \sum x_i < \hat{C} \end{cases}$$

Вычислим \hat{C} и p из уравнения:

$$\begin{aligned} P_{\lambda_0}(l(\bar{x}, \lambda_0, \lambda_1) > C) + p \cdot P_{\lambda_0}(l(\bar{x}, \lambda_0, \lambda_1) = C) = \\ = P_{\lambda_0}\left(\sum_1^n x_i > \hat{C}\right) + p \cdot P_{\lambda_0}\left(\sum_1^n x_i = \hat{C}\right) = \alpha_1 = 0.1 \\ x_i \rightarrow P_{ois}(\lambda_0) \\ \sum x_i \rightarrow P_{ois}(n\lambda_0) \end{aligned}$$

Подбором среди целых чисел можно найти такое наибольшее \hat{C} и α_0 , что

$$\begin{aligned} \alpha_0 = P_{\lambda_0}\left(\sum_1^n x_i > \hat{C}\right) = 1 - P_{n\lambda_0}(\hat{C}) - p_{n\lambda_0}(\hat{C}) < \alpha_1 \\ p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_0}\left(\sum_1^n x_i = A\right)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_0}(A)} \end{aligned}$$

В результате расчета получим: $\alpha_0 = 0.09867$; $\hat{C} = 41$; $p = 0.03499$

$$\sum_1^n x_i = 35$$

$35 < 41 \Rightarrow$ Таким образом, принимаем гипотезу H_0

Теперь поменяем местами основную и альтернативную гипотезы.

$$H_0 : \lambda = \lambda_1 = 1.40$$

$$H_1 : \lambda = \lambda_0 = 0.70$$

$$l(\bar{x}, 1.4, 0.7) = \frac{L(\bar{x}, 0.7)}{L(\bar{x}, 1.4)} = \left(\frac{1}{2}\right)^{\sum x_i} \cdot \exp(n * (\lambda_1 - \lambda_0)) = \left(\frac{1}{2}\right)^{\sum x_i} \cdot \exp(0.7n)$$

$$ll(\bar{x}, \lambda_0, \lambda_1) = - \sum x_i \cdot \ln\left(\frac{1}{2}\right) + 0.7n < \ln C$$

$$\sum x_i < \frac{\ln C - 0.7n}{\ln\left(\frac{1}{2}\right)}$$

$$\hat{C} = \frac{\ln C - 0.7n}{\ln\left(\frac{1}{2}\right)}$$

Тогда критерий примет вид:

$$\phi(\bar{x}) = \begin{cases} 0, \text{ if } \sum x_i < \hat{C} \\ p, \text{ if } \sum x_i = \hat{C} \\ 1, \text{ if } \sum x_i > \hat{C} \end{cases}$$

Вычислим \hat{C} и p из уравнения:

$$P_{\lambda_1}\left(\sum_1^n x_i > \hat{C}\right) + p \cdot P_{\lambda_1}\left(\sum_1^n x_i = \hat{C}\right) = \alpha_1 = 0.1$$

$$p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_1}\left(\sum_1^n x_i = A\right)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_1}(A)}$$

В результате расчета получим: $\alpha_0 = 0.081593$; $\hat{C} = 58$; $p = 1.049973$

$$\sum_1^n x_i = 35$$

$35 < 58 \Rightarrow$ Таким образом, отвергаем гипотезу H_0

При замене основной и альтернативной гипотезы меняется также гипотеза, которая принимается. Но так как изменение происходит со сменой гипотез местами, решение не меняется. \square

Задача 8. В пунктах (с) - (f) заменить семейство распределений Пуассона на семейство геометрических распределений.

Решение.

$$P_{\lambda}(X = k) = \frac{\lambda^k}{(\lambda + 1)^{k+1}}, k = 0, 1, \dots$$

□

Задача 9. В предположении, что исходные наблюдения являются выборкой из геометрического распределения, построить оценку максимального правдоподобия параметра λ , а также оценку λ по методу моментов. Найти смещение оценок

Решение.

Плотность геометрического распределения имеет вид:

$$P_{\lambda}(X = k) = \frac{\lambda^k}{(\lambda + 1)^{k+1}}$$

- Метод максимального правдоподобия

$$l(\bar{x}, \lambda) = \prod_1^n \frac{\lambda^{x_i}}{(\lambda + 1)^{x_i+1}} = \frac{\lambda^{\sum_1^n x_i}}{(\lambda + 1)^{\sum_1^n x_i + n}}$$

$$ll(\bar{x}, \lambda) = \ln \lambda \cdot \sum_1^n x_i - \ln(\lambda + 1) \sum_1^n x_i - n \ln(\lambda + 1)$$

$$\frac{\partial ll}{\partial \lambda} = \frac{1}{\lambda} \sum_1^n x_i - \frac{1}{\lambda + 1} \sum_1^n x_i - \frac{n}{\lambda + 1}$$

$$\frac{\partial ll}{\partial \lambda} = 0 \rightarrow \hat{\lambda} = \frac{1}{n} \sum_1^n x_i = \bar{x} = 0.7$$

- Метод моментов

$$M_1 = \mathbb{X} = \lambda$$

$$M_1^* = \hat{X}$$

$$\hat{\lambda} = \bar{X}$$

Чтобы найти смещение оценки, найдем:

$$\mathbb{E}\hat{\lambda} = \mathbb{E}\hat{X} = \frac{1}{n} \sum_1^n \mathbb{E}(x_i) = \frac{n\lambda}{n} = \lambda \Rightarrow \text{оценки несмещенные}$$

□

Задача 10. Построить асимптотический доверительный интервал уровня значимости $\alpha_1 = 0.10$ для параметра λ на базе оценки максимального правдоподобия.

Решение.

$$\frac{\partial^2 l}{\partial \lambda^2} = \frac{1}{\lambda^2} \sum_1^n x_i + \frac{1}{(\lambda + 1)^2} \sum_1^n x_i + \frac{n}{(\lambda + 1)^2}$$

$$\hat{I} = -\frac{\partial^2 l}{\partial \lambda^2}(\hat{\lambda}) = -\frac{\partial^2 l}{\partial \lambda^2}(\hat{X}) = n\left(\frac{1}{\bar{X}} - \frac{1}{\bar{X} + 1}\right) = 42.017$$

$$\sigma^2(\hat{\lambda}) = \hat{I}^{-1} = 0.024$$

$$\sigma = \sqrt{\hat{I}^{-1}} = 0.154$$

Доверительный интервал будет иметь вид

$$[\hat{\lambda} - x_\alpha \sigma, \hat{\lambda} + x_\alpha \sigma]$$

$$x_\alpha = \phi^{-1}\left(1 - \frac{\alpha_1}{2}\right) = 1.645$$

Получен доверительный интервал $[0.4467, 0.9534]$

□

Задача 11. Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с геометрическим распределением с параметром $\lambda_0 = 0.70$. Проверить гипотезу на уровне значимости $\alpha_1 = 0.10$. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Решение.

$$H_0 : X_1, \dots, X_n \sim \text{Geom}\left(\frac{1}{0.7 + 1}\right) = \text{Geom}\left(\frac{1}{1.7}\right)$$

Построим таблицу оценки методом χ^2 .

x_i	0	1	2	3	4	Σ
m_i	25	18	5	1	1	50
p_i	0.58823529	0.24221453	0.09973540	0.04106752	0.02874726	1
m'_i	29.411765	12.110727	4.986770	2.053376	1.437363	50
$m_i - m'_i$	-4.41176471	5.88927336	0.01323021	-1.05337580	-0.43736306	0
$\frac{(m_i - m'_i)^2}{m'_i}$	0.6617647	2.86387	0.00003510055	0.5403787	0.1330815	$\chi^2_{\text{набл}}$

Итого

$$\chi_{\text{набл}}^2 = \sum_1^k \frac{(m_i - m'_i)^2}{m'_i} = 4.19913$$

$$l = k - r - 1 = 5 - 1 - 1 = 3$$

$$\chi_{\text{кр}}^2 = \chi_{\alpha;l}^2 = \chi_3^2 \text{ на уровне значимости } 0.1 = 6.251389$$

$\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2 \Rightarrow$ гипотеза H_0 принимается, выборка принадлежит геометрическому распределению. Наибольшее значение уровня значимости, при котором еще нет оснований отвергнуть данную гипотезу = 0.2407491 \square

Задача 12. Построить критерий значимости χ^2 проверки сложной гипотезы согласия с геометрическим распределением. Проверить гипотезу на уровне значимости $\alpha_1 = 0.10$. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Решение.

Сложная гипотеза H_0 имеет вид:

$$H_0 : X_1, \dots, X_n \sim \text{Geom} \left(\frac{1}{1 + \lambda} \right)$$
$$\sum_1^r \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)} \longrightarrow \chi_{k-r-1}^2$$

Метод минимизации хи-квадрат:

$$\underset{\lambda}{\operatorname{argmin}} \sum_1^k \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)}$$

Задача реализована в R следующим скриптом:

```
P <- -function(a){  
  p <- -0  
  p[1] <- -pgeom(0, a)  
  p[2] <- -pgeom(1, a) - sum(p)  
  p[3] <- -pgeom(2, a) - sum(p)
```

$$p[4] < -pgeom(3, a) - sum(p)$$

$$p[5] < -1 - sum(p)$$

$$p\}$$

$$X2 < -function(a)\{g < -n \cdot P(a); f < -(nu - g)^2/g; sum(f)\}$$

$$nu < -c(25, 18, 5, 1, 1)$$

$$XM < -nlm(X2, 1/(1 + 0.70))$$

$$\text{Получили оптимальную } \hat{\lambda} = \frac{1}{0.5733551} - 1 = 0.7441197$$

Построим таблицу оценки методом χ^2 .

x_i	0	1	2	3	4	Σ
m_i	25	18	5	1	1	50
p_i	0.58823529	0.24221453	0.09973540	0.04106752	0.02874726	1
m'_i	29.411765	12.110727	4.986770	2.053376	1.437363	50
$m_i - m'_i$	-4.41176471	5.88927336	0.01323021	-1.05337580	-0.43736306	0
$\frac{(m_i - m'_i)^2}{m'_i}$	0.6617647	2.863870	0.00003510055	0.5403787	0.1330815	$\chi^2_{\text{набл}}$

В результате вычислений получим, что $\chi^2_{\text{набл}} = 4.135312 < \chi^2_{\text{крит}} = 4.6$

Таким образом, гипотеза принимается.

Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = 0.1264819 □

Часть 2

В результате эксперимента получены данные.

2.953	2.962	2.922	3.026	2.987	2.964	2.898	2.777	2.977	3.102
3.019	3.040	2.939	2.969	3.170	3.021	2.943	2.998	3.075	2.968
2.919	2.960	3.137	3.077	2.967	3.146	3.081	3.002	2.896	2.863
2.984	3.151	2.863	2.948	2.946	3.021	3.067	3.206	2.926	3.082
2.965	3.180	3.105	3.084	2.885	3.225	2.878	3.106	3.062	3.138

$\alpha_2 = 0.20$	$c = 2.90$	$d = 3.06$	$h = 0.05$
$a_0 = 2.60$	$\sigma_0 = 0.10$	$a_1 = 3.00$	$\sigma_1 = 0.10$

Задача 13. Построить вариационный ряд, эмпирическую функцию распределения, гистограмму и полигон частот с шагом h .

Решение.

Вариационный ряд:

2.777 2.863 2.863 2.878 2.885 2.896 2.898 2.919 2.922 2.926 2.939 2.943 2.946 2.948 2.953
 2.960 2.962 2.964 2.965 2.967 2.968 2.969 2.977 2.984 2.987 2.998 3.002 3.019 3.021 3.021 3.026
 3.040 3.062 3.067 3.075 3.077 3.081 3.082 3.084 3.102 3.105 3.106 3.137 3.138 3.146 3.151 3.170
 3.180 3.206 3.225

□

Эмпирическая функция распределения

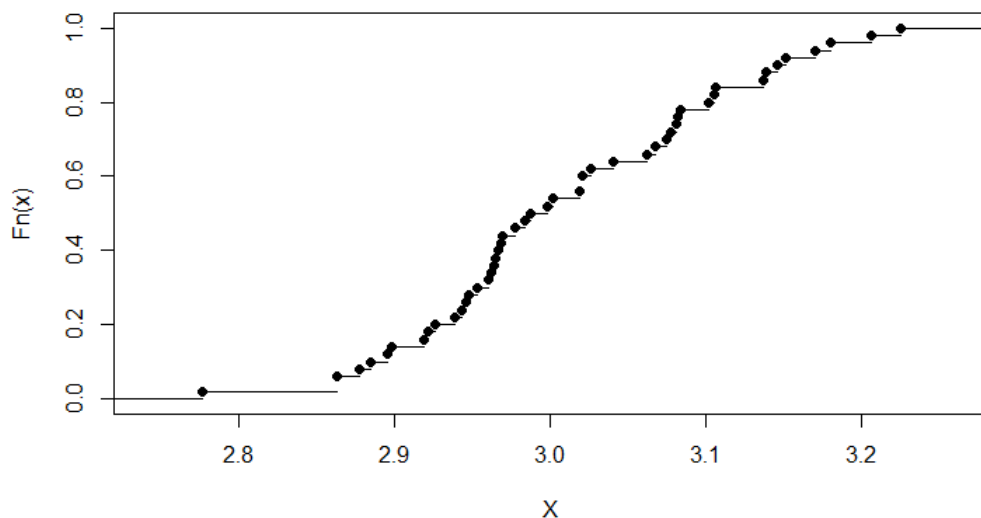


Рис. 3. Эмпирическая функция распределения

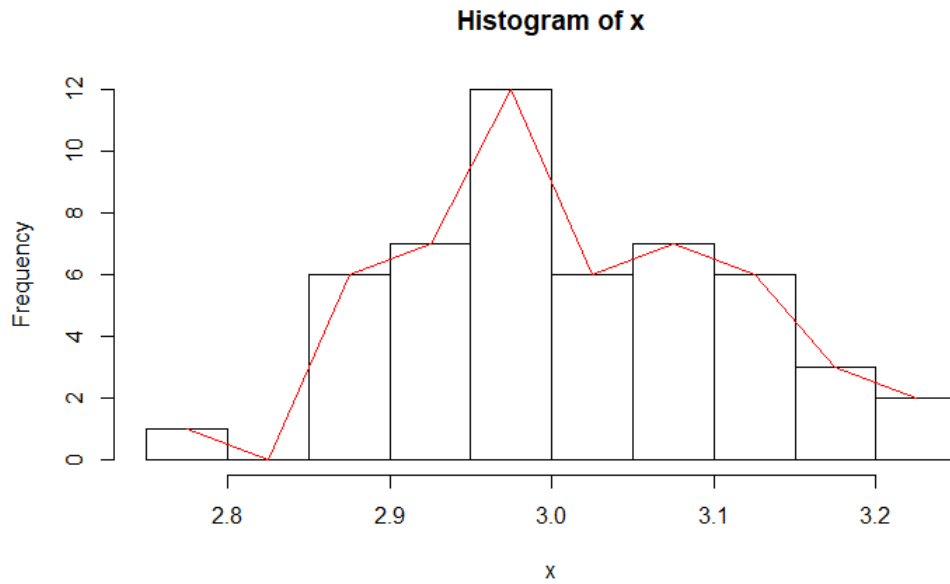


Рис. 4. Гистограмма и полигон частот с шагом $h = 0.05$

Задача 14. Вычислить выборочные аналоги следующих характеристик:

- математического ожидания
- дисперсии
- медианы
- асимметрии
- эксцесса
- вероятности $P(X \in [c, d])$

Решение.

- Математическое ожидание:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^n (x_i) = 3.0116$$

- Дисперсия:

$$D_B = \bar{x}^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.00970184$$

- Медиана:

$$Me = 2.993$$

- Ассиметрия:

$$As = \frac{\mu_3^*}{\sigma^3} = 0.1669468$$

- Эксцесс:

$$Ex = \frac{\mu_4^*}{\sigma^4} - 3 = -0.5054119$$

- Вероятность:

$$P(x \in [c, d]) = P(x \in [2.90, 3.06]) = F(3.06) - F(2.90) = 0.5$$

□

Задача 15. В предположении, что исходные наблюдения являются выборкой из нормального распределения, построить оценку максимального правдоподобия параметров (a, σ^2) , и соответствующие оценки по методу моментов. Найти смещение оценок.

Решение.

Плотность нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

- Метод максимального правдоподобия

Функция правдоподобия:

$$L(\vec{x}, a, \sigma^2) = \frac{1}{\sigma^n \sqrt{(2\pi)^n}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2\right)$$

Прологарифмируем:

$$LL(\vec{x}, a, \sigma^2) = \frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2$$

$$\left\{ \begin{array}{l} \frac{\partial LL(\vec{x}, a, \sigma^2)}{\partial a} = \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - na \right) = 0 \\ \frac{\partial LL(\vec{x}, a, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - a)^2 = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \hat{a} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = S^2 \end{array} \right.$$

Найдем значения $\hat{a} = \bar{x} = 3.0116$ - выборочное среднее и $\hat{\sigma}^2 = S^2 = 0.00970184$ - выборочная дисперсия.

- Метод моментов

В случае нормального распределения имеем $a'_1 = \mathbb{E}(x) = a$ и $a'_2 = \mathbb{E}(x^2) = \sigma^2 + a^2$

Уравнения моментов принимают вид:

$$\tilde{a} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} = 3.0116$$

$$\tilde{a}^2 + \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \bar{x}^2$$

Откуда моментные оценки:

$$\tilde{a} = \bar{x} = 3.0116$$

$$\tilde{\sigma}^2 = \bar{x}^2 - \bar{x}^2 = \frac{(\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i)}{n} = S^2 = 0.00970184$$

Оценки являются несмещенными.

□

Задача 16. Построить доверительные интервалы уровня значимости α_2 для параметров (a, σ^2)

Решение.

Параметр a :

$$\sqrt{n-1} \left(\frac{\bar{x} - a}{s} \right) \sim S_{n-1}$$

$$x_\alpha : S_{n-1}(x_\alpha) = 1 - \frac{\alpha_2}{2}$$

Тогда получаем:

$$P \left(-x_\alpha \leq \sqrt{n-1} \left(\frac{\bar{x} - a}{s} \right) \leq x_\alpha \right) = 1 - \alpha_2 = P \left(\bar{x} - \frac{x_\alpha s}{\sqrt{n-1}} \leq a \leq \bar{x} + \frac{x_\alpha s}{\sqrt{n-1}} \right)$$

ДИ уровня значимости $\alpha_2 = 0.20$ для a :

$$[2.9997; 3.0235]$$

Параметр σ^2 :

$$\frac{ns^2}{\sigma^2} \sim \chi_{n-1}^2$$

Тогда выберем $x_{1\alpha}, x_{2\alpha}$ - квантили распределения χ_{n-1}^2 уровня $\frac{\alpha_2}{2}$ и $1 - \frac{\alpha_2}{2}$ соответственно, тогда $p \left(\frac{nS^2}{x_{1\alpha}} \leq \sigma^2 \leq \frac{nS^2}{x_{2\alpha}} \right) = 1 - \alpha_2$
 ДИ уровня значимости $\alpha_2 = 0.20$ для σ^2 :

$$[0.0078178; 0.0131728]$$

□

Задача 17. С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с нормальным распределением с параметрами a_0, σ_0^2 . Проверить гипотезу на уровне значимости α_2 . Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

Решение.

Простая гипотеза $H_0 : a = a_0 = 2.60, \sigma^2 = \sigma_0^2 = 0.10$

Согласно теореме Колмогорова, при справедливости гипотезы:

$\sqrt{n} \sup |F_n(x) - F_0(x)| \rightarrow K(x)$, где $K(x)$ - функция Колмогорова

$$K(C_{\alpha_2}) = 1 - \alpha_2 = 1 - 0.2 = 0.8$$

$$C_{\alpha_2} = C_{0.2} = 1.0493$$

$$H_0 : X_1, \dots, X_n \sim N(2.60, 0.10)$$

С помощью R вычислим:

$$\sup |F_n(x) - F_0(x)| = 0.97573$$

$$\sqrt{n} \sup |F_n(x) - F_0(x)| = 6.658375 > 1.0493$$

Отвергаем гипотезу H_0

$$p - value = 2.2 \cdot 10^{-16}$$

Наибольшее значение уровня значимости, на котором нет оснований отвергнуть гипотезу

□

Задача 18. Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с нормальным распределением с параметрами (a_0, σ_0^2) . Проверить гипотезу на уровне значимости α_2 . Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

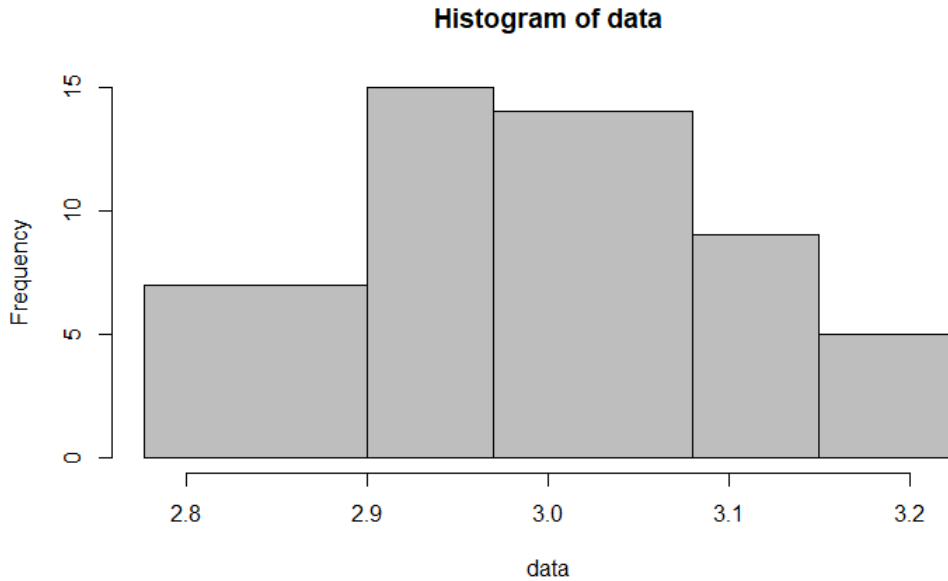
Решение.

$$H_0 : X_1, \dots, X_n \sim N(2.6, 0.1)$$

$$\sum_{i=1}^k \frac{(m_i - m'_i)^2}{m_i} \rightarrow \chi_{\text{кр}}^2$$

$$\chi_{\text{кр}}^2 = \chi_4^2 = 5.988617$$

Перестроим гистограмму частот, выбрав следующие точки: $(-\text{Inf}, 2.9, 2.97, 3.08, 3.15, \text{Inf})$



Интервал	$(-\infty; 2.9]$	$(2.9; 2.97]$	$(2.97; 3.08]$	$(3.08; 3.15]$	$(3.15; \infty)$	Σ
m_i	7	15	14	9	5	50
p_i	0.9987	0.0012	0.0001	~ 0	~ 0	1
$\frac{(m_i - m'_i)^2}{m'_i}$	36	3592	36605	2092090	26330240	$\chi_{\text{набл}}^2$

$$\chi_{\text{набл}}^2 = 28462569 \gg \chi_{\text{кр}}^2 = 5.988617$$

$$P(\chi_{\text{набл}}^2 > 28462569) \rightarrow 0$$

Гипотеза отвергается, а точность чисел не позволяет вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть гипотезу (оно крайне близко к 0). \square

Задача 19. Построить критерий значимости χ^2 проверки сложной гипотезы согласия с нормальным распределением. Проверить гипотезу на уровне значимости α_2 . Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Решение.

$$H_0 : X_1, \dots, X_n \sim N(a\sigma^2)$$

$$\sum_{i=1}^k \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \rightarrow \chi_{\text{кр}}^2$$

Метод минимизации хи-квадрат:

$$\underset{a, \sigma^2}{\operatorname{argmin}} \sum_{i=1}^k \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \rightarrow \chi_{\text{кр}}^2$$

Задача реализована в R с помощью скрипта:

```
P <- -function(a){
p <- -0
p[1] <- -pnorm(2.9, a[1], a[2])
p[2] <- -pnorm(2.97, a[1], a[2]) - sum(p)
p[3] <- -pnorm(3.08, a[1], a[2]) - sum(p)
p[4] <- -pnorm(3.15, a[1], a[2]) - sum(p)
p[5] <- -1 - sum(p)
p}
X2 <- -function(a){g <- -n * P(a); f <- -(nu - g)^2/g; sum(f)}
nu <- -c(7, 15, 14, 9, 5)
a <- -c(mean(data), sqrt(var(data)))
XM <- -nlm(X2, a)
```

С помощью функции *nlm* был найден минимум χ^2 :

$$\chi_{\text{набл}}^2 = 3.928807, a = 3.0045, \sigma = 0.1086$$

$$\chi_{\text{набл}}^2 > \chi_{\text{кр}}^2 = 3.218876$$

Сложная гипотеза согласия с нормальным распределением отвергается.

$$P(x > 3.928807) = 0.1402395$$

Гипотеза может быть принята, если уровень значимости

$$\alpha \leq 0.1402395$$

□

Задача 20. Построить наиболее мощный критерий проверки простой гипотезы о нормальности с параметром $(a, \sigma^2) = (a_0, \sigma_0^2)$ при альтернативе нормальности с параметром $(a, \sigma^2) = (a_1, \sigma_1^2)$. Проверить гипотезу на уровне значимости α_2 . Что получится, если поменять местами основную и альтернативную гипотезы?

Решение.

Отношение правдоподобия ($\sigma_0 = \sigma_1$):

$$\begin{aligned} l(x) = \frac{L(x, a_1, \sigma_1^2)}{L(x, a_0, \sigma_0^2)} &= \frac{\sigma_0^n}{\sigma_1^n} \exp \left(\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - a_0)^2 - \sum_{i=1}^n (x_i - a_1)^2 \right) \right) = \\ &= \exp \left(\frac{1}{2\sigma_0^2} \sum_{i=1}^n -2x_i a_0 + a_0^2 + 2x_i a_1 - a_1^2 \right) \end{aligned}$$

Логарифмируем:

$$\exp \left(\frac{1}{2\sigma_0^2} \sum_{i=1}^n -2x_i a_0 + a_0^2 + 2x_i a_1 - a_1^2 \right) > c \rightarrow \frac{1}{2\sigma_0^2} \sum_{i=1}^n (a_0^2 - a_1^2) + 2x_i(a_1 - a_0) > \log c$$

Отсюда:

$$\frac{n(a_0^2 - a_1^2)}{2\sigma_0^2} + \frac{a_1 - a_0}{\sigma_0^2} \sum_{i=1}^n x_i > \log c$$

Подставив известные $n = 50, \sigma_0 = 0.10, a_0 = 2.60, a_1 = 3.00$ получаем:

$$-5600 + 40 \sum_{i=1}^n x_i > \log c \rightarrow \sum_{i=1}^n x_i > \frac{\log c + 5600}{40} = c^*$$

Тогда критерий примет вид:

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i > c^* \\ p, & \sum_{i=1}^n x_i = c^* \\ 0, & \sum_{i=1}^n x_i < c^* \end{cases}$$

С помощью R определим c^* и получим:

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i > 129 \\ p, & \sum_{i=1}^n x_i = 129 \\ 0, & \sum_{i=1}^n x_i < 129 \end{cases}$$

В нашем случае $\sum_{i=1}^n x_i = 150 > 129$, поэтому принимается основная гипотеза о нормальности с параметром (a_0, σ_0^2) .

Если поменять местами основную и альтернативную гипотезы, то будет принята альтернативная гипотеза о нормальности с параметром (a_1, σ_1^2) . \square

Задача 21. В пунктах (с) - (g) заменить семейство нормальных распределений на двухпараметрическое семейство распределений Лапласа.

Решение.

Плотность распределения Лапласа:

$$f(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}}{\sigma}|x - a|\right)$$

\square

Задача 22. В предположении, что исходные наблюдения являются выборкой из распределения семейства Лапласа, построить оценку максимального правдоподобия параметров (a, σ^2) , и соответствующие оценки по методу моментов. Найти смещение оценок.

Решение.

- Метод максимального правдоподобия

$$\begin{aligned}
 l(\vec{x}, a, \sigma^2) &= \frac{1}{\sigma^n 2^{n/2}} \exp\left(-\frac{\sqrt{2}}{\sigma} \sum_{i=1}^n |x_i - a|\right) \\
 ll(\vec{x}, a, \sigma^2) &= -n \log(\sigma) - \frac{n}{2} \log(2) - \frac{\sqrt{2}}{\sigma} \sum_{i=1}^n |x_i - a| = \\
 &= -n \log(\sigma) - \frac{n}{2} \log(2) - \frac{\sqrt{2}}{\sigma} \sum_{i=1}^k (a - x_{(i)}) - \frac{\sqrt{2}}{\sigma} \sum_{i=k+1}^n (x_{(i)} - a) = \\
 &= -n \log(\sigma) - \frac{n}{2} \log(2) - \frac{\sqrt{2}}{\sigma} k a + \frac{\sqrt{2}}{\sigma} \sum_{i=1}^k x_{(i)} - \frac{\sqrt{2}}{\sigma} \sum_{i=k+1}^n x_{(i)} + \frac{\sqrt{2}}{\sigma} (n - k - 1) a = \\
 &= -n \log(\sigma) - \frac{n}{2} \log(2) + \frac{\sqrt{2}}{\sigma} \sum_{i=1}^k x_{(i)} - \frac{\sqrt{2}}{\sigma} \sum_{i=k+1}^n x_{(i)} + \frac{\sqrt{2}}{\sigma} (n - k - 1) a + \frac{\sqrt{2}}{\sigma} (n - 2k - 1) a
 \end{aligned}$$

$$\frac{\partial ll(\vec{x}, a, \sigma^2)}{\partial a} = \frac{\sqrt{2}}{\sigma} (n - 2k - 1)$$

$$\frac{\sqrt{2}}{\tilde{\sigma}} (n - 2k - 1) = 0$$

$$k = \frac{n - 1}{2} - \text{выборочная медиана}$$

$$\tilde{a} \in (x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}) = (x_{(25)}, x_{(26)}) = (2.987, 2.998) = 2.9925$$

$$\frac{\partial ll(\vec{x}, a, \sigma^2)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sqrt{2}}{\sigma^2} \sum_{i=1}^n |x_i - a|$$

$$-\frac{n}{\tilde{\sigma}} + \frac{\sqrt{2}}{\tilde{\sigma}^2} \sum_{i=1}^n |x_i - \tilde{a}| = 0$$

$$\tilde{\sigma} = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{a}|$$

Оценка максимального правдоподобия:

$$(\tilde{a}, \tilde{\sigma}) = (2.9925, 0.006599938)$$

• Метод моментов

$$\mathbb{E}(x) = a \rightarrow \hat{a} = \bar{x} = 3.0116$$

$$\mathbb{D}(x) = \frac{2}{\frac{2}{\sigma^2}} = \sigma^2 \rightarrow \hat{\sigma}^2 = s^2 = 0.00970184$$

$$(\hat{a}, \hat{\sigma}^2) = (\tilde{a}, \tilde{\sigma}^2) = (3.0116, 0.00970184)$$

$$\mathbb{E}(\hat{a}) = \mathbb{E}(\bar{x}) = a \rightarrow \text{несмещенная оценка}$$

$$\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}(s^2) = \frac{n-1}{n}\sigma^2 \rightarrow \mathbb{E}(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \rightarrow \text{смещенная оценка}$$

□

Задача 23. Построить доверительные интервалы уровня значимости α_2 для параметров (a, σ^2)

Решение.

Параметр a :

$$\frac{\sum_{i=1}^n x_i - na}{\sqrt{n\sigma^2}} \rightarrow N(0, 1); \frac{\sum_{i=1}^n x_i - na}{\sqrt{ns^2}} \rightarrow N(0, 1)$$

$$\text{Выберем } t_\gamma : \phi(t_\gamma) = 1 - \frac{\alpha_2}{2} \Rightarrow t_\gamma = 1.281552$$

$$P\left(-t_\gamma \leq \frac{\sqrt{n}(\bar{x} - a)}{s} \leq t_\gamma\right) = 1 - \alpha_2 = P\left(\bar{x} - \frac{t_\gamma s}{\sqrt{n}} \leq x \leq \bar{x} + \frac{t_\gamma s}{\sqrt{n}}\right)$$

Отсюда асимптотический ДИ уровня значимости 0.2:

$$[2.993748, 3.029452]$$

Параметр σ^2 :

$$\sqrt{n}(\tilde{\sigma} - \sigma) \sim N\left(0, \frac{\sigma^2}{2}\right)$$

$$\frac{\sqrt{2n}}{\frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i - \tilde{a}|} \left(\frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i - \tilde{a}| - \sigma \right) \sim N(0, 1)$$

Найдем ДИ:

$$P \left(-t_\gamma \leq \frac{\sqrt{2n}}{\frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i - \tilde{a}|} \left(\frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i - \tilde{a}| - \sigma \right) \leq t_\gamma \right) = 1 - \alpha_2 =$$

$$= P \left(\frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i - \tilde{a}| - \frac{t_\gamma \sum_{i=1}^n |x_i - \tilde{a}|}{n^{\frac{3}{2}}} \leq \sigma^2 \leq \frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i - \tilde{a}| + \frac{t_\gamma \sum_{i=1}^n |x_i - \tilde{a}|}{n^{\frac{3}{2}}} \right)$$

Вычислим ДИ:

$$[0.00903340, 0.01679993]$$

□

Задача 24. С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с распределением из семейства Лапласа с параметрами a_0, σ_0^2 . Проверить гипотезу на уровне значимости α_2 . Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

Решение.

Простая гипотеза $H_0 : a = a_0 = 2.60, \sigma^2 = \sigma_0^2 = 0.10$

Согласно теореме Колмогорова, при справедливости гипотезы:

$\sqrt{n} \sup |F_n(x) - F_0(x)| \rightarrow K(x)$, где $K(x)$ - функция Колмогорова

$$K(C_{\alpha_2}) = 1 - \alpha_2 = 1 - 0.2 = 0.8$$

$$C_{\alpha_2} = C_{0.2} = 1.0493$$

$$H_0 : X_1, \dots, X_n \sim Laplace(2.60, 0.10)$$

С помощью R вычислим:

$$\sup |F_n(x) - F_0(x)| = 0.9390869$$

$$\sqrt{n} \sup |F_n(x) - F_0(x)| = 6.640347 > 1.0493$$

Отвергаем гипотезу H_0

$$p - value = 2.2 \cdot 10^{-16}$$

Наибольшее значение уровня значимости, на котором нет оснований отвергнуть гипотезу

□

Задача 25. Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением из семейства Лапласа с параметрами (a_0, σ_0^2) . Проверить гипотезу на уровне значимости α_2 . Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

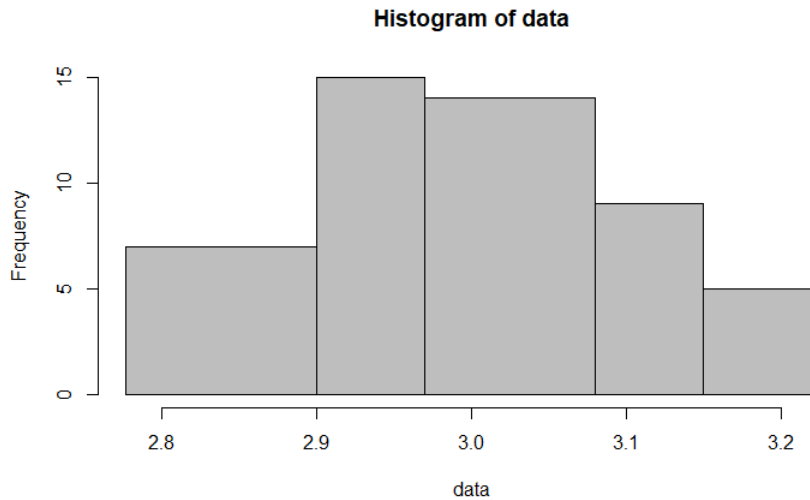
Решение.

$$H_0 : X_1, \dots, X_n \sim Laplace(2.6, 0.1)$$

$$\sum_{i=1}^k \frac{(m_i - m'_i)^2}{m_i} \rightarrow \chi_{кр}^2$$

$$\chi_{кр}^2 = \chi_4^2 = 5.988617$$

Перестроим гистограмму частот, выбрав следующие точки: $(-Inf, 2.9, 2.97, 3.08, 3.15, Inf)$



Интервал	$(-\infty; 2.9]$	$(2.9; 2.97]$	$(2.97; 3.08]$	$(3.08; 3.15]$	$(3.15; \infty)$	Σ
m_i	7	15	14	9	5	50
p_i	0.9928152020	0.0045149597	0.0021063547	0.0003540957	0.0002093880	1
$\frac{(m_i - m'_i)^2}{m'_i}$	36.62785	966.91238	1833.14041	4557.05246	2377.92240	$\chi_{набл}^2$

$$\chi_{набл}^2 = 9771.655 > \chi_{кр}^2 = 5.988617$$

$$P(\chi_{набл}^2 > 9771.655) \rightarrow 0$$

Гипотеза отвергается, а точность чисел не позволяет вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть гипотезу (оно крайне близко к 0). \square

Задача 26. Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением из семейства Лапласа. Проверить гипотезу на уровне значимости α_2 . Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Решение.

$$H_0 : X_1, \dots, X_n \sim Laplace(a\sigma^2)$$

$$\sum_{i=1}^k \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \rightarrow \chi_{кр}^2$$

Метод минимизации хи-квадрат:

$$\underset{a, \sigma^2}{argmin} \sum_{i=1}^k \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \rightarrow \chi_{кр}^2$$

С помощью функции *nlm* был найден минимум χ^2 :

$$\chi_{набл}^2 = 7.530788, a = 2.9898297, \sigma = 0.1375121$$

$$\chi_{набл}^2 > \chi_{кр}^2 = 3.218876$$

Сложная гипотеза согласия с нормальным распределением отвергается.

$$P(x > 7.530788) = 0.02315849$$

Гипотеза может быть принята, если уровень значимости

$$\alpha \leq 0.02315849$$

□