МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра ВМ-2

ОТЧЕТ

по индивидуальному заданию №2
по дисциплине «Статический анализ»
Вариант №13

Студент гр. 8382	____________________	Мирончик П.Д.
Преподаватель	____________________	Малов С.В.

Санкт-Петербург
2020

ПОСТАНОВКА ЗАДАЧИ

Вар. 13 (83822020)

1. В результате эксперимента получены данные, приведенные в таблице 1.

a) Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.

b) Вычислить выборочные аналоги следующих числовых характеристик:
(i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса,
(vi) вероятности $\mathbf{P}(X \in [a, b])$.

c) 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.

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

e) 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.

f) 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.

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

h) В пунктах (c)–(f) заменить семейство распределений Пуассона на семейство геометрических распределений

$$\mathbf{P}_\lambda(X = k) = \frac{\lambda^k}{(\lambda + 1)^{k+1}}, \ k = 0, 1, \ldots .$$

Таблица 1 $\alpha_1 = 0.10$; $a = 0.00$; $b = 1.79$; $\lambda_0 = 0.60$; $\lambda_1 = 1.40$.
0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 2 0 0 2 0 1 0 3 0 2 0 0 3 0 3 3 0 4 0 3 1 0 2 0 0 0 0 1 0 1 0 1 0
0 0 0

2. В результате эксперимента получены данные, приведенные в таблице 2.

a) Построить вариационный ряд, эмпирическую функцию распределения, гистограмму и полигон частот с шагом h.

b) Вычислить выборочные аналоги следующих числовых характеристик:
(i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса,
(vi) вероятности $\mathbf{P}(X \in [c, d])$.

c) В предположении, что исходные наблюдения являются выборкой из показательного распределения, построить оценку максимального правдоподобия параметра λ и соответствующую оценку по методу моментов. Найти смещение оценок.

d) Построить асимптотический доверительный интервал уровня значимости α_2 для параметра λ на базе оценки максимального правдоподобия.

e) С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с показательным распределением с параметром λ_0. Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

f) Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с показательным распределением с параметром λ_0. Проверить гипотезу на уровне α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

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

h) Построить наиболее мощный критерий проверки простой гипотезы о показательности с параметром $\lambda = \lambda_0$ при альтернативе показательности с параметром $\lambda = \lambda_1$. Проверить гипотезу на уровне значимости α_2. Что получится, если поменять местами основную и альтернативную гипотезы?

i) В пунктах (c)–(h) заменить семейство показательных распределений на семейство гамма-распределений с плотностями $f(x) = \dfrac{\sqrt{\lambda} e^{-\lambda x/2}}{\sqrt{2\pi x}}$ (использовать таблицу распределений χ^2_1).

Таблица 2 $\alpha_2 = 0.01$; $c = 0.00$; $d = 2.00$; $h = 0.40$; $\lambda_0 = 0.50$; $\lambda_1 = 1.00$.
0.589 0.596 0.215 0.664 0.323 0.005 2.487 0.662 0.584 0.031 1.136 1.494 2.698 1.516 0.839 0.186 2.173 1.014 0.198 0.344 2.269
0.209 1.609 0.803 0.228 1.382 1.763 0.413 2.474 0.034 3.198 3.552 0.165 0.570 0.484 1.558 1.381 0.117 0.034 0.536 2.455 1.765
0.973 0.044 3.889 0.327 0.013 0.394 0.141 0.354

ХОД РЕШЕНИЯ

1.

a) Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.

Построение ряда:

```
x <- c(0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 0, 2, 0, 1, 0,
3, 0, 2, 0, 0, 3, 0, 3, 3, 0, 4, 0, 3, 1, 0, 2, 0, 0, 0, 0, 1, 0, 1,
0, 1, 0, 0, 0, 0)
vararr <- sort(x)
```

Полученный вариационный ряд

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

Построение эмпирической ф.р.:

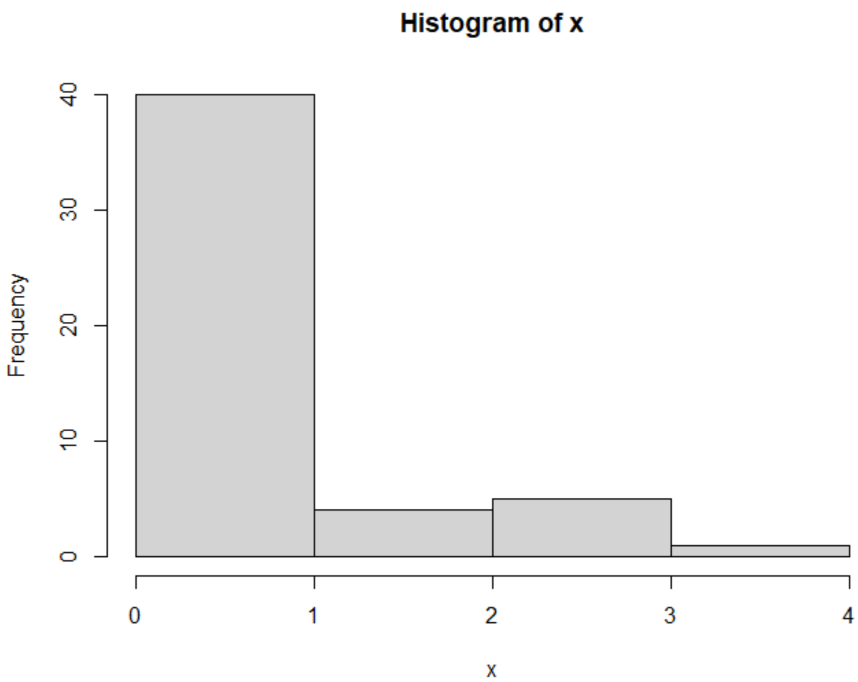
```
F <- function(x, t) {
  z <- x[x<t]
  return(length(z) / length(x))
}
```

Полученные значения

```
0,          x <= 0
0.64,       0 < x <=1,
0.8,        1 < x <= 2,
0.88,       2 < x <= 3,
0.98,       3 < x <= 4
1,          x > 4
```

Гистограмма частот:

```
hist(x, breaks = c(min(x):max(x)), right = TRUE)
```



b) Вычислить выборочные аналоги следующих числовых характеристик:

(i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асиметрии, (v) эксцесса, (vi) вероятности P(X in [a,b])

(i) математическое ожидание

```
meanX <- sum(x) / length(x)
```

получили 0.7

(ii) дисперсия

```
variance <- sum(x^2) / length(x) - meanX^2
```

получили 1.21

(iii) медиана

```
if (length(x) %% 2 != 0) {
  median <- vararr[(length(vararr) - 1) / 2]
} else {
  median <- (vararr[length(vararr) / 2] + vararr[length(vararr) / 2 +
1]) / 2
}
```

получили 0

(iv) ассиметрия

```
assym <- sum((x - meanX)^3) / (length(x) * variance^(3/2))
```

получили 1.424493

(v) эксцесс

```
excess <- (sum((x - meanX)^4) / (length(x) * variance^2)) - 3
```

получили 0.7932518

(vi) вероятность P(X in [a,b])

```
Px <- function (a, b) {
  return(F(x, b) - F(x, a))
}
```

a=0, b=1.79

получим 0.8

c) 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.

ОМП:

```
library("maxLik")
LL<-function(t){sum(dpois(x,t[1],log=TRUE))}
maxLL<-maxNR(LL,start=c(1))
OMP<-maxLL$estimate
```

полученное ОМП: 0.7

оценка λ по методу моментов:

```
mean<-sum(x)/length(x)
```

получили также 0.7

Плотность распределения Пуассона: $p(k) = \frac{e^{-\lambda}\lambda^k}{k!}$

Метод моментов:

$$E(x_1) = \lambda$$

$$\bar{x} = E(x_1)$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$E(\hat{\lambda}) = E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{n\lambda}{n} = \lambda = E(\lambda)$$

Значит оценка несмещенная

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

Так как x_i имеет распределение Пуассона, то $I_1(\lambda) = \frac{1}{\lambda}$. Из этого следует, что $I(\lambda) = n * I_1(\lambda) = \frac{n}{\lambda}$.

По методу максимального правдоподобия:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$\sqrt{I(\lambda)}(\hat{\lambda} - \lambda) => N(0,1)$$

$$\sqrt{n * I_1(\lambda)}(\hat{\lambda} - \lambda) => N(0,1)$$

$$\sqrt{\frac{n}{\bar{x}}}(\bar{x} - \lambda) => N(0,1),\ \alpha_1 = 0{,}10$$

$$p(T_1(\vec{x}) \le \lambda \le T_2(\vec{x})) = 1 - \alpha_1$$

$$p\left(-x_\alpha \le \sqrt{\frac{n}{\bar{x}}}(\bar{x} - \lambda) \le x_\alpha\right) = \Phi(x_\alpha) - \Phi(-x_\alpha) = 2 * \Phi(x_\alpha) - 1 = 1 - \alpha_1$$

где $\Phi(x_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_\alpha} e^{-\frac{t^2}{2}} dt$ - квантиль порядка x_α стандартного нормального закона распределения.

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

$$p\left(\bar{x} - x_\alpha \sqrt{\frac{\bar{x}}{n}} \leq \lambda \leq \bar{x} + x_\alpha \sqrt{\frac{\bar{x}}{n}}\right) = 1 - \alpha_1$$

Код:

```
TI <- function(a1) {
  mean<-sum(x)/length(x)
  xa<-qnorm (1-a1/2)
  return(c(
    mean-xa*sqrt(mean/length(x)),
    mean+xa*sqrt(mean/length(x))
  ))
}
```

Результат для a1=0.1: [0.5053783, 0.8946217] 3-способа

e) Используя гистограмму частот, построить критерий значимости $\chi 2$ 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.

Таблица частот:

```
table(x)
```

получим

0	1	2	3	4
32	8	4	5	1

Простая гипотеза H_0: $p(x) = \frac{\lambda_0^x}{x!} e^{-\lambda_0}, \ \lambda_0 = 0.6$

x	**n_i**	**p_i**	**np_i**	**n_i-np_i**	$\frac{(n_i - np_i)^2}{np_i}$
0	32	0.549	27.45	4.55	0.754
1	8	0.329	16.45	-8.45	4.341

2	4	0.099	4.95	-0.95	0.182
3	5	0.020	1	4	16
>=4	1	0.003	0.15	0.85	4.817
sum	50	1	50	0	26.094

$\chi^2_{0.1,4} = 7{,}7794$, следовательно мы не принимаем гипотезу.

Наибольший уровень значимости:

```
r<-5
Xi2<-26.094
maxval <- pchisq(Xi2,r-1,lower.tail = FALSE);
```

Получим 0.00003 (по таблице распределения это 26.11493)

f) Построить критерий значимости χ2 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.

$\cancel{\lambda = \bar{x} = 0.7}$ Мультиномиальная ОМП

x	**n_i**	**p_i**	**np_i**	**n_i-np_i**	$\frac{(n_i - np_i)^2}{np_i}$
0	32	0.497	24.829	7.171	2.0709206
1	8	0.348	17.38	-9.38	5.0627763
2	4	0.122	6.083	-2.083	0.7133776
3	5	0.0284	1.419	3.581	9.0324038
>=4	1	0.005	0.15	0.752	2.2742241
sum	50	1	50	0	19.1537

Объединить зоны

$\chi^2_{0.1,3} = 6{,}251$, значит не принимаем гипотезу.

Наибольший уровень значимости:

```
r<-5
Xi2<-19.1537
maxval <- pchisq(Xi2,r-2,lower.tail = FALSE);
print(maxval)
```

результат: 0.00025 ((по таблице распределения: $\chi^2_{0.00025,3} = 19.18788$)

g) Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром λ = λ₀ при альтернативе пуассоновости с параметром λ = λ₁. Проверить гипотезу на уровне значимости α₁. Что получится, если поменять местами основную и альтернативную гипотезы?

Основная гипотеза $H_0: \lambda = \lambda_0 = 0.6$, альтернативная гипотеза $H_1: \lambda = \lambda_1 = 1.4$.

Согласно лемме Неймана-Пирсона:

$$\varphi(x) = \begin{cases} 0, l(x; \lambda_0, \lambda_1) < \text{C} \\ p, l(x; \lambda_0, \lambda_1) = \text{C}\ , p \in [0,1) \\ 1, l(x; \lambda_0, \lambda_1) > \text{C} \end{cases}$$

Функция правдоподобия для распределения Пуассона (была найдена ранее):

$$L(\vec{x}, \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} * e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

Найдём $l(x; \lambda_0, \lambda_1)$:

$$l(x; \lambda_0, \lambda_1) = \frac{L(x; \lambda_1)}{L(x; \lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^{n} x_i} e^{n*(\lambda_0 - \lambda_1)} = (2.33)^{\sum_{i=1}^{n} x_i} e^{-0.8n}$$

Прологарифмируем полученное выражение:

$$\left(\sum_{i=1}^{n} x_i\right) * \log(2.33) - 0.8n = \log\,(C)$$

$$\sum_{i=1}^{n} x_i = \frac{\log(C) + 0.8n}{\log(2.33)}$$

Обозначим $\tilde{\text{C}} = \frac{\log(C)+0.8n}{\log(2.33)}$. Тогда критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < \tilde{C} \\ p, \sum_{i=1}^{n} x_i = \tilde{C} \\ 1, \sum_{i=1}^{n} x_i > \tilde{C} \end{cases}$$

Из этого следует уравнение из которого можно найти p и $\tilde{C}$:

$$1 * P_{\lambda_0}(l(x;\lambda_0,\lambda_1) > C) + p * P_{\lambda_0}(l(x;\lambda_0,\lambda_1) = C) = 1 * P_{\lambda_0}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) +$$

$$+p * P_{\lambda_0}\left(\sum_{i=1}^{n} x_i = \tilde{C}\right) = \alpha_1 = 0{,}1$$

Из того, что $x_i \sim Pois(\lambda_0)$, можно сделать вывод, что $\sum_{i=1}^{n} x_i \sim Pois(n\lambda_0)$. Воспользуемся тем, что $\sum_{i=1}^{n} x_i \in N_0$ и найдем перебором $\tilde{C}$:

```
lambda<-0.6
a1<-0.1
c<-0
a0<-ppois(c,lambda*length(x))
while (a0<a1)
{
  c<-c+1;
  a0<-ppois(c,lambda*length(x))
}
c<-c-1
print(c)
p<-(a1-ppois(c,lambda*length(x)))/dpois(c,lambda*length(x))
print(p)
print(a0)
```

Критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i > 22 \\ 0.743752, \sum_{i=1}^{n} x_i = 22 \\ 1, \sum_{i=1}^{n} x_i < 22 \end{cases}$$

Так как сумма $\sum_{i=1}^{n} x_i = 35\ (> 22)$, **принимаем альтернативную гипотезу H_1**.

Поменяем местами основную и альтернативную теории и проведём аналогичные вычисления:

Найдём $l(x; \lambda_1, \lambda_0)$:

$$l(x; \lambda_0, \lambda_1) = \frac{L(x; \lambda_1)}{L(x; \lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^{n} x_i} e^{n*(\lambda_0 - \lambda_1)} = (0.43)^{\sum_{i=1}^{n} x_i} e^{0.8n}$$

Прологарифмируем полученное выражение:

$$-\left(\sum_{i=1}^{n} x_i\right) * \log(2.33) + 0.8n = \log{(C)}$$

$$\sum_{i=1}^{n} x_i = \frac{0.8n - \log(C)}{\log(2.33)}$$

Обозначим $\tilde{C} = \frac{0.8n - \log(C)}{\log(2.33)}$. Тогда критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < \tilde{C} \\ p, \sum_{i=1}^{n} x_i = \tilde{C} \\ 1, \sum_{i=1}^{n} x_i > \tilde{C} \end{cases}$$

Из этого следует уравнение:

$$1 * P_{\lambda_0}(l(x; \lambda_0, \lambda_1) > C) + p * P_{\lambda_0}(l(x; \lambda_0, \lambda_1) = C) = 1 * P_{\lambda_0}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) +$$

$$+ p * P_{\lambda_0}\left(\sum_{i=1}^{n} x_i = \tilde{C}\right) = \alpha_1 = 0{,}1$$

Из того, что $x_i \sim Pois(\lambda_0)$, можно сделать вывод, что $\sum_{i=1}^{n} x_i \sim Pois(n\lambda_0)$. Воспользуемся тем, что $\sum_{i=1}^{n} x_i \in N_0$ и найдем перебором $\tilde{C}$:

```
lambda<-1.4
a1<-0.1
```

```
c<-0
a0<-ppois(c,lambda*length(x))
while (a0<a1)
{
  c<-c+1;
  a0<-ppois(c,lambda*length(x))
}
c<-c-1
print(c)
p<-(a1-ppois(c,lambda*length(x)))/dpois(c,lambda*length(x))
print(p)
print(a0)
```

Критерий примет вид:

$$\varphi(x) = \begin{cases} 0, & \sum_{i=1}^{n} x_i > 58 \\ 1.049733, & \sum_{i=1}^{n} x_i = 58 \\ 1, & \sum_{i=1}^{n} x_i < 58 \end{cases}$$

Так как сумма $\sum_{i=1}^{n} x_i = 35\ (< 58)$, **принимаем основную гипотезу** $\boldsymbol{H_0}$.

h) В пунктах (c)-(f) заменить семейство распределений Пуассона на семейство геометрических распределений.

h-c) 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.

Нахождение оценки максимального правдоподобия параметра λ:

Плотность геометрического распределения: $p(x = k) = \frac{\lambda^k}{(\lambda+1)^{k+1}}, k = 0,1, \ldots$

$$p(x = k) = \left(\frac{\lambda}{\lambda + 1}\right)^k \left(\frac{1}{\lambda + 1}\right), k = 0,1, \ldots$$

Обозначим $q = \left(\frac{\lambda}{\lambda+1}\right)$, и $p = \left(\frac{1}{\lambda+1}\right), k = 0,1, \ldots$

$$L(\vec{x},\lambda)=\left(\frac{\lambda}{\lambda+1}\right)^{\sum_{i=1}^{n}x_i}\left(\frac{1}{\lambda+1}\right)^{n},k=0,1,\dots$$

$$LL(\vec{x},\lambda)=\log(\lambda)\sum_{i=1}^{n}x_i-\log\,(\lambda+1)\sum_{i=1}^{n}x_i-n\log(\,\lambda+1)$$

$$\frac{\partial}{\partial\lambda}LL(\vec{x},\lambda)=\frac{\sum_{i=1}^{n}x_i}{\lambda}-\frac{\sum_{i=1}^{n}x_i}{\lambda+1}-\frac{n}{\lambda+1}=0$$

$$\hat{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}=\bar{x}=0.7$$

Метод моментов:

Плотность геометрического распределения: $p(x=k)=\frac{\lambda^k}{(\lambda+1)^{k+1}},k=0,1,\dots$

$$E(x_1)=\frac{q}{p}=\frac{\left(\frac{\lambda}{\lambda+1}\right)}{\left(\frac{1}{\lambda+1}\right)}=\lambda$$

$$\bar{x}=E(x_1)$$

$$\hat{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}=\bar{x}=0.7$$

Нахождение смещения оценок:

$$E(\hat{\lambda})=E(\bar{x})=\frac{1}{n}\sum_{i=1}^{n}E(x_i)=\frac{n\lambda}{n}=\lambda=E(\lambda)$$

Следовательно, оценка $\hat{\lambda}=\bar{x}$ несмещённая.

h-d) Построить асимптотический доверительный интервал уровня значимости α1 для параметра λ на базе оценки максимального правдоподобия.

x_i -> Geom(λ)

Найдём информацию Фишера:

$$p(x,\lambda)=\frac{\lambda^x}{(\lambda+1)^{x+1}}$$

$$\ln\big(p(x,\lambda)\big)=x*\ln(\lambda)-(x+1)*\ln\,(\lambda+1)$$

$$\frac{\partial^2}{\partial\lambda^2}\ln\big(p(x,\lambda)\big)=-\frac{x}{\lambda^2}-\frac{(x+1)}{(\lambda+1)^2}$$

$E(x) = \lambda$

$$I_1(\lambda) = -E_\lambda\left(\frac{\partial^2 \ln\big(p(x,\lambda)\big)}{\partial \lambda^2}\right) = -E_\lambda\left(-\frac{x}{\lambda^2} - \frac{(x+1)}{(\lambda+1)^2}\right) = \frac{1}{\lambda} - \frac{1}{\lambda+1} = \frac{1}{\lambda(\lambda+1)}$$

$$I(\lambda) = n * I_1(\lambda) = \frac{n}{\lambda(\lambda+1)}$$

Оценка максимального правдоподобия параметра λ:

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

$\sqrt{I(\lambda)}\,\big(\hat{\lambda} - \lambda\big) => N(0,1)$

Эксперимент регулярен, значит, подстановка ОМП вместо параметра в информацию Фишера не нарушает асимптотической нормальности.

$\sqrt{n * I_1(\lambda)}\,\big(\hat{\lambda} - \lambda\big) => N(0,1)$

$\sqrt{\frac{n}{\bar{x}(\bar{x}+1)}}\;(\bar{x} - \lambda) => N(0,1),$ при $\alpha_1 = 0{,}10$

$p\big(T_1(\vec{x}) \le \lambda \le T_2(\vec{x})\big) = 1 - \alpha_1$

$$p\left(-x_\alpha \le \sqrt{\frac{n}{\bar{x}(\bar{x}+1)}}(\bar{x} - \lambda) \le x_\alpha\right) = \Phi(x_\alpha) - \Phi(-x_\alpha) = 2 * \Phi(x_\alpha) - 1 =$$

$= 1 - \alpha_1$

где $\Phi(x_\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_\alpha} e^{-\frac{t^2}{2}}dt$ - квантиль порядка x_α стандартного нормального закона распределения.

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

$$p\left(\bar{x} - x_\alpha\sqrt{\frac{\bar{x}(\bar{x}+1)}{n}} \le \lambda \le \bar{x} + x_\alpha\sqrt{\frac{\bar{x}(\bar{x}+1)}{n}}\right) = 1 - \alpha_1$$

Выполним код для a1=0.1

```
TIf <- function(a1) {
  mean<-sum(x)/length(x)
  xa<-qnorm (1-a1/2)
  return(c(
    mean-xa*sqrt((mean*(mean+1))/length(x)),
    mean+xa*sqrt((mean*(mean+1))/length(x))
  ))
}
```

получим [0.4462443, 0.9537557]

f-e) Используя гистограмму частот, построить критерий значимости $\chi 2$ проверки простой гипотезы согласия с геометрическим распределением с параметром λ_0. Проверить гипотезу на уровне значимости α_1. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Таблица частот имеет вид

0	1	2	3	4
32	8	4	5	1

Простая гипотеза H_0: $p(x) = \frac{{\lambda_0}^x}{(\lambda_0+1)^{x+1}}, \lambda_0 = 0.6$

x	**n_i**	**p_i**	**np_i**	**n_i-np_i**	$\frac{(n_i - np_i)^2}{np_i}$
0	32	0.625	31.2500000	0.75000000	0.0180000000
1	8	0.234	11.7187500	-3.7187500	1.1800833333
2	4	0.088	4.3945312	-0.3945312	0.0354201389
3	5	0.033	1.6479492	3.35205078	6.8183195891
>=4	1	0.020	0.9887695	0.01123047	0.0001275559
sum	50	1	50	0	8.051951

Для данного эксперимента $\chi^2_{0.1,4} = 7{,}7794$, что меньше полученного значения, следовательно, мы не принимаем гипотезу H_0.

Наибольший уровень значимости:

```
r<-5
Xi2<-8.051951
maxval <- pchisq(Xi2,r-1,lower.tail = FALSE)
print(maxval)
```

Полученное значение 0.0897 (по таблице распределения: $\chi^2_{0.138,4} = 8.052$).

h-f) Построить критерий значимости $\chi 2$ проверки сложной гипотезы согласия с геометрическим распределением. Проверить

гипотезу на уровне значимости α1. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

$\lambda = \bar{x} = 0.7$

Основная гипотеза H0: X ~ Geom(λ).

x	**n_i**	**p_i**	**np_i**	**n_i-np_i**	$\frac{(n_i - np_i)^2}{np_i}$
0	32	0.58823529	29.411765	2.5882353	0.2277647
1	8	0.24221453	12.110727	-4.1107266	1.3952981
2	4	0.09973540	4.986770	-0.9867698	0.1952596
3	5	0.04106752	2.053376	2.9466242	4.2284487
>=4	1	0.02874726	1.437363	-0.4373631	0.1330815
sum	50	1	50	0	6.179853

Для данного эксперимента $\chi^2_{0.1,3} = 6{,}251$, что больше полученного значения, следовательно, мы принимаем гипотезу H_0.

Наибольший уровень значимости:

```
r<-5
Xi2<-6.179853
maxval <- pchisq(Xi2,r-2,lower.tail = FALSE)
print(maxval)
```

получим 0.103 (по таблице распределения: $\chi^2_{0.103,3} = 6.179854$).

2.

а) Построить вариационный ряд, эмпирическую функцию распределения, гистограмму и полигон частот с шагом h.

Вариационный ряд:

```
x = c(0.589, 0.596, 0.215, 0.664, 0.323, 0.005, 2.487, 0.662, 0.584,
0.031, 1.136, 1.494, 2.698, 1.516, 0.839, 0.186, 2.173, 1.014, 0.198,
0.344, 2.269, 0.209, 1.609, 0.803, 0.228, 1.382, 1.763, 0.413, 2.474,
0.034, 3.198, 3.552, 0.165, 0.570, 0.484, 1.558, 1.381, 0.117, 0.034,
0.536, 2.455, 1.765, 0.973, 0.044, 3.889, 0.327, 0.013, 0.394, 0.141,
0.354)
vararr = sort(x)
```

Получим

```
0.005 0.013 0.031 0.034 0.034 0.044 0.117 0.141 0.165 0.186 0.198
0.209 0.215 0.228 0.323 0.327 0.344 0.354 0.394 0.413 0.484 0.536
```

```
0.570 .584 0.589 0.596 0.662 0.664 0.803 0.839 0.973 1.014 1.136 1.381
1.382 1.494 1.516 1.558 1.609 1.763 1.765 2.173 2.269 2.455 2.474
2.487 2.698 3.198 3.552 3.889
```

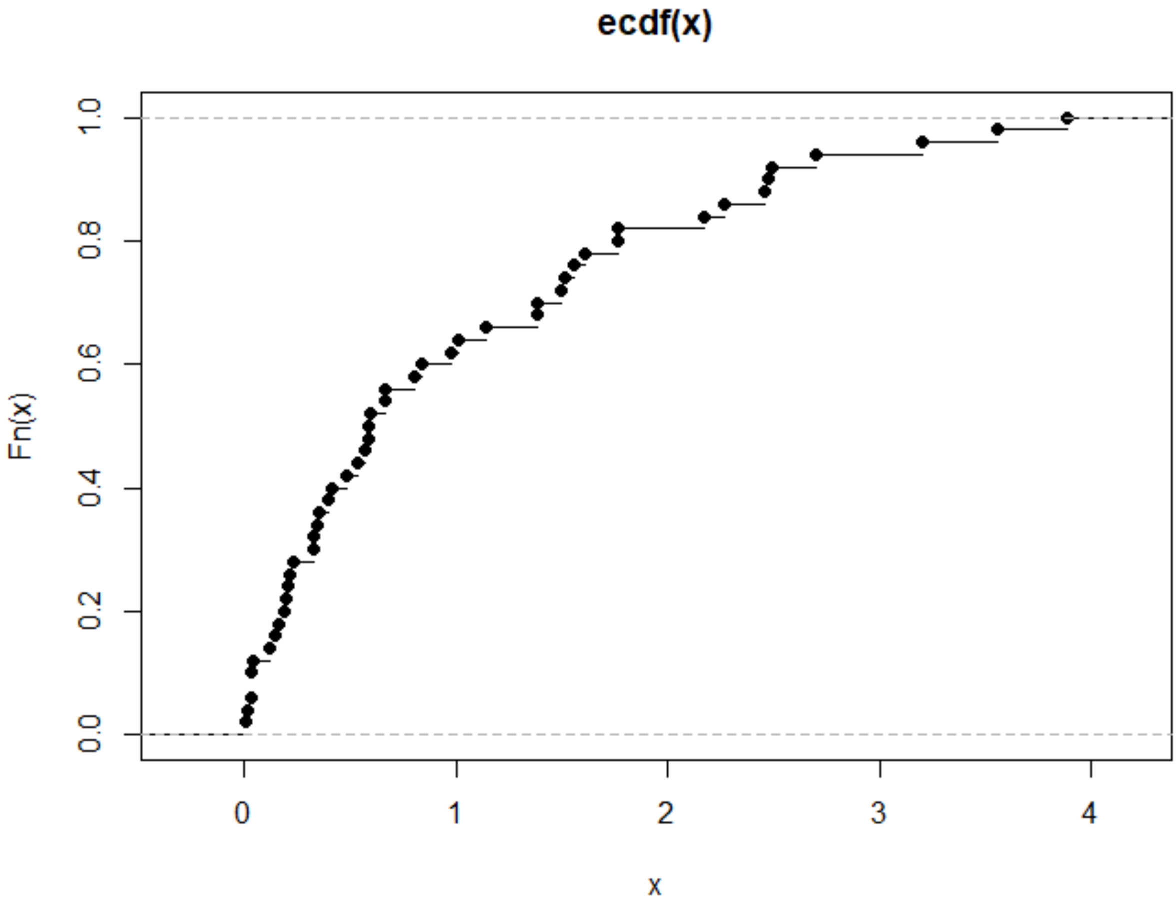
Эмпирическая функция распределения:

```
x < 0.005: 0
x < 0.013: 0.02
x < 0.031: 0.04
x < 0.034: 0.06
x < 0.044: 0.1
x < 0.117: 0.12
x < 0.141: 0.14
x < 0.165: 0.16
x < 0.186: 0.18
x < 0.198: 0.2
x < 0.209: 0.22
x < 0.215: 0.24
x < 0.228: 0.26
x < 0.323: 0.28
x < 0.327: 0.3
x < 0.344: 0.32
x < 0.354: 0.34
x < 0.394: 0.36
x < 0.413: 0.38
x < 0.484: 0.4
x < 0.536: 0.42
x < 0.57: 0.44
x < 0.584: 0.46
x < 0.589: 0.48
x < 0.596: 0.5
x < 0.662: 0.52
x < 0.664: 0.54
x < 0.803: 0.56
x < 0.839: 0.58
x < 0.973: 0.6
x < 1.014: 0.62
x < 1.136: 0.64
x < 1.381: 0.66
x < 1.382: 0.68
x < 1.494: 0.7
x < 1.516: 0.72
x < 1.558: 0.74
x < 1.609: 0.76
x < 1.763: 0.78
x < 1.765: 0.8
x < 2.173: 0.82
x < 2.269: 0.84
x < 2.455: 0.86
x < 2.474: 0.88
x < 2.487: 0.9
x < 2.698: 0.92
x < 3.198: 0.94
x < 3.552: 0.96
x < 3.889: 0.98
x >= 3.889: 1
```

График:

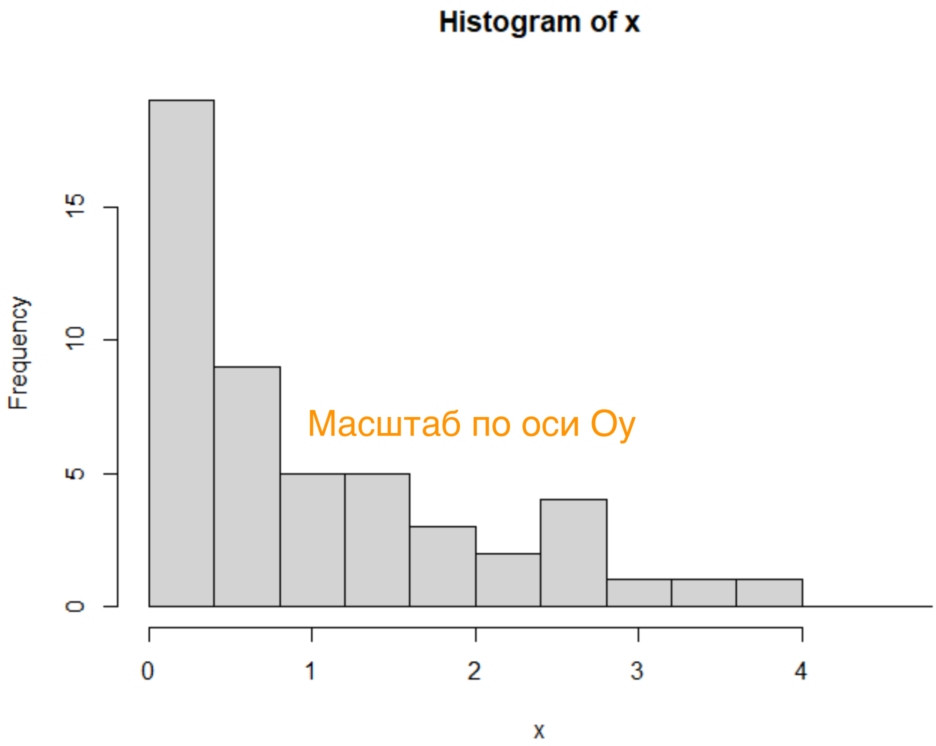
```
plot(ecdf(x))
```

Гистограмма



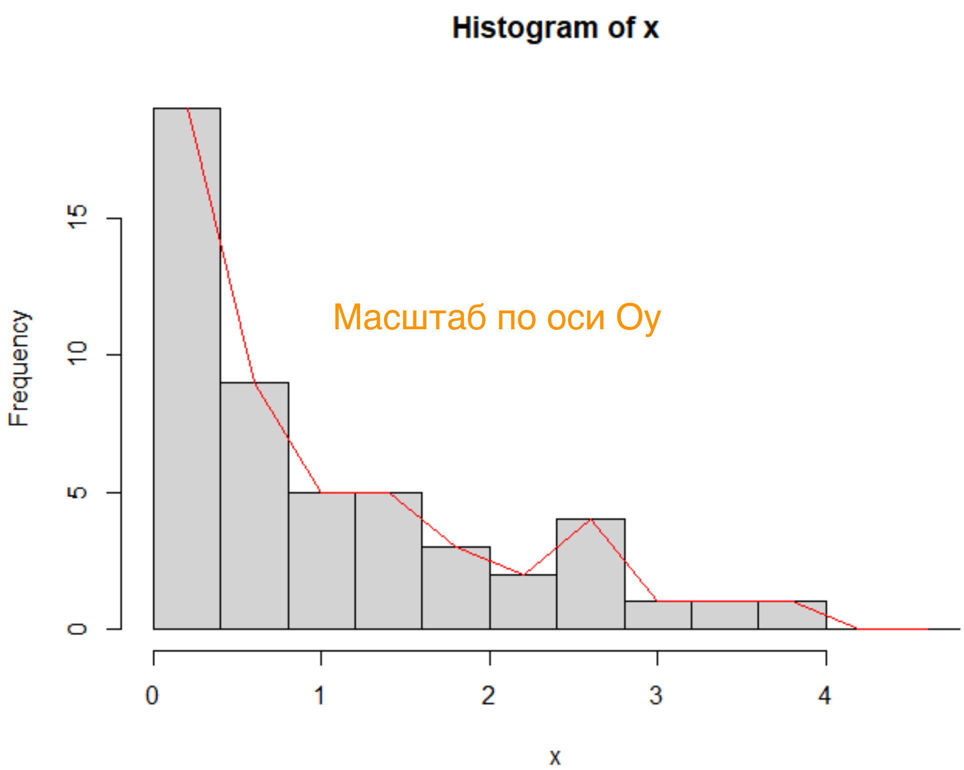
Гистограмма частот:

```
h = 0.4
hist(x,breaks=seq(0, max(x)+1, by=h))
```



Построение полигона:

```
h<-0.4
pol<-hist(x,breaks=seq(0, max(x)+1, by=h))
lines(pol$counts~pol$mids, col="red")
```



b) Вычислить выборочные аналоги следующих числовых характеристик: (i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса, (vi) вероятности P(X [c, d]).

Математическое ожидание: 1.01776

```
mean <- sum(x)/length(x)
```

Дисперсия: 0.9938275

```
variance<-sum(x^2)/length(x)-mean^2
```

Медиана: 0.589

```
if(length(vararr) %% 2 != 0){
  median<-vararr[(length(vararr)-1)/2]
} else {
  median<-(vararr[(length(vararr))/2]+vararr[(length(vararr)+2)/2])/2
}
```

Асимметрия: 1.137623

```
assymmetry<-sum((x-mean)^3)/(length(x)*variance^(3/2))
```

Эксцесс: 0.4342106

```
excess<-(sum((x-mean)^4)/(length(x)*variance^2))-3
```

Вероятность P(X [c,d]): 0.82

```
c <- 0
d <- 2
P <- F(d)-F(c)
```

c) В предположении, что исходные наблюдения являются выборкой из показательного распределения, построить оценку максимального правдоподобия параметра λ и соответствующую оценку по методу моментов. Найти смещение оценок.

Нахождение оценки максимального правдоподобия параметра λ:

Плотность показательного распределения: $p(x) = \lambda e^{-\lambda x}\overline{11}_{\{x \geq 0\}}$

$$L(\vec{x}, \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\overline{11}_{\{x \geq 0\}}$$

$$LL(\vec{x}, \lambda) = n * \log(\lambda) - \lambda * \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial\lambda}LL(\vec{x},\lambda)=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i=0$$

$$\hat{\lambda}=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\bar{x}}=0.9825499$$

<u>Метод моментов:</u>

Плотность показательного распределения: $p(x)=\lambda e^{-\lambda x}!\,\overline{11}_{\{x\geq 0\}}$

$$E(x_1)=\frac{1}{\lambda}$$

$$\bar{x}=E(x_1)$$

$$\hat{\lambda}=\frac{n}{\sum_{i=1}^{n}x_i}=\frac{1}{\bar{x}}=0.9825499$$

<u>Нахождение смещения оценок:</u>

$$\sum_{i=1}^{n}x_i\sim\Gamma(n,1/\lambda)$$

$$E(\hat{\lambda})=n\int_0^\infty\frac{1}{x}*\frac{\lambda^n}{\Gamma(n)}*x^{n-1}*e^{-x\lambda}dx=n\int_0^\infty\frac{\lambda^n}{(n-1)!}*x^{n-2}*$$

$$*\,e^{-x\lambda}dx=\frac{n\lambda}{(n-1)!}\int_0^\infty(x\lambda)^{n-2}*e^{-x\lambda}dx=\frac{n\lambda(n-2)!}{(n-1)!}=\frac{n\lambda}{n-1}$$

Несмещённая оценка:

$$\hat{\lambda}=\frac{n-1}{n}*\frac{1}{\bar{x}}=0.9628989$$

d) Построить асимптотический доверительный интервал уровня значимости α_2 для параметра λ на базе оценки максимального правдоподобия.

Найдем информацию Фишера:

$$p(x,\lambda)=\lambda e^{-\lambda x}\overline{11}_{\{x\geq 0\}}$$

$$\log(p(x,\lambda))=\log(\lambda)-\lambda x$$

$$\frac{\partial^2\log(p(x,\lambda))}{\partial\lambda^2}=-\frac{1}{\lambda^2}$$

$$I_1(\lambda)=-E_\lambda\left(-\frac{1}{\lambda^2}\right)=\frac{1}{\lambda^2}$$

$$I(\lambda) = \frac{n}{\lambda^2}$$

Оценка максимального правдоподобия параметра λ:

$$\hat{\lambda} = \frac{1}{\bar{x}}.$$

$$\sqrt{I(\lambda)}\left(\hat{\lambda} - \lambda\right) \Rightarrow N(0{,}1)$$

Эксперимент регулярен, значит, подстановка ОМП вместо параметра в информацию Фишера не нарушает асимптотической нормальности:

$$\sqrt{n * I_1(\lambda)}\left(\hat{\lambda} - \lambda\right) \Rightarrow N(0{,}1)$$

$$\sqrt{\frac{n}{\lambda^2}}\left(\frac{1}{\bar{x}} - \lambda\right) \Rightarrow N(0{,}1)$$

$$p\left(T_1(\vec{x}) \le \lambda \le T_2(\vec{x})\right) = 1 - \alpha_2$$

$$p\left(-x_\alpha \le \sqrt{n\bar{x}^2}\left(\frac{1}{\bar{x}} - \lambda\right) \le x_\alpha\right) = \Phi(x_\alpha) - \Phi(-x_\alpha) = 2 * \Phi(x_\alpha) - 1 = 1 - \alpha_2$$

где $\Phi(x_\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_\alpha} e^{-\frac{t^2}{2}} dt$ - квантиль порядка x_α стандартного нормального закона распределения.

$$x_\alpha = \Phi^{-1}\left(1 - \frac{\alpha_2}{2}\right)$$

$$p\left(\frac{1}{\bar{x}} - x_\alpha\sqrt{\frac{1}{n\bar{x}^2}} \le \lambda \le \frac{1}{\bar{x}} + x_\alpha\sqrt{\frac{1}{n\bar{x}^2}}\right) = 1 - \alpha_2$$

```
TI = function(a2) {
  mean<-sum(x)/length(x)
  xa<-qnorm (1-a2/2)
  return(c(
    1/mean-xa/sqrt(mean*mean*length(x)),
    1/mean+xa/sqrt(mean*mean*length(x))
  ))
}
print(TI(0.01))
```

Получим [0.6246293, 1.3404705]

e) С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с показательным распределением с параметром λ_0. Проверить гипотезу на уровне

значимости α_2. Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

Гипотеза:

$H_0: \lambda = \lambda_0 = 0.50$

Согласно теореме Колмогорова, при справедливости гипотезы:

$\sqrt{n} \sup\limits_x |F_n(x) - F(x)| => K, K -$ распределение Колмогорова

Обозначим $D_n = \sup\limits_x |F_n(x) - F(x)|$

$p(\sqrt{n} D_n < \lambda_{\alpha_2}) = 1 - \alpha_2$

Согласно таблице распределения Колмогорова, $\lambda_{\alpha 2} = 1{,}63$.

Вычислим величину D_n:

```
lambda0 <- 0.5;
Fy<-ecdf(x)
Fexp <- pexp(x,lambda0)
Diff <- array(dim=50)
for(i in 1:length(x)){
  Diff[i]<-abs(Fy(x[i])-Fexp[i])
}
D<-max(Diff)
Dn<-D*sqrt(length(x))
print(Dn)
```

Где достигается sup, табличка

Получим $\sqrt{n} D_n = 1.962$. $\sqrt{n} D_n > \lambda_{\alpha_2}$ – значит, гипотеза H_0 не принимается. Наибольшее значение уровня значимости, при котором мы все еще могли бы принять гипотезу $\approx 0{,}00092$.

f) Используя гистограмму частот, построить критерий значимости $\chi 2$ проверки простой гипотезы согласия с показательным распределением с параметром λ_0. Проверить гипотезу на уровне α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Простая гипотеза H_0: $p(x) = \lambda_0 e^{-\lambda_0 x} \overline{1} 1_{\{x \geq 0\}}, \lambda_0 = 0{,}5$

Граница	**n_i**	**p_i**	**np_i**	**n_i-np_i**	$\frac{(n_i - np_i)^2}{np_i}$
0.4	19	0.18126925	9.063462	9.9365377	10.89371553

0.8	9	0.14841071	7.420535	1.5794646	0.33618984
1.2	5	0.12150841	6.075420	-1.0754205	0.19036201
1.6	5	0.09948267	4.974134	0.0258664	0.00013451
2	3	0.08144952	4.072476	-1.0724761	0.28243384
2.4	2	0.06668523	3.334261	-1.3342615	0.53392743
2.8	4	0.05459725	2.729862	1.2701376	0.59096368
3.2	1	0.04470045	2.235022	-1.2350223	0.68244513
3.6	1	0.03659763	1.829881	-0.8298815	0.37636497
>=3.6	1	0.16529889	8.264944	-7.2649444	6.38593736
sum	50	1	50	0	20.27247

Объединить зоны

Для этого эксперимента $\chi^2_{0.01,9} = 21.66599$, значит мы принимаем гипотезу H_0

Наибольший уровень значимости:

```
r<-10
Xi2<-20.27247
maxval <-pchisq(Xi2,r-2,lower.tail = FALSE);
print(maxval)
```

Полученное значение: 0.016 (по таблице распределения: $\chi^2_{0.016,9} =$ 20.27247).

g) Построить критерий проверки значимости χ2 сложной гипотезы согласия с показательным распределением. Проверить гипотезу на уровне α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

$\cancel{\lambda = \frac{1}{\bar{x}} = 0.9825499}$ Сформулировать гипотезу

Основная гипотеза H0: X ~ Exp(λ)

Граница	n_i	p_i	np_i
0.4	19	0.32498473	16.2492363
0.8	9	0.21936965	10.9684827
1.2	5	0.14807787	7.4038934
1.6	5	0.09995482	4.9977411
2	3	0.06747103	3.3735516
2.4	2	0.04554398	2.2771988

2.8	4	0.03074288	1.5371440
3.2	1	0.02075191	1.0375957
3.6	1	0.01400786	0.7003929
>=3.6	1	0.02909527	1.4547635
sum	50	1	50

sum($\frac{(\boldsymbol{n_i - np_i})^2}{\boldsymbol{np_i}}$) = 5.892289

Для данного эксперимента $\chi^2_{0.01,8} = 20.09024$, что больше полученного значения, следовательно, мы принимаем гипотезу H_0.

Наибольший уровень значимости:

```
r<-10
Xi2<-5.892289
maxval <- pchisq(Xi2,r-2,lower.tail = FALSE);
print(maxval)
```

Полученное значение: 0.6592958 (по таблице распределения: $\chi^2_{0.6592958,8} = 5.892289$).

h) Построить наиболее мощный критерий проверки простой гипотезы о показательности с параметром λ = λ₀ при альтернативе показательности с параметром λ = λ₁. Проверить гипотезу на уровне значимости α₂. Что получится, если поменять местами основную и альтернативную гипотезы?

Основная гипотеза $H_0: \lambda = \lambda_0 = 0{,}5$, альтернативная гипотеза $H_1: \lambda = \lambda_1 = 1$.

Согласно лемме Неймана-Пирсона:

$$\varphi(x) = \begin{cases} 0, l(x; \lambda_1, \lambda_0) < \mathrm{C} \\ 1, l(x; \lambda_1, \lambda_0) > \mathrm{C} \end{cases}$$

Функция правдоподобия для показательного распределения (была найдена ранее):

$$L(\vec{x}, \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \overline{11}_{\{x \geq 0\}}$$

Найдём $l(x; \lambda_1, \lambda_0)$:

$$l(x; \lambda_0, \lambda_1) = \frac{L(x; \lambda_1)}{L(x; \lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^n e^{\sum_{i=1}^n x_i * (\lambda_0 - \lambda_1)} = (2)^n e^{-0.5 * \sum_{i=1}^n x_i}$$

Прологарифмируем полученное выражение:

$$n * \log(2) - 0.5 * \left(\sum_{i=1}^{n} x_i\right) = \log\,(C)$$

$$\sum_{i=1}^{n} x_i = \frac{n * \log(2) - log\,(C)}{0.5}$$

Обозначим $\tilde{C} = \frac{n * \log(2) - log\,(C)}{0.5}$. Тогда критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < \tilde{C} \\ 1, \sum_{i=1}^{n} x_i > \tilde{C} \end{cases}$$

Из этого следует уравнение:

$$P_{\lambda_0}(l(x; \lambda_0, \lambda_1) > C) = P_{\lambda_0}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) = \alpha_2 = 0{,}01$$

$x_i \sim Exp(\lambda_0)$, $\sum_{i=1}^{n} x_i \sim \Gamma(n, 1/\lambda_0)$.

Найдём $\tilde{C}$ из уравнения:

$$E\varphi\,(\overrightarrow{x}) = P(\sum_{i=1}^{n} x_i < \tilde{C}) = \alpha_2$$

$\tilde{C}$ - квантиль распределения $\Gamma(n, 1/\lambda_0)$ уровня α_2:

```
lambda<-0.5
a2<-0.01
C<- qgamma(a2,50,lambda)
```

Критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < 70.06489 \\ 1, \sum_{i=1}^{n} x_i > \ 70.06489 \end{cases}$$

Так как сумма $\sum_{i=1}^{n} x_i = 50.888\ (< 70.06489)$, принимаем основную гипотезу H_0.

Поменяем местами основную и альтернативную теории и проведём аналогичные вычисления:

Найдём $l(x;\lambda_1,\lambda_0)$:

$$l(x;\lambda_0,\lambda_1)=\frac{L(x;\lambda_1)}{L(x;\lambda_0)}=\left(\frac{\lambda_0}{\lambda_1}\right)^n e^{\sum_{i=1}^n x_i*(\lambda_1-\lambda_0)}=(0.5)^n e^{0.5*\sum_{i=1}^n x_i}$$

Прологарифмируем полученное выражение:

$$-\mathrm{n}*\log(2)+0.5*\left(\sum_{i=1}^n x_i\right)=\log{(C)}$$

$$\sum_{i=1}^n x_i=\frac{\mathrm{n}*\log(2)+\log{(C)}}{0.5}$$

Обозначим $\tilde{\mathrm{C}}=\frac{\mathrm{n}*\log(2)+\log{(C)}}{0.5}$. Тогда критерий примет вид:

$$\varphi(x)=\begin{cases}0, \sum_{i=1}^n x_i<\tilde{\mathrm{C}}\\1, \sum_{i=1}^n x_i>\ \tilde{\mathrm{C}}\end{cases}$$

Из этого следует уравнение:

$$P_{\lambda_0}(l(x;\lambda_0,\lambda_1)>C)=P_{\lambda_0}\left(\sum_{i=1}^n x_i>\tilde{C}\right)=\alpha_2\ =0{,}01$$

$x_i \sim Exp(\lambda_1)$, $\sum_{i=1}^n x_i \sim \Gamma(n,1/\lambda_1)$.

Найдём $\tilde{C}$ из уравнения:

$$E\varphi\ \overrightarrow{(x)}=P(\sum_{i=1}^n x_i<\tilde{\mathrm{C}})=\alpha_2$$

$\tilde{C}$ - квантиль распределения $\Gamma(n,1/\lambda_1)$ уровня α_2:

```
lambda<-1
a2<-0.01
C<- qgamma(a2,50,lambda)
```

Критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < 35.03245 \\ 1, \sum_{i=1}^{n} x_i > \ 35.03245 \end{cases}$$

Так как сумма $\sum_{i=1}^{n} x_i = 50.888\ (> 35.03245)$, принимаем альтернативную гипотезу H_1.

i) В пунктах (c)-(h) заменить семейство показательных распределений на семейство гамма-распределений.

i-c) В предположении, что исходные наблюдения являются выборкой из гамма-распределения, построить оценку максимального правдоподобия параметра λ и соответствующую оценку по методу моментов. Найти смещение оценок.

Нахождение оценки максимального правдоподобия параметра λ:

Плотность гамма распределения: $p(x) = \frac{\sqrt{\lambda} e^{-\lambda x/2}}{\sqrt{2\pi x}}$

$$L(\vec{x}, \lambda) = \frac{(\sqrt{\lambda})^n e^{-\lambda \sum_{i=1}^{n} x_i/2}}{(\sqrt{2\pi})^n \prod_{i=1}^{n} \sqrt{x_i}}$$

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

$$\frac{\partial}{\partial\lambda} LL(\vec{x}, \lambda) = \frac{n}{2\lambda} - \frac{1}{2}\sum_{i=1}^{n} x_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}} = 0.9825499$$

Метод моментов:

Плотность гамма распределения: $p(x) = \frac{\sqrt{\lambda} e^{-\lambda x/2}}{\sqrt{2\pi x}}$

$$E(x_1) = k * \theta = \frac{1}{2} * \frac{2}{\lambda} = \frac{1}{\lambda}$$

$$\bar{x} = E(x_1)$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}} = 0.9825499$$

Нахождение смещения оценок:

$$x_i \sim \Gamma\left(\frac{1}{2}, \frac{2}{\lambda}\right)$$

$$\sum_{i=1}^{n} x_i \sim \Gamma\left(\frac{n}{2}, \frac{2}{\lambda}\right).$$

$$E\left(\frac{n}{\sum_{i=1}^{n} x_i}\right) = n\int_0^\infty \frac{1}{x}\frac{\left(\frac{\lambda}{2}\right)^{\frac{n}{2}} * x^{\frac{n}{2}-1} * e^{-\frac{\lambda x}{2}}}{\Gamma\left(\frac{n}{2}\right)}dx$$

$$= \frac{\lambda n}{2}\int_0^\infty \frac{1}{x}\frac{\left(\frac{\lambda}{2}\right)^{\frac{n}{2}-1} * x^{\frac{n}{2}-1} * e^{-\frac{\lambda x}{2}}}{\Gamma\left(\frac{n}{2}\right)}dx =$$

$$= \frac{\lambda n * \Gamma\left(\frac{n}{2} - 1\right)}{2 * \Gamma\left(\frac{n}{2}\right)} = \frac{\lambda n}{(n-2)}$$

Несмещённая оценка:

$$\hat{\lambda} = \frac{(n-2)}{n}\frac{1}{\bar{x}} = 0.9432479$$

i-d) Построить асимптотический доверительный интервал уровня значимости α_2 для параметра λ на базе оценки максимального правдоподобия.

Согласно центральной предельной теореме, при больших n гамма-распределение может быть приближено нормальным распределением:

$$\Gamma(k, \theta) \approx \mathrm{N}(k\theta, k\theta^2), \text{ при } k \to \infty.$$

В нашем случае:

$$\Gamma\left(\frac{1}{2}, \frac{2}{\lambda}\right) \approx \mathrm{N}\left(\frac{1}{\lambda}, \frac{2}{\lambda^2}\right)$$

По лемме Фишера:

$$\frac{\sqrt{n}(\bar{x} - E(x_i))}{\sqrt{D(x_i)}} \Rightarrow N(0, 1)$$

$$E(x) = \frac{1}{\lambda}, D(x) = \frac{2}{\lambda^2}$$

Следовательно:

$$\frac{\sqrt{n}\left(\bar{x}-\frac{1}{\lambda}\right)}{\bar{x}\sqrt{2}} \Rightarrow N(0,1)$$

Найдем область:

$$P\left(-x_\alpha \leq \sqrt{n}\frac{\bar{x}-\frac{1}{\lambda}}{\bar{x}\sqrt{2}} < x_\alpha\right) = \Phi(\mathrm{x}_\alpha) - \Phi(-\mathrm{x}_\alpha) = 2\Phi(\mathrm{x}_\alpha) - 1 = 1 - \alpha_2$$

где $\Phi(x_\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_\alpha} e^{-\frac{t^2}{2}}dt$ - квантиль порядка x_α стандартного нормального закона распределения.

$$x_\alpha = \Phi^{-1}\left(1-\frac{\alpha_2}{2}\right)$$

$$-x_\alpha \leq \sqrt{n}\frac{\bar{x}-{}^1/_\lambda}{\bar{x}\sqrt{2}} \leq x_\alpha$$

$$\bar{x}-\frac{x_\alpha\bar{x}\sqrt{2}}{\sqrt{n}} \leq \frac{1}{\lambda} \leq \bar{x}+\frac{x_\alpha\bar{x}\sqrt{2}}{\sqrt{n}}$$

$$\left(\bar{x}+\frac{x_\alpha\bar{x}\sqrt{2}}{\sqrt{n}}\right)^{-1} \leq \lambda \leq \left(\bar{x}-\frac{x_\alpha\bar{x}\sqrt{2}}{\sqrt{n}}\right)^{-1}$$

Вычислим:

```
TIi = function(a2) {
  mean<-sum(x)/length(x)
  xa<-qnorm (1-a2/2)
  return(c(
    1/(mean+xa*mean*sqrt(2/length(x))),
    1/(mean-xa*mean*sqrt(2/length(x)))
  ))
}
print(TIi(0.01))
```

Результат: [0.6484768, 2.0265692]

i-e) С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с гамма распределением с параметром λ_0. Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

Гипотеза:

$H_0: \lambda = \lambda_0 = 0.5$

Согласно теореме Колмогорова, при справедливости гипотезы:

$\sqrt{n} \sup_x |F_n(x) - F(x)| => K, K -$ распределение Колмогорова

Обозначим $D_n = \sup_x |F_n(x) - F(x)|$

$p(\sqrt{n} D_n < \lambda_{\alpha_2}) = 1 - \alpha_2$

Согласно таблице распределения Колмогорова, $\lambda_{\alpha 2} = 1,63$.

Вычислим величину D_n:

```
lambda0 = 0.5;
Fy = ecdf(x)
lambda = 0.5
Fgamma = pgamma(x, 0.5, lambda / 2)
Diff = array(dim=50)
for(i in 1:length(x)){
  Diff[i] = abs(Fy(x[i])-Fgamma[i])
}
D = max(Diff)
Dn<-D*sqrt(length(x))
```

Полученное значение $\sqrt{n} D_n = 1.311351$. $\sqrt{n} D_n < \lambda_{\alpha_2}$ – значит, гипотеза H_0 принимается. Наибольший уровень значимости по таблице распределений Колмогорова будет примерно равен 0,064

i-f) Используя гистограмму частот, построить критерий значимости χ2 проверки простой гипотезы согласия с гамма распределением с параметром λ_0. Проверить гипотезу на уровне α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Простая гипотеза H_0: $p(x) = \frac{\sqrt{\lambda_0} e^{-\lambda_0 x/2}}{\sqrt{2\pi x}}, \lambda_0 = 0,5$

Граница	n_i	p_i
0.4	19	0.34527915
0.8	9	0.12763159
1.2	5	0.08851123
1.6	5	0.06748466
2	3	0.05378286
2.4	2	0.04398883
2.8	4	0.03659811

3.2	1	0.03082036
3.6	1	0.02619072
>=3.6	1	0.17971249
sum	50	1

$$\sum \frac{(n_i - np_i)^2}{np_i} = 12.09439$$

Для данного эксперимента $\chi^2_{0.01,9} = 21.66599$, что больше полученного значения, следовательно, мы принимаем гипотезу H_0.

Наибольший уровень значимости: 0.208

```
r<-10
Xi2<-39.90235
maxval <-pchisq(Xi2,r-1,lower.tail = FALSE)
```

i-g) Построить критерий проверки значимости $\chi 2$ сложной гипотезы согласия с гамма распределением. Проверить гипотезу на уровне α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

$$\lambda = \frac{1}{\bar{x}} = 0.9825499$$

Основная гипотеза H_0: $X \sim \Gamma(1/2, 2/\lambda)$.

Граница	$\mathbf{n_i}$	$\mathbf{p_i}$
0.4	19	0.46928376
0.8	9	0.15541598
1.2	5	0.09775279
1.6	5	0.06764224
2	3	0.04893627
2.4	2	0.03633663
2.8	4	0.02744720
3.2	1	0.02098589
3.6	1	0.01619183
>=3.6	1	0.06000742
sum	50	1

$$\sum \frac{(n_i - np_i)^2}{np_i} = 8.375704$$

Для данного эксперимента $\chi^2_{0.01,8} = 20.09024$, что больше полученного значения, следовательно, мы принимаем гипотезу H_0.

Наибольший уровень значимости: 0.3976567

```
pchisq(8.375704,8,lower.tail = FALSE)
```

i-h) Построить наиболее мощный критерий проверки простой гипотезы о гаммовости с параметром $\lambda = \lambda_0$ при альтернативе гаммовости с параметром $\lambda = \lambda_1$. Проверить гипотезу на уровне значимости α_2. Что получится, если поменять местами основную и альтернативную гипотезы?

Основная гипотеза $H_0 : \lambda = \lambda_0 = 0.5$, альтернативная гипотеза $H_1 : \lambda = \lambda_1 = 1$.

Согласно лемме Неймана-Пирсона:

$$\varphi(x) = \begin{cases} 0, l(x; \lambda_1, \lambda_0) < \text{C} \\ 1, l(x; \lambda_1, \lambda_0) > \text{C} \end{cases}$$

Функция правдоподобия для гамма-распределения

$$L(\vec{x}, \lambda) = \frac{(\sqrt{\lambda})^n e^{-\frac{\lambda}{2}\sum_{i=1}^n x_i}}{(\sqrt{2\pi})^n \prod_{i=1}^n \sqrt{x_i}}$$

Найдём $l(x; \lambda_1, \lambda_0)$:

$$l(x; \lambda_0, \lambda_1) = \frac{L(x; \lambda_1)}{L(x; \lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{n/2} e^{\frac{(\lambda_0 - \lambda_1)}{2} * \sum_{i=1}^n x_i} = (2)^{n/2} e^{-0.25 \sum_{i=1}^n x_i}$$

Прологарифмируем полученное выражение:

$$\left(\frac{\text{n}}{2}\right) * \log(2) - 0.25 * \left(\sum_{i=1}^n x_i\right) = \log\,(C)$$

$$\sum_{i=1}^n x_i = \frac{\left(\frac{\text{n}}{2}\right) * \log(2) - \log(C)}{0.25}$$

Обозначим $\tilde{\text{C}} = \frac{\left(\frac{\text{n}}{2}\right) * \log(2) - \log(C)}{0.25}$. Тогда критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < \tilde{C} \\ 1, \sum_{i=1}^{n} x_i > \tilde{C} \end{cases}$$

Из этого следует уравнение:

$$P_{\lambda_0}(l(x;\lambda_0,\lambda_1) > C) = P_{\lambda_0}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) = \alpha_2 = 0{,}01$$

Из того, что $x_i \sim \Gamma\left(\frac{1}{2}, \frac{2}{\lambda_0}\right) \sim N\left(\frac{1}{\lambda_0}, \frac{2}{\lambda_0^2}\right)$, можно сделать вывод, что $\sum_{i=1}^{n} x_i \sim N\left(\frac{n}{\lambda_0}, \frac{2n}{\lambda_0^2}\right)$.

Найдём $\tilde{C}$ из уравнения:

$$E\varphi\,\overrightarrow{(x)} = P(\sum_{i=1}^{n} x_i < \tilde{C}) = \alpha_2$$

$\tilde{C}$ - квантиль распределения $N\left(\frac{1}{\lambda_0}, \frac{2}{\lambda_0^2}\right)$ уровня α_2:

```
lambda<-0.5
a2<-0.01
C<-qnorm(a2,50/lambda,sqrt(2*50)/lambda)
```

Критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < 53.47304 \\ 1, \sum_{i=1}^{n} x_i > 53.47304 \end{cases}$$

Так как сумма $\sum_{i=1}^{n} x_i = 50.888\ (< 53.47304)$, принимаем основную гипотезу H_0.

Поменяем местами основную и альтернативную теории и проведём аналогичные вычисления:

Найдём $l(x;\lambda_1,\lambda_0)$:

$$l(x;\lambda_0,\lambda_1) = \frac{L(x;\lambda_0)}{L(x;\lambda_1)} = \left(\frac{\lambda_0}{\lambda_1}\right)^{n/2} e^{\frac{(\lambda_1-\lambda_0)}{2} * \sum_{i=1}^{n} x_i} = (0.5)^{n/2} e^{0.25 \sum_{i=1}^{n} x_i}$$

Прологарифмируем полученное выражение:

$$-\left(\frac{\text{n}}{2}\right) * \log(2) + 0{,}25 * \left(\sum_{i=1}^{n} x_i\right) = \log\,(C)$$

$$\sum_{i=1}^{n} x_i = \frac{\left(\frac{\text{n}}{2}\right) * \log(2) + \log(C)}{0{,}25}$$

Обозначим $\tilde{\text{C}} = \frac{\left(\frac{\text{n}}{2}\right) * \log(2) + \log(C)}{0{,}25}$. Тогда критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < \tilde{\text{C}} \\ 1, \sum_{i=1}^{n} x_i > \tilde{\text{C}} \end{cases}$$

Из этого следует уравнение:

$$P_{\lambda_1}(l(x; \lambda_0, \lambda_1) > C) = P_{\lambda_1}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) = \alpha_2 = 0{,}01$$

Из того, что $x_i \sim \Gamma\left(\frac{1}{2}, \frac{2}{\lambda_1}\right) \sim N\left(\frac{1}{\lambda_1}, \frac{2}{\lambda_1^2}\right)$, можно сделать вывод, что $\sum_{i=1}^{n} x_i \sim N\left(\frac{n}{\lambda_1}, \frac{2n}{\lambda_1^2}\right)$.

Найдём $\tilde{C}$ из уравнения:

$$E\varphi\,\overrightarrow{(x)} = P(\sum_{i=1}^{n} x_i < \tilde{\text{C}}) = \alpha_2$$

$\tilde{C}$ - квантиль распределения $N\left(\frac{1}{\lambda_1}, \frac{2}{\lambda_1^2}\right)$ уровня α_2:

```
lambda<-1
a2<-0.01
C<-qnorm(a2,50/lambda,sqrt(2*50)/lambda)
```

Критерий примет вид:

$$\varphi(x) = \begin{cases} 0, \sum_{i=1}^{n} x_i < 26.73652 \\ 1, \sum_{i=1}^{n} x_i > 26.73652 \end{cases}$$

Так как сумма $\sum_{i=1}^{n} x_i = 50.888\ (> 26.73652)$, принимаем альтернативную гипотезу H_1.