$\Theta_{UR} := \Theta$ — множество допустимых значений вектора неизвестны параметров без учёта ограничений;

$\Theta_R := \{\theta \in \Theta : g(\theta) = 0\}$ — множество допустимых значений вектор неизвестных параметров с учётом ограничений;

$\widehat{\theta}_{UR} \in \Theta_{UR}$ — точка максимума функции l на множестве Θ_{UR};

$\widehat{\theta}_R \in \Theta_R$ — точка максимума функции l на множестве Θ_R.

Тогда для тестирования гипотезы H_0 можно воспользоваться одной из следующих ниже статистик:

– статистика отношения правдоподобия

$$LR := -2(l(\widehat{\theta}_R) - l(\widehat{\theta}_{UR})) \overset{a}{\sim} \chi^2_r;$$

– статистика Вальда

$$W := g'(\widehat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial \theta'}(\widehat{\theta}_{UR}) \cdot I^{-1}(\widehat{\theta}_{UR}) \cdot \frac{\partial g'}{\partial \theta}(\widehat{\theta}_{UR})\right]^{-1} \cdot g(\widehat{\theta}_{UR}) \overset{a}{\sim} \chi^2_r;$$

– статистика множителей Лагранжа

$$LM := \left[\frac{\partial l}{\partial \theta}(\widehat{\theta}_R)\right]' \cdot I^{-1}(\widehat{\theta}_R) \cdot \left[\frac{\partial l}{\partial \theta}(\widehat{\theta}_R)\right] \overset{a}{\sim} \chi^2_r.$$

Задача 5.1. Дядя Вова (Владимир Николаевич) и Скрипач (Гедеван) зарабатывают на Плюке чатлы, чтобы купить гравицапу. Число заработанных за i-й день чатлов имеет пуассоновское распределение, заработки за разные дни независимы. За прошедшие 100 дней они заработали 250 чатлов.

1. Оцените параметр λ пуассоновского распределения методом максимального правдоподобия.
2. Сколько дней им нужно давать концерты, чтобы оценка вероятности купить гравицапу составила 0.99? Гравицапа стоит пол кц или 2200 чатлов.
3. Постройте 95% доверительный интервал для λ.
4. Проверьте гипотезу о том, что средний дневной заработок равен 2 чатла, с помощью теста отношения правдоподобия, теста Вальда, теста множителей Лагранжа.

Решение. Пусть X_i — число заработанных за i-й день чатлов. По условию $X_i \sim Pois(\lambda)$.

Реализация случайной выборки: $\sum\limits_{i=1}^{100} x_i = 250.$

Функция максимального правдоподобия:

$$L(\lambda, x_1, \ldots, x_{100}) = \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdot \ldots \cdot \mathbb{P}(X_{100} = x_{100}) =$$

$$= \frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{x_{100}}}{x_{100}!} \cdot e^{-\lambda} = \frac{\lambda^{\sum x_i}}{x_1! \cdot \ldots \cdot x_{100}!} e^{-100\lambda}.$$

Логарифмическая функция максимального правдоподобия:

$$l(\lambda, x_1, \ldots, x_{100}) = \sum x_i \cdot \log \lambda - 100\lambda - \log\,(x_1! \cdot \ldots \cdot x_{100}!)\,.$$

1. $\widehat{\lambda} = 2.5$

$$\frac{\partial l(\cdot)}{\partial \lambda} = \frac{\sum x_i}{\lambda} - 100.$$

Из условия первого первого порядка:

$$\frac{250}{\widehat{\lambda}} - 100 = 0,$$

$$\widehat{\lambda} = 2.5.$$

Заметим, что $\widehat{\lambda} > 0$, следовательно, то, что мы выполняли безусловную оптимизацию вместо условной (хотя должны были, учитывая, что $\lambda > 0$), не повлияло на решение.

2. $k = 838$, где k — минимальное необходимое число дней.

$$\widehat{\mathbb{P}}\left(\sum_{i=1}^{k} X_i > 2200\right) = 0.99,$$

$$\widehat{\mathbb{P}}\left(\frac{\sum_{i=1}^{k} X_i - \widehat{\lambda} \cdot k}{\sqrt{\widehat{\lambda} \cdot k}} > \frac{2200 - \widehat{\lambda} \cdot k}{\sqrt{\widehat{\lambda} \cdot k}}\right) = 0.99.$$

Из ЦПТ:

$$\frac{\sum_{i=1}^{k} X_i - \lambda \cdot k}{\sqrt{\lambda \cdot k}} \sim \mathcal{N}(0, 1).$$

Значит:

$$\frac{2200 - \widehat{\lambda} \cdot k}{\sqrt{\widehat{\lambda} \cdot k}} = x_{0.99},$$

где $x_{0.99} \approx 2.3263$ — 99 % квантиль нормального распределения. Решая квадратное уравнение, получим:

$$n \approx 837.4228.$$

Значит, для того, чтобы с вероятностью не меньше 99 % купить гравицапу, необходимо дать, как минимум, 838 концертов.

3. $\lambda \in [2.190102, 2.809898]$.

Так как оценки метода максимального правдоподобия асимптотически нормальны, то:

$$\widehat{\lambda} \overset{\text{ac}}{\sim} \mathcal{N}\left(\lambda, \operatorname{Var}\widehat{\lambda}\right).$$

Для оценки дисперсии можем использовать оценку информационной матрица Фишера:

$$\widehat{\operatorname{Var}\lambda} = \left(\widehat{I}\left(\widehat{\lambda}\right)\right)^{-1},$$

$$\widehat{\operatorname{Var}\lambda} = -\left(-\frac{\sum x_i}{\widehat{\lambda}^2}\right)^{-1} = \frac{\widehat{\lambda}^2}{\sum x_i} = 0.025.$$

Границы 95 % доверительного интервала:

```
left <- 2.5 + qnorm(0.025)*sqrt(0.025)
right <- 2.5 + qnorm(0.975)*sqrt(0.025)

cbind(left, right)

##          left    right
## [1,] 2.190102 2.809898
```

Итак:

$$\lambda \in [2.190102, 2.809898].$$

4. $$\begin{cases} H_0 : \lambda = 2, \\ H_a : \lambda \neq 2. \end{cases}$$

Значение логарифмической функции правдоподобия в неограниченной модели:

$$l(\widehat{\lambda}_{UR}) = 250 \cdot \log 2.5 - 100 \cdot 2.5 - \log\left(x_1! \cdot \ldots \cdot x_{100}!\right) \approx$$
$$\approx -20.92732 - \log\left(x_1! \cdot \ldots \cdot x_{100}!\right).$$

Значение логарифмической функции правдоподобия в ограниченной модели:

$$l(\widehat{\lambda}_{UR}) = 250 \cdot \log 2 - 100 \cdot 2 - \log\left(x_1! \cdot \ldots \cdot x_{100}!\right) \approx$$
$$\approx -26.7132 - \log\left(x_1! \cdot \ldots \cdot x_{100}!\right).$$

Применяем **тест отношения правдоподобия**:

$$LR = 2\left(l(\widehat{\lambda}_{UR}) - l(\widehat{\lambda}_R)\right) \overset{H_0}{\sim} \chi^2_q.$$

Наблюдаемое значение LR-статистики:

$$LR = 2\Big(-20.92732 - \log\left(x_1! \cdot \ldots \cdot x_{100}!\right) - \\ - \left(-26.7132 - \log\left(x_1! \cdot \ldots \cdot x_{100}!\right)\right)\Big) = 5.78588.$$

При 5 % уровне значимости гипотеза отвергается в пользу альтернативной:

```
p.value <- 1 - pchisq(5.78588, df = 1)
p.value

## [1] 0.01615541

p.value > 0.05 # если TRUE, то гипотеза не отвергается

## [1] FALSE
```

Применяем **тест Вальда**:

$$W = \left(g\left(\widehat{\lambda}_{UR}\right)\right)' \cdot \left(\frac{\partial g}{\partial \lambda'}\left(\widehat{\lambda}_{UR}\right) \cdot \widehat{I}^{-1}\left(\widehat{\lambda}_{UR}\right) \cdot \frac{\partial g'}{\partial \lambda}\left(\widehat{\lambda}_{UR}\right)\right)^{-1} \cdot g\left(\widehat{\lambda}_{UR}\right) \overset{H_0}{\sim} \chi^2_q.$$

В нашей задаче:

$$g\left(\widehat{\lambda}_{UR}\right) = \widehat{\lambda}_{UR} - 2 = 0.5,$$

$$\frac{\partial g}{\partial \lambda}\left(\widehat{\lambda}_{UR}\right) = 1,$$

$$\left(\widehat{I}\left(\widehat{\lambda}_{UR}\right)\right)^{-1} = -0.025,$$

$$W = 0.5 \cdot (0.025)^{-1} \cdot 0.5 = 10.$$

При 5 % уровне значимости гипотеза отвергается в пользу альтернативной:

Применяем **тест множителей Лагранжа**:

$$LM = \left(\frac{\partial l}{\partial \theta}\left(\widehat{\theta}_R\right)\right)' \cdot \widehat{I}^{-1}\left(\widehat{\theta}_R\right) \cdot \left(\frac{\partial l}{\partial \theta}\left(\widehat{\theta}_R\right)\right) \overset{H_0}{\sim} \chi^2_q.$$

```
p.value <- 1 - pchisq(10, df = 1)
p.value

## [1] 0.001565402

p.value > 0.05 # если TRUE, то гипотеза не отвергается

## [1] FALSE
```

Здесь:

$$\frac{\partial l}{\partial \lambda}\left(\widehat{\lambda}_R\right) = \frac{250}{2} - 100 = 25,$$

$$\left(\widehat{I}\left(\widehat{\lambda}_R\right)\right)^{-1} = \frac{4}{250} = 0.016,$$

$$LM = 25 \cdot 0.016 \cdot 25 = 10.$$

При 5 % уровне значимости гипотеза отвергается в пользу альтернативной:

```
p.value <- 1 - pchisq(10, df = 1)
p.value

## [1] 0.001565402

p.value > 0.05 # если TRUE, то гипотеза не отвергается

## [1] FALSE
```

Задача 5.2. Инопланетянин Капп совершил вынужденную посадку на Землю. Каждый день он выходит на связь со своей далёкой планетой. Продолжительность каждого сеанса связи имеет экспоненциальное распределение с параметром λ. Прошедшие 100 сеансов связи в сумме длились 11 часов.

1. Оцените параметр λ экспоненциального распределения методом максимального правдоподобия.
2. Постройте 95% доверительный интервал для λ.
3. Проверьте гипотезу о том, что средняя продолжительность сеанса связи равна 5 минутам с помощью теста отношения правдоподобия, теста Вальда, теста множителей Лагранжа.