# 1 Conditional densities and expectations

From the probability theory (NMSA 333) we know, that the conditional expectation of $Y$ for given $X$ is defined as
$$\mathsf{E}\,(Y \mid X) = \mathsf{E}\,(Y \mid \sigma(X)),$$
where $\sigma(X)$ is sigma-algebra generated with the random variable $X$. In what follows we concentrate on the situation when the random vector $(X, Y)^\mathsf{T}$ has a joint density $f_{XY}(x, y)$ with respect to the two-dimensional Lebesgue measure.

**Conditional density** of the random random $Y$ for given $X$ is defined for $f_X(x) > 0$ as
$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)},$$
where $f_X(x)$ is the marginal density of $X$.

**Conditional expectation:**
$$\mathsf{E}\,(Y \mid X = x) = \int y\, f_{Y|X}(y|x)\, dy.$$
It is known that $\mathsf{E}\,Y$ is „the best" estimator of $Y$ (when the quadratic loss function is minimized), when one know only the marginal distribution of $Y$. Analogously $\mathsf{E}\,(Y \mid X = x)$ is „the best" estimator $Y$ with the knowledge of the joint distribution of $(Y, X)^\mathsf{T}$ and the realisation of $X$.

Be careful. While $\mathsf{E}\,(Y \mid X = x)$ is a function that is defined on the support of $X$, the conditional expectation $\mathsf{E}\,(Y \mid X)$ is a random variable that is a function of $X$.

**Some useful properties of conditional expectation:** Let $h_1 : \mathbb{R}^2 \to \mathbb{R}$, $h_2 : \mathbb{R}^2 \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ are measurable functions. Then

(i) $\mathsf{E}\,(a \mid X) = a$ for an arbitrary $a \in \mathbb{R}$.

(ii) $\mathsf{E}\big(\mathsf{E}\,(Y \mid X)\big) = \mathsf{E}\,Y$.

(iii) $\mathsf{E}\,(a_1\, h_1(X, Y) + a_2 h_2(X, Y) \mid X) = a_1\, \mathsf{E}\,(h_1(X, Y) \mid X) + a_2\, \mathsf{E}\,(h_2(X, Y) \mid X)$ for an arbitrary $a_1, a_2 \in \mathbb{R}$.

(iv) $\mathsf{E}\,(\psi(X) h_1(X, Y) \mid X) = \psi(X)\, \mathsf{E}\,(h_1(X, Y) \mid X)$.

**Variance decomposition with the help of conditioning:**

❖ $\mathsf{var}(Y) = \mathsf{E}\big[\mathsf{var}\,(Y \mid X)\big] + \mathsf{var}\big(\mathsf{E}\,(Y \mid X)\big)$.

Proof:
$$\begin{aligned}
\mathsf{var}(Y) &= \mathsf{E}\,Y^2 - \big[\mathsf{E}Y\big]^2 = \mathsf{E}\left[\mathsf{E}\left(Y^2 \mid X\right)\right] - \big[\mathsf{E}Y\big]^2 \\
&= \mathsf{E}\left[\mathsf{var}\,(Y \mid X)\right] + \mathsf{E}\left[\mathsf{E}\,(Y \mid X)\right]^2 - \left[\mathsf{E}\{\mathsf{E}\,(Y \mid X)\}\right]^2 \\
&= \mathsf{E}\left[\mathsf{var}\,(Y \mid X)\right] + \mathsf{var}\big(\mathsf{E}\,(Y \mid X)\big).
\end{aligned}$$

### Example 1. $f(x,y) = x + y$

Let $(X,Y)^\mathsf{T}$ be a random vector with the density

$$f(x,y) = (x+y)\mathbb{I}_M, \quad M = \{(x,y) : 0 \le x \le 1, 0 \le y \le 1\}.$$

(i) Calculate $\mathsf{E}\,(\,XY \mid X = x\,)$.
(ii) Calculate $\mathsf{E}\,(\,XY \mid X\,)$.
(iii) Calculate $\mathsf{E}\,\big(\,XY^2 \,\big|\, X\,\big)$.
(iv) Calculate $\mathsf{E}\,\big(\,XY^2 \,\big|\, X^2\,\big)$.

### Example 2. Conditionally normal distribution

Consider the random vector $(Y,X)^\mathsf{T}$. Let $Y$ given $X$ have the normal distribution with the expectation $2\,X^3$ and the variance $3\,X^2$. Further let $X$ have the uniform distribution on the interval $(0,1)$.

(i) Calculate $\mathsf{E}\big[\frac{Y}{X^2}|X\big]$.
(ii) Calculate $\mathsf{E}\,\frac{Y}{X^2}$.
(iii) Calculate $\mathsf{E}\,Y$.
(iv) Calculate $\mathsf{var}(Y)$.

### Example 3. Conditional expectation of the distribution on a rectangle

Let the random vector $(X,\,Y)^\mathsf{T}$ follow the distribution given by the density

$$f(x,\,y) = \begin{cases} \dfrac{1}{x}\exp\Big(-\dfrac{y}{x}\Big), & 1 < x < 2,\ y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Calculate $\mathsf{E}\big(Y \,\big|\, X = t\big)$ a $\mathsf{E}\big(Y \,\big|\, X\big)$.
(ii) Calculate $\mathsf{E}\left(Y \,\middle|\, \log\Big(\dfrac{X-1}{2-X}\Big) = t\right)$ a $\mathsf{E}\left(Y \,\middle|\, \log\Big(\dfrac{X-1}{2-X}\Big)\right)$.
(iii) Calculate $\mathsf{E}\left(\dfrac{Y}{X^6} \,\middle|\, \log\Big(\dfrac{X-1}{2-X}\Big)\right)$

### Example 4. Conditionally uniform distribution

Consider a random vector $(Y,X)^\mathsf{T}$. Let $Y$ given $X$ have uniform distribution $\mathsf{R}(0,X^2+1)$. Further let $X$ have normal distribution $\mathsf{N}(0,1)$.

(i) Calculate $\mathsf{E}\big[Y|\exp\{X\}\big]$.
(ii) Calculate $\mathsf{E}\,Y$.
(iii) Calculate $\mathsf{var}(Y)$.

## 2 Sufficient statistics

Let the random vector $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ have density $f(\boldsymbol{x};\boldsymbol{\theta})$ with respect to a $\sigma$-finite measure $\mu$, where $\boldsymbol{\theta} \in \Theta$ is an unknown parameter.

**Definition 1.** We say that the statistic $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{X})$ is **sufficient** for the parameter $\boldsymbol{\theta}$, if the conditional distribution $\boldsymbol{X}$ given $\boldsymbol{S}$ does not depend on $\boldsymbol{\theta}$.

Thus the sufficient statistic contains all the available information about $\boldsymbol{\theta}$ that is in the random vector $\boldsymbol{X}$. The following theorem is useful when searching for sufficient statistics.

**Theorem 1** (Fisher-Neyman factorization theorem)**.** *The statistic $\boldsymbol{S}$ is sufficient if and only if there exist a non-negative measurable functions $g(\boldsymbol{s};\boldsymbol{\theta})$ and $h(\boldsymbol{x})$, such that*

$$f(\boldsymbol{x};\boldsymbol{\theta}) = g\big(\boldsymbol{S}(\boldsymbol{x});\boldsymbol{\theta}\big)\, h(\boldsymbol{x}).$$

In applications we search for sufficient statistics that are in some sense 'minimal'. This is motivation for the following definition.

**Definition 2.** We say that the sufficient statistic $\boldsymbol{S}(\boldsymbol{X})$ is **minimal**, if for each sufficient statistic $\boldsymbol{T}(\boldsymbol{X})$ there exists a function $\boldsymbol{g}$ such that $\boldsymbol{S}(\boldsymbol{X}) = \boldsymbol{g}\big(\boldsymbol{T}(\boldsymbol{X})\big)$.

The following theorem can be useful to find the minimal sufficient statistic.

**Theorem 2** (Lehmann-Scheffé theorem about a minimal sufficient statistic)**.** *Let $\boldsymbol{S}$ be a sufficient statistic and the set $M = \{\boldsymbol{x} : f(\boldsymbol{x};\boldsymbol{\theta}) > 0\}$ does not depend on $\boldsymbol{\theta}$. For $\boldsymbol{x}, \boldsymbol{y} \in M$ introduce*

$$h(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) = \frac{f(\boldsymbol{x};\boldsymbol{\theta})}{f(\boldsymbol{y};\boldsymbol{\theta})}.$$

*Let $h(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$ implies that $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{S}(\boldsymbol{y})$. Then $\boldsymbol{S}(\boldsymbol{X})$ is minimal.*

**Definition 3.** We say that the statistic $\boldsymbol{S}$ is **complete**, if for each measurable function $w(\boldsymbol{S})$ the following implication holds

$$\Big\{\mathsf{E}_{\boldsymbol{\theta}}\, w(\boldsymbol{S}) = 0 \text{ for each } \boldsymbol{\theta} \in \Theta\Big\} \Longrightarrow \Big\{w(\boldsymbol{S}) = 0 \text{ almost surely for each } \boldsymbol{\theta} \in \Theta\Big\}.$$

### Example 5. Geometric distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the geometric distribution, i.e.

$$\mathsf{P}(X_i = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \dots$$

Find, if $S(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for parameter $p$.

(i) With the help of the definition of the sufficient statistic.

(ii) With the help of the Fisher-Neyman factorization theorem.

### Example 6. Poisson distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the Poisson distribution, i.e.

$$\mathsf{P}(X_i = k) = \frac{\lambda^k \, \mathrm{e}^{-\lambda}}{k!}, \qquad k = 0, 1, 2, \dots$$

Find, if $S(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is sufficient statistic for the parameter $\lambda$.

(i) With the help of the definition the sufficient statistic.

(ii) With the help of the Fisher-Neyman factorization theorem.

(iii) Show that $X_1 + X_2$ is a complete statistic.

### Example 7. Uniform discrete distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the uniform discrete distribution, i.e.

$$\mathsf{P}(X_i = k) = \frac{1}{M}, \qquad k = 1, 2, \dots, M,$$

where $M \in \mathbb{N}$. Find, if $S(\boldsymbol{X}) = \max_{1 \le i \le n} X_i$ is a sufficient statistic for the parameter $M$.

(i) With the help of definition sufficient statistic.

(ii) With the help of Fisher-Neyman factorization theorem.

### Example 8. Zero mean Gaussian distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the normal distribution $\mathsf{N}(0, \sigma^2)$. Check, if the following statistics are sufficient for the parameter $\sigma^2$.

$$\text{(i) } \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{X}, \qquad \text{(ii) } \boldsymbol{T}(\boldsymbol{X}) = \big(|X_1|, \dots, |X_n|\big)^\mathsf{T}, \qquad \text{(iii) } T(\boldsymbol{X}) = \sum_{i=1}^n X_i, \qquad \text{(iv) } T(\boldsymbol{X}) = \sum_{i=1}^n |X_i|,$$

$$\text{(v) } T(\boldsymbol{X}) = \sum_{i=1}^n X_i^2, \qquad \text{(vi) } T(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n X_i^2, \qquad \text{(vii) } \boldsymbol{T}(\boldsymbol{X}) = \left(\frac{1}{n}\sum_{i=1}^{n-1} X_i^2, X_n^2\right)^\mathsf{T}.$$

### Example 9. Bernoulli distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from Bernoulli distribution, i.e.

$$\mathsf{P}(X_i = 1) = p, \qquad \mathsf{P}(X_i = 0) = 1 - p.$$

Define $S(\boldsymbol{X}) = \sum_{i=1}^n X_i$.

(i) Show that $S(\boldsymbol{X})$ is sufficient for parameter $p$.

(ii) Show that $S(\boldsymbol{X})$ is even minimal sufficient statistic for parameter $p$.

(iii) From the definition prove that $T(\boldsymbol{X}) = X_1$ is complete statistic for parameter $p$. Is the statistic $T(\boldsymbol{X})$ sufficient?

(iv) From the definition show that $S(\boldsymbol{X})$ is a complete statistic for the parameter $p$.

## Example 10. Gaussian distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be random sample from the normal distribution $\mathsf{N}(\mu, \sigma^2)$.

(i) Find minimal sufficient statistic for $(\mu, \sigma^2)^\mathsf{T}$.

## Example 11. Uniform distribution $\mathsf{R}(0, \theta)$

Let $X_1, \ldots, X_n$ be a random sample from the uniform distribution $\mathsf{R}(0, \theta)$ with the density

$$f(x) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta > 0$.

(i) Show that the statistic $X_{(n)} = \max_{1 \le i \le n} X_i$ is sufficient and complete.

(ii) Show that the statistic $X_1$ is complete, but it is not sufficient.

## Example 12. Uniform distribution $\mathsf{R}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$

Let $X_1, \ldots, X_n$ be a random sample from the uniform distribution $\mathsf{R}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ with the density

$$f(x) = \begin{cases} 1, & \theta - \frac{1}{2} < x < \theta + \frac{1}{2}, \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta \in \mathbb{R}$.

(i) Show that $S(\mathbf{X}) = \big(X_{(1)}, X_{(n)}\big)^\mathsf{T}$ is a sufficient statistic for the parameter $\theta$.

(ii) Show that $S(\mathbf{X})$ is not complete.

## Example 13. Pareto distribution

Let $X_1, \ldots, X_n$ be a random sample from Pareto distribution with the density

$$f(x) = \frac{\beta\, \alpha^\beta}{x^{\beta+1}}\, \mathbb{I}_{\{x > \alpha\}}, \qquad \text{where } \beta > 0, \alpha > 0.$$

(i) Find a non-trivial sufficient statistic for the parameter $\boldsymbol{\theta} = (\alpha, \beta)^\mathsf{T}$.

## Example 14. „Curved normal“ $N(\mu, \mu^2)$

Let $X_1, \ldots, X_n$ b a random sample form the Gaussian distribution $\mathsf{N}(\mu, \mu^2)$, where $\mu \in \mathbb{R}$.

(i) Find a minimal sufficient statistic.

(ii) Is the statistic from (i) complete?

## Example 15. Multinomial distribution

We are modelling the number of children born in days of the week with the help of multinomial distribution $\mathsf{M}(n, p_1, \ldots, p_7)$, i.e.

$$\mathsf{P}\big(X_1 = x_1, \ldots, X_7 = x_7\big) = \frac{n!}{x_1! \cdots x_7!} p_1^{x_1} \cdots p_7^{x_7}, \quad \text{where } \sum_{i=1}^{7} x_i = n, \quad \sum_{i=1}^{7} p_i = 1.$$

(i) Is the vector $\boldsymbol{X} = (X_1, \ldots, X_7)$ the minimal sufficient statistic for the vector parameter $\mathbf{p} = (p_1, \ldots, p_7)^\mathsf{T}$? If yes, would it be possible to decrease the dimension of the statistic so that it is still minimal sufficient?

(ii) Find the minimal sufficient statistic (for the parameters of the model) provided that $p_1 = p_2 = \ldots = p_5$ and $p_6 = p_7$.

(iii) Find a minimal sufficient statistic provided that children the probabilities for each of the days of the week is the same, i.e. $p_1 = \ldots = p_7$.

## Example 16. Zero mean Gaussian distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from the normal ditribution $\mathsf{N}(0, \sigma^2)$. Show that the following statistics are not complete.

(i) $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$,

(ii) $T(\boldsymbol{X}) = \sin(X_1) - 1$.

## Example 17. Beta distribution

Let $X_1, \ldots, X_n$ be a random sample from the Beta distribution with the density

$$f(x) = \begin{cases} \dfrac{x^{a-1}(1-x)^{b-1}}{B(a,\, b)}, & 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $a > 0$, $b > 0$ are unknown parameters and $B(a,\, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx$ is a Beta function in points $a$ and $b$.

(i) Find a minimal sufficient statistic for the parameter $(a, b)^\mathsf{T}$.

## Example 18. Two independent samples from the Gausiian distribution

Let $X_1, \ldots, X_n$ be a random sample from the distribution $\mathsf{N}(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_m$ be a random sample from the distribution $\mathsf{N}(\mu_2, \sigma^2)$. The random samples are independent.

(i) Show that

$$S(\boldsymbol{X}, \boldsymbol{Y}) = \left( \sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2, \sum_{i=1}^{m} Y_i, \sum_{i=1}^{m} Y_i^2 \right)^\mathsf{T}$$

is a sufficient statistic.

(ii) Show that the statistic $S(\boldsymbol{X}, \boldsymbol{Y})$ is not complete.

# 3 The use of sufficient statistics in the estimation theory

Let the distribution of our data (represented by random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$) is known up to an unknown parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\mathsf{T}$, which belongs to the parametric space $\Theta$.

**Definition 4.** We say that estimator $T = T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is the best unbiased estimator of the parametric function $a(\boldsymbol{\theta})$, if for each other unbiased estimator $\widetilde{T} = \widetilde{T}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ it holds that

$$\mathsf{var}_{\boldsymbol{\theta}}\big(T\big) \leq \mathsf{var}_{\boldsymbol{\theta}}\big(\widetilde{T}\big), \quad \text{pro} \quad \forall \boldsymbol{\theta} \in \Theta.$$

As we see below, the complete sufficient statistic plays an important role when searching for the best unbiased estimator. The complete sufficient statistic can be easily found in the exponential systems.

**Theorem 3** (About exponential systems)**.** *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent identically distributed random vectors with the density of exponential type, i.e.*

$$f(\boldsymbol{x}; \boldsymbol{\theta}) = q(\boldsymbol{\theta})\, h(\boldsymbol{x}) \, \exp\left\{ \sum_{j=1}^{k} \theta_j R_j(\boldsymbol{x}) \right\},$$

*where $h(\boldsymbol{x}) \geq 0$ a $q(\boldsymbol{\theta}) > 0$. Suppose, that parameteric space contains nondegenerated $k$-dimensional interval. Put*

$$\boldsymbol{S} = (S_1, \ldots, S_k)^\mathsf{T}, \qquad \textit{where} \qquad S_j = \sum_{i=1}^{n} R_j(\boldsymbol{X}_i), \qquad j = 1, \ldots, k.$$

*Then $\boldsymbol{S}$ is a complete sufficient statistic for the parameter $\boldsymbol{\theta}$.*

The following theorem says that the estimator can be „improved“ by conditioning on the sufficient statistic.

**Theorem 4** (Rao-Blackwell theorem)**.** *Let $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be a sufficient statistic and $a(\boldsymbol{\theta})$ is a parametric function that is to be estimated. Let $T = T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be an estimator such that $\mathsf{E}_{\boldsymbol{\theta}}\, T^2 < \infty$ for all $\boldsymbol{\theta} \in \Theta$. Denote $u(\boldsymbol{S}) = \mathsf{E}\,[T|\boldsymbol{S}]$. Then it holds that*

$$\mathsf{E}\, u(\boldsymbol{S}) = \mathsf{E}\, T, \qquad \mathsf{E}\left[T - a(\boldsymbol{\theta})\right]^2 \geq \mathsf{E}\left[u(\boldsymbol{S}) - a(\boldsymbol{\theta})\right]^2,$$

*where the equality holds if and only if $T = u(\boldsymbol{S})$ almost surely.*

First Lehmann-Scheffé theorem says, that if an unbiased estimator is conditioned on the complete sufficient statistic than we get the best unbiased estimate.

**Theorem 5** (The first Lehmann-Scheffé theorem)**.** *Suppose that $T = T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is an unbiased estimator of the parametric function $a(\boldsymbol{\theta})$ such that $\mathsf{E}_{\boldsymbol{\theta}}\, T^2 < \infty$ for all $\boldsymbol{\theta} \in \Theta$. Let $\boldsymbol{S}$ be a complete sufficient statistic for the parameter $\boldsymbol{\theta}$. Define $u(\boldsymbol{S}) = \mathsf{E}\,[T|\boldsymbol{S}]$. Then $u(\boldsymbol{S})$ is the unique best unbiased estimator of $a(\boldsymbol{\theta})$.*

The second Lehmann-Scheffé theorem says that if an unbiased estimator is a function of a complete sufficient statistic then the estimator is the best unbiased estimator.

**Theorem 6** (The second Lehmann-Scheffé theorem)**.** *Let $\boldsymbol{S}$ be a complete sufficient statistic for the parameter $\boldsymbol{\theta}$. Let $g$ be a function such that statistic $W = g(\boldsymbol{S})$ is an unbiased estimator of the parametric function $a(\boldsymbol{\theta})$. Further let $\mathsf{E}_{\boldsymbol{\theta}}\, W^2 < \infty$ for all $\boldsymbol{\theta} \in \Theta$. Then $W$ is the unique best unbiased estimator of $a(\boldsymbol{\theta})$.*

## Example 19. Geometric distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from a geometric distribution, i.e.

$$\mathsf{P}(X_i = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \ldots$$

where $p \in (0, 1)$.

(i) Show that estimator $T(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i = 0\}$ is unbiased estimator of the parameter $p$.

(ii) With the help of sufficient statistic $S(\boldsymbol{X}) = \sum_{i=1}^n X_i$ and Rao-Blackwell theorem „improve“ the estimator $T(\boldsymbol{X})$.

(iii) Is the estimator derived in (ii) the best unbiased estimator of the parameter $p$?

(iv) Analogously as above find the best unbiased estimator of the parametric function $p(1-p)$.

## Example 20. Special multinomial distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from the following version of multinomial distribution

$$\mathsf{P}(X_i = -1) = \mathsf{P}(X_i = 1) = p, \qquad \mathsf{P}(X_i = 0) = 1 - 2\,p,$$

where $p \in (0, \frac{1}{2})$.

(i) Show that the estimator $T(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i = 1\}$ is an unbiased estimator of the parameter $p$.

(ii) Show that $S(\boldsymbol{X}) = \sum_{i=1}^n \mathbb{I}\{X_i \neq 0\}$ is a sufficient statistic for the parameter $p$.

(iii) With the help of $S(\boldsymbol{X})$ and Rao-Blackwell theorem „improve“ the estimator $T(\boldsymbol{X})$.

(iv) Is the estimator found in (iii) the best unbiased estimator of the parameter $p$?

## Example 21. Bernoulli distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from the Bernoulli distribution, i.e.

$$\mathsf{P}(X_i = 1) = p, \qquad \mathsf{P}(X_i = 0) = 1 - p.$$

(i) Find the best unbiased estimator of the parameter $p$.

(ii) Find the best unbiased estimator of the parametric function $p(1-p)$.

## Example 22. Poisson distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from the Poisson distribution with the parameter $\lambda$.

(i) Find the best unbiased estimator of the parameter $\lambda$.

(ii) Find the best unbiased estimator of the parametric function $\mathrm{e}^{-\lambda}$.

### Example 23. Gaussian distribution

Let $X_1, \ldots, X_n$ be a random sample from the Gaussian distribution with the density

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\tfrac{(x-\mu)^2}{2\,\sigma^2} \right\}, \qquad x \in \mathbb{R}.$$

Consider the estimator $\widetilde{\sigma}_n = a_n \sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2}$, where $a_n = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}$.

(i) Show that $S_n^2$ is the best unbiased estimator of the parameter $\sigma^2$.
(ii) Show that $\widetilde{\sigma}_n$ is the best unbiased estimator of $\sigma$.
(iii) Is the sample median the best unbiased estimator of the parameter $\mu$?
(iv) Show that $\overline{X}_n + u_\alpha \widetilde{\sigma}_n$ is the best unbiased estimator of the parametric function $\mu + u_\alpha \sigma$.
(v) Find the best unbiased estimator of the parametric function $\mu^2$.

*Hint. Note that the density of the Gaussian distribution can be written in the form*

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\tfrac{x^2}{2\,\sigma^2} + -\tfrac{2\,x\,\mu}{2\,\sigma^2} \right\} \exp\left\{ -\tfrac{\mu^2}{2\,\sigma^2} \right\}.$$

*Now use Theorem 3 to find that the complete sufficient statistic is given by* $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$.

### Example 24. „Curved normal“ $N(\mu, \mu^2)$

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from the Gaussian distribution with the density

$$f(x; \mu) = \frac{1}{\sqrt{2\pi\mu^2}} \exp\left\{ -\tfrac{(x-\mu)^2}{2\,\mu^2} \right\}, \qquad x \in \mathbb{R}, \qquad \mu > 0.$$

Introduce $T_1(\boldsymbol{X}) = \overline{X}_n$ a $T_2(\boldsymbol{X}) = a_n \sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2}$, where $a_n = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}$.

(i) Show that $T_1(\boldsymbol{X})$ i $T_2(\boldsymbol{X})$ are the unbiased estimators $\mu$ and each of the estimators is a function of the minimal sufficient statistic.
(ii) Show that the variances of the estimators $T_1(\boldsymbol{X})$ and $T_2(\boldsymbol{X})$ are different.

### Example 25. Estimator of the shift in an exponential distribution

Let the random sample $X_1, \ldots, X_n$ come from the distribution with the density

$$f_X(x; \delta) = \begin{cases} \lambda\, e^{-\lambda(x-\delta)}, & x \in (\delta, \infty), \\ 0, & \text{otherwise}, \end{cases}$$

where $\delta \in \mathbb{R}$ and $\lambda$ is known.

(i) Find the best unbiased estimator of the parameter $\delta$.

*Hint: Show that* $\min_{1 \le i \le n} X_i$ *is the complete sufficient statistic and calculate its expectation. From this find a correction so that the estimator is unbiased.*

## Example 26. Estimator of $\lambda$ in exponential distribution

Let $X_1, \ldots, X_n$ be a random sample from exponential distribution with the density

$$f(x; \lambda) = \lambda\, e^{-\lambda\, x}\, \mathbb{I}_{(0,\infty)}(x).$$

(i) Find the best unbiased estimator of the parameter $\lambda$.

(ii) Find the best unbiased estimator of the parametric function $\lambda^k$.

*Hint for (i): Search for the estimator which is a multiple of* $\frac{1}{\overline{X}_n}$. *You can make use of the fact that* $\sum_{i=1}^n X_i$ *has a Gamma distribution with the density* $f(x) = \frac{\lambda^n\, x^{n-1}\, \mathrm{e}^{-\lambda x}}{\Gamma(n)}\, \mathbb{I}_{(0,\infty)}(x)$.

## Example 27. Estimator of $\theta$ in a uniform distribution

Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution $U(0, \theta)$ with the density

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta > 0$.

(i) Is the estimator $\widetilde{\theta}_n = 2\, \overline{X}_n$ the best unbiased estimator of the parameter $\theta$?

(ii) If the answer in (i) is negative then find the best unbiased estimator of the parameter $\theta$.

## Example 28. General multinomial distribution

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent identically distributed random vectors with the multinomial distribution $\mathsf{M}(1; p_1, \ldots, p_K)$, where

$$\mathsf{P}\big(\boldsymbol{X}_1 = (x_1, \ldots, x_K)\big) = p_1^{x_1} \cdots p_K^{x_K},$$

with

$$x_i \in \{0, 1\}, \qquad 0 < p_i < 1, \qquad i = 1, \ldots, K,$$

and

$$\sum_{i=1}^{K} x_i = 1, \qquad \sum_{i=1}^{K} p_i = 1.$$

(i) Find the complete sufficient statistic for the parameter $\boldsymbol{p} = (p_1, \ldots, p_K)^{\mathsf{T}}$.

(ii) Find the best unbiased estimator of the parametric function $a(\boldsymbol{p}) = p_1\, p_2$.

# 4 Method of maximum likelihood - introduction

Let the joint density function of our observations $\mathbf{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ be $p(\boldsymbol{x}; \boldsymbol{\theta})$ (with respect to a $\sigma$-finite measure $\mu$), which depends on an unknown parameter $\boldsymbol{\theta} \in \Theta$. By the **likelihood** we understand a (random) function of the parameter $\boldsymbol{\theta}$:

$$L_n(\boldsymbol{\theta}) = p(\mathbf{X}; \boldsymbol{\theta}),$$

Note si that if the distribution of our observations $\mathbf{X}$ is discrete, then the likelihood $L_n(\boldsymbol{\theta})$ is in fact the probability of our observed data view as a function of the parameter $\boldsymbol{\theta}$.

The maximal likelihood estimator is defined as

$$\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta}).$$

Usually the estimator $\widehat{\boldsymbol{\theta}}_n$ is searched as an argument of the maximum of logarithmic likelihood (**log-likelihood**) $\ell_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta})$. If the density $p(\boldsymbol{x}; \boldsymbol{\theta})$ is „sufficiently smooth“ then the estimator is often searched as a root of the **likelihood equation**

$$\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

In many applications we assume that $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are independent identically distributed random vectors with the density $f(\boldsymbol{x}; \boldsymbol{\theta})$ with respect to a $\sigma$-finite measure $\mu$. Then

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{X}_i; \boldsymbol{\theta}) \quad \text{and} \quad \ell_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\boldsymbol{X}_i; \boldsymbol{\theta}).$$

**Unidimensional parameter**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent identically distributed random vectors from the distribution with the density $f(\boldsymbol{x}; \theta)$ with respect to a $\sigma$-finite measure $\mu$. Then under appropriate regularity assumptions (requiring among others that the support of density $f(\boldsymbol{x}; \theta)$ does not depend on the unknown parameter $\theta$) the maximum likelihood estimator is asymptotically normal and it satisfies

$$\sqrt{n}\,\big(\widehat{\theta}_n - \theta\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, 1/J(\theta)\big), \tag{1}$$

where $J(\theta)$ is the **Fisher information** about parameter $\theta$ in (one) random vector $\boldsymbol{X}_1$. This Fisher information is defined as

$$J(\theta) = \mathsf{E}\Big[\frac{\partial \log f(\boldsymbol{X}_1; \theta)}{\partial \theta}\Big]^2,$$

nevertheless it is usually easier to calculate it as

$$J(\theta) = -\mathsf{E}\Big[\frac{\partial^2 \log f(\boldsymbol{X}_1; \theta)}{\partial \theta^2}\Big].$$

Thus we get that the asymptotic variance (i.e. the variance of the asymptotic distribution) of the maximal likelihood estimator under some appropriate regularity assumptions satisfies

$$\mathsf{avar}\big(\widehat{\theta}_n\big) = \tfrac{1}{n\,J(\theta)}.$$

**Estimator of a transformed parameter.** Sometimes we are interested in a maximal likelihood estimator of a parametric function $g(\theta)$. Let $\widehat{\theta}_n$ be maximal likelihood estimator of the parameter $\theta$. Then $g(\widehat{\theta}_n)$ is the maximal likelihood estimator of the parametric function $g(\theta)$. Moreover if $\widehat{\theta}_n$ satisfies (1) and $g$ is continuously differentiable on the parameter space, then the asymptotic distribution of $g(\widehat{\theta}_n)$ follows from the $\Delta$-method and it holds

$$\sqrt{n}\,\big(g(\widehat{\theta}_n) - g(\theta)\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, [g'(\theta)]^2/J(\theta)\big).$$

Thus

$$\mathsf{avar}\big(g(\widehat{\theta}_n)\big) = \tfrac{[g'(\theta)]^2}{n\,J(\theta)}.$$

## Example 29. Poisson distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the Poisson distribution with the parameter $\lambda$.

(i) Find the maximal likelihood estimator of the parameter $\lambda$ a derive its asymptotic distribution.

(ii) Find the maximal likelihood estimator of the parametric function $\mathrm{e}^{-\lambda}$ a derive its asymptotic distribution.

## Example 30. Exponential distribution

Let the random sample $X_1, \dots, X_n$ come from the distribution with the density

$$f_X(x;\lambda) = \begin{cases} \lambda\,\mathrm{e}^{-\lambda\,x}, & x > 0, \\ 0, & \text{otherwise}, \end{cases}$$

where $\lambda > 0$.

(i) Find the maximal likelihood estimator $\widehat{\lambda}_n$ of the parameter $\lambda$.

(ii) Derive the asymptotic distribution of the estimator found in (i).

## Example 31. Geometric distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the geometric distribution, i.e.

$$\mathsf{P}(X_i = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \dots,$$

where $p \in (0,1)$.

(i) Find the maximal likelihood estimator of the parameter $p$ and derive its asymptotic distribution.

(ii) Find the maximal likelihood estimator of the parametric function $p(1-p)$ a derive its asymptotic distribution.

## Example 32. Uniform distribution $\mathsf{U}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$

Let $X_1, \dots, X_n$ be random sample from uniform distribution $\mathsf{R}\big(\theta - \frac{1}{2}, \theta + \frac{1}{2}\big)$ with the density

$$f(x;\theta) = \begin{cases} 1, & \theta - \frac{1}{2} \le x \le \theta + \frac{1}{2}, \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta \in \mathbb{R}$.

(i) Find the maximal likelihood estimator of the parameter $\theta$.

(ii) Show that the estimator is (weakly) consistent.

## Example 33. Logistic distribution

Let $X_1, \dots, X_n$ be random sample from the logistic distribution with the density

$$f(x;\theta) = \frac{\mathrm{e}^{-(x-\theta)}}{\left(1 + \mathrm{e}^{-(x-\theta)}\right)^2}, \quad x \in \mathbb{R},$$

where $\theta \in \mathbb{R}$.

(i) Find the likelihood equation for the estimator of the parameter $\theta$ and show that the equation has exactly one root.

(ii) Find the asymptotic distribution of the estimator from (i).

## Example 34. Weibullovo distribution

Let $X_1, \dots, X_n$ be random sample from the Weibull distribution with the density

$$f(x;\theta) = \begin{cases} \theta\, x^{\theta-1}\, \mathrm{e}^{-x^\theta}, & x > 0 \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta > 0$.

(i) Write down the likelihood equation for the maximum likelihood estimator of the parameter $\theta$ and show that this equation has a unique root.

(ii) Find the asymptotic distribution of the estimator from (i).

## 5 Neyman-Pearson theorem

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from the distribution with the density $f(\boldsymbol{x}; \boldsymbol{\theta})$ with respect to a $\sigma$-finite measure $\nu$. We are interested in testing hypothesis $H_0 : \boldsymbol{\theta}_X = \boldsymbol{\theta}_0$ against the alternative $H_1 : \boldsymbol{\theta}_X = \boldsymbol{\theta}_1$, where $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$. Put

$$T_n = \frac{\prod_{i=1}^n f(\boldsymbol{X}_i; \boldsymbol{\theta}_1)}{\prod_{i=1}^n f(\boldsymbol{X}_i; \boldsymbol{\theta}_0)},$$

and consider the test of the form

$$T_n \geq c, \tag{2}$$

where $c$ is such a constant so that the test has the level $\alpha$. Then the Neyman-Pearson theorem says that the test with the critical region (2) maximizes the power (i.e. it minimizes the probability of the type II error) among all tests with the level $\alpha$. We also say that such a test is the **most powerful** test.

It is worth noting that $T_n = \frac{L_n(\boldsymbol{\theta}_1)}{L_n(\boldsymbol{\theta}_0)}$, where $L_n(\boldsymbol{\theta})$ is a likelihood at $\boldsymbol{\theta}$.

### Example 35. Poisson distribution

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with the parameter $\lambda$.

(i) Find the most powerful test of the hypotheses

$$H_0 : \lambda_X = \lambda_0, \qquad H_1 : \lambda_X = \lambda_1,$$

where $\lambda_1 > \lambda_0$. Note that this test does not depend on $\lambda_1$

(ii) Modify the test from (i) for a situation when $\lambda_1 < \lambda_0$.

### Example 36. Bernoulli distribution

Let $X_1, \ldots, X_n$ be random sample from a Bernoulli distribution with the parameter $p$.

(i) Find the most powerful test of the hypotheses

$$H_0 : p_X = p_0, \qquad H_1 : p_X = p_1,$$

where $p_1 > p_0$. Does the test depend on the specific choice of the value $p_1$?

(ii) Modify the test from (i) for the situation that $p_1 < p_0$?

### Example 37. Exponential distribution

Let $X_1, \ldots, X_n$ be a random sample from the exponential distribution with the parameter $\lambda$.

(i) Find the most powerful test of the hypotheses

$$H_0 : \lambda_X = \lambda_0, \qquad H_1 : \lambda_X = \lambda_1,$$

where $\lambda_1 > \lambda_0$.

(ii) Modify the test from (i) for a situation when $\lambda_1 < \lambda_0$.

## 6 Method of the maximum likelihood - the vector parameter

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent and identically distributed random vectors (or variables) from the distribution with the density $f(\boldsymbol{x}; \boldsymbol{\theta})$ with respect to a $\sigma$-finite measure $\mu$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^{\mathsf{T}}$ is unknown parameter. Denote the true value of the parameter as $\boldsymbol{\theta}_X$. Then under appropriate regularity assumptions (see for instance Chapter 7.6.5 of the book Anděl: Základy matematické statistiky, 2007, MATFYZPRESS) is the maximum likelihood estimator ($\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{n1}, \dots, \widehat{\theta}_{np})^{\mathsf{T}}$) asymptotically normal and it satisfies

$$\sqrt{n}\,\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_X\big) \xrightarrow[n\to\infty]{d} \mathsf{N}_p\big(\mathbf{0}_p, \boldsymbol{J}^{-1}(\boldsymbol{\theta}_X)\big), \tag{3}$$

where $\boldsymbol{J}(\boldsymbol{\theta})$ is the Fisher information matrix about the parameter $\boldsymbol{\theta}$ in the random vector (veličině) $\boldsymbol{X}_1$.

**Estimation of the asymptotic variance.** Note that (3) implies that the asymptotic variance of maximal likelihood estimator is (in regular cases)

$$\mathsf{avar}(\widehat{\boldsymbol{\theta}}_n) = \tfrac{1}{n}\,\boldsymbol{J}^{-1}(\boldsymbol{\theta}_X).$$

As a consistent estimator of $\boldsymbol{J}(\boldsymbol{\theta}_X)$ we usually use either $\boldsymbol{J}(\widehat{\boldsymbol{\theta}}_n)$ or the empirical Fisher information matrix at the point $\widehat{\boldsymbol{\theta}}_n$, i.e.

$$I_n(\widehat{\boldsymbol{\theta}}_n) = -\frac{1}{n}\,\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathsf{T}}}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n}. \tag{4}$$

### Confidence interval for $\theta_{Xk}$

In applications we are usually interested in confidence intervals for $\theta_{Xk}$ (i.e. for the $k$-th coordinate of the parameter $\boldsymbol{\theta}_X$), where $k = 1, \dots, p$. Denote $\widehat{\theta}_{nk}$ the $k$-th component of the maximal likelihood estimator $\widehat{\boldsymbol{\theta}}_n$. If the asymptotic normality result (3) and $\widehat{\boldsymbol{J}} \xrightarrow[n\to\infty]{P} \boldsymbol{J}(\boldsymbol{\theta}_X)$ hold, then the asymptotic (two-sided) confidence interval is given by

$$\bigg(\widehat{\theta}_{nk} - \tfrac{u_{1-\alpha/2}\sqrt{\widehat{J}^{kk}}}{\sqrt{n}}, \widehat{\theta}_{nk} + \tfrac{u_{1-\alpha/2}\sqrt{\widehat{J}^{kk}}}{\sqrt{n}},\bigg), \tag{5}$$

where $\widehat{J}^{kk}$ is the $k$-th diagonal element of the matrix $\widehat{\boldsymbol{J}}^{-1}$ (i.e. of the inverse matrix of the estimated Fisher information matrix).

### Example 38. Lognormal distribution

Let $X_1, \dots, X_n$ be a random sample from the lognormal distribution with the density

$$f(x; \mu, \sigma^2) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

(i) Find the maximal likelihood estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n, \widehat{\sigma}_n^2)^{\mathsf{T}}$ of the vector parameter $\boldsymbol{\theta} = (\mu, \sigma^2)^{\mathsf{T}}$.

(ii) Derive the asymptotic distribution of the estimator from (i).

(iii) Find the confidence interval for the parameter $\mu$.

### Example 39. Uniform distribution $\mathsf{U}(a, b)$

Let $X_1, \dots, X_n$ be a random sample from the uniform distribution $\mathsf{U}(a, b)$ with the density

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise}, \end{cases}$$

where $a < b$.

(i) Find the maximal likelihood estimator of the vector parameter $(a, b)^\mathsf{T}$.

(ii) Show that the estimator from (i) is (weakly) consistent.

(iii) Calculate

$$\lim_{n\to\infty} \mathsf{P}\big(n\,(\widehat{b}_n - b) \le x\big)$$

and with the help of this result find the the limit distribution of the estimator $\widehat{b}_n$.

### Example 40. Gaussian linear regression model

Suppose you observe independent and identically distributed random vectors $\binom{\boldsymbol{X}_1}{Y_1}, \dots, \binom{\boldsymbol{X}_n}{Y_n}^\mathsf{T}$, where $\boldsymbol{X}_i = (X_{i1}, \dots, X_{ip})^\mathsf{T}$. Let the conditional distribution of $Y_i$ given $\boldsymbol{X}_i$ is Gaussian with the mean $\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}_i$ and variance $\sigma^2$ (for $i = 1, \dots, n$), where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\mathsf{T}$. Further let the distribution $\boldsymbol{X}_i$ does not depend on the parameters $\boldsymbol{\beta}$ and $\sigma^2$. Finally let $\mathsf{E}\,\boldsymbol{X}_i\boldsymbol{X}_i^\mathsf{T}$ be a finite matrix that is not singular.

(i) Find the maximal likelihood estimator of the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\mathsf{T}, \sigma^2)^\mathsf{T}$.

(ii) Derive the asymptotic distribution of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n = \big(\widehat{\boldsymbol{\beta}}_n^\mathsf{T}, \widehat{\sigma}_n^2\big)^\mathsf{T}$ from (i).

(iii) From (ii) deduce the asymptotic distribution of the estimator $\widehat{\boldsymbol{\beta}}_n$.

### Example 41. Model of the logistic regression

Suppose you observe independent and identically distributed random vectors $\binom{\boldsymbol{X}_1}{Y_1}, \dots, \binom{\boldsymbol{X}_n}{Y_n}^\mathsf{T}$, where

$$\mathsf{P}\big(Y_1 = 1 \,|\, \boldsymbol{X}_1\big) = \frac{\exp\{\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}_1\}}{1 + \exp\{\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}_1\}}, \qquad \mathsf{P}\big(Y_1 = 0 \,|\, \boldsymbol{X}_1\big) = \frac{1}{1 + \exp\{\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}_1\}},$$

and the distribution of $\boldsymbol{X}_1$ does not depend on the unknown vector parameter $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\mathsf{T}$. Further let $\mathsf{E}\,\frac{\exp\{\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}_1\}}{(1+\exp\{\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}_1\})^2}\boldsymbol{X}_i\boldsymbol{X}_i^\mathsf{T}$ be a finite matrix that is non-singular.

(i) Derive the asymptotic distribution of the maximal likelihood estimator parameter $\boldsymbol{\beta}$.

(ii) Find the two-sided confidence interval for the parameter $\beta_1$.

# 7 Method of maximum likelihood - asymptotic tests (without nuissance parameters)

### Asymptotic tests for a vector parameter

The null hypothesis $H_0 : \boldsymbol{\theta}_X = \boldsymbol{\theta}_0$ against the alternative $H_1 : \boldsymbol{\theta}_X \neq \boldsymbol{\theta}_0$ can be tested with Wald test, Rao score test or the likelihood ratio test.

Analogously as previously denote $\ell_n(\boldsymbol{\theta})$ the logarithmic likelihood and $\boldsymbol{U}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ its derivative. Further let $\widehat{\boldsymbol{J}}$ be an estimator of $\boldsymbol{J}(\boldsymbol{\theta}_0)$ (Fisher information matrix of one observation at the point of the null hypothesis). Define the following test statistics

$$
\begin{aligned}
W_n &= n\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)^{\mathsf{T}} \widehat{\boldsymbol{J}}\, \big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \qquad &&\text{(Wald test)},\\
R_n &= \frac{1}{n} \left[\boldsymbol{U}_n(\boldsymbol{\theta}_0)\right]^{\mathsf{T}} \widehat{\boldsymbol{J}}^{-1} \boldsymbol{U}_n(\boldsymbol{\theta}_0) \qquad &&\text{(Rao score test)},\\
LR_n &= 2\left(\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\boldsymbol{\theta}_0)\right) \qquad &&\text{(Likelihood ratio test)}.
\end{aligned}
$$

Note that we need the estimator $\widehat{\boldsymbol{J}}$. In Wald test we usually use $\boldsymbol{J}(\widehat{\boldsymbol{\theta}}_n)$ or the empirical Fisher information matrix at the point $\widehat{\boldsymbol{\theta}}_n$, see (4). On the other hand in Rao score test (whose test statistic is sometimes denoted also as $LM_n$) we usually use $\boldsymbol{J}(\boldsymbol{\theta}_0)$ or the empirical Fisher information matrix at the point $\boldsymbol{\theta}_0$. The reason is that then to perform the Rao score test we do not need to calculate the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$.

Under appropriate regularity assumptions (see e.g. Chapter 7.6.5 of the book Anděl: Základy matematické statistiky, 2007, MATFYZPRESS) and under the null hypothesis each of the three tests has asymptotically $\chi^2$-distribution with $p$ degrees of freedom. The large values of the test statistic speak against the null hypothesis. That is why we reject the null hypothesis if the test test statistic is greater (or equal) to $(1-\alpha)$-quantile of $\chi^2$-distribution with $p$ degree of freedom.

### One-dimensional parameter $\theta$

In this special case the test statistics are of the form

$$
\begin{aligned}
W_n &= n\big(\widehat{\theta}_n - \theta_0\big)^2\, \widehat{J} \qquad &&\text{(Wald test)},\\
R_n &= \frac{[U_n(\theta_0)]^2}{n\,\widehat{J}} \qquad &&\text{(Rao score test)},\\
LR_n &= 2\left(\ell_n(\widehat{\theta}_n) - \ell_n(\theta_0)\right) \qquad &&\text{(Likelihood ratio test)}.
\end{aligned}
$$

Under the null hypothesis $H_0 : \theta_X = \theta_0$ each of the test statistics has (under appropriate regularity assumptions) asymptotically $\chi^2$-distribution with one degrees of freedom.

## Example 42. Exponential distribution

Let $X_1, \ldots, X_n$ be a random sample from the distribution

$$
f(x;\lambda) = \begin{cases} \lambda\, \mathrm{e}^{-\lambda x}, & x > 0,\\ 0, & \text{otherwise}, \end{cases}
$$

where $\lambda > 0$.

(i) Derive Wald test, Rao score test a likelihood ratio test for testing the null hypothesis $H_0 : \lambda_X = \lambda_0$ against the alternative $H_1 : \lambda_X \neq \lambda_0$.

### Example 43. Geometric distribution

Consider independent identically distributed random variables $X_1, \dots, X_n$ from a geometric distribution i.e.

$$\mathsf{P}(X_i = k) = p(1-p)^k, \qquad k = 0, 1, 2, \dots$$

where $p \in (0, 1)$ be unknown parameter.

(i) Derive Wald test, Rao score test and the likelihood ratio test for testing the null hypothesis, that $p_X = p_0$ against two-sided alternative $p_X \neq p_0$.

### Example 44. Gaussian distribution

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ be a random sample from the Gaussian distribution $\mathsf{N}(\mu, \sigma^2)$.

(i) Derive Wald test, Rao score test and likelihood ratio test for testing the null hypothesis $H_0 : (\mu, \sigma^2)^\mathsf{T} = (0, 1)^\mathsf{T}$ against $H_1 : (\mu, \sigma^2)^\mathsf{T} \neq (0, 1)^\mathsf{T}$.

### Example 45. Regression in exponential distribution

Let $(X_1, Y_1)^\top, \dots, (X_n, Y_n)^\top$ be independent and identically distributed random vectors. Let the conditional distribution of $Y$ for given $X$ is has the density

$$f_{Y|X}(y|x;\beta) = \beta\, x\, \exp\big\{-\beta\, x\, y\big\}\, \mathbb{I}\{y > 0\},$$

where $\beta > 0$ is an unknown parameter. Further suppose that the distribution of $X$ does not depend on $\beta$.

(i) Find the maximal likelihood estimator for unknown parameter $\beta$.

(ii) Derive Wald test, Rao score test and the likelihood ratio test for testing $H_0 : \beta_X = \beta_0$ against the alternative $H_1 : \beta_X \neq \beta_0$.

### Example 46. Logistic distribution

Let $X_1, \dots, X_n$ be a random sample from the logistic distribution with the density

$$f(x;\theta) = \frac{\mathrm{e}^{-(x-\theta)}}{\big(1 + \mathrm{e}^{-(x-\theta)}\big)^2}, \quad x \in \mathbb{R},$$

where $\theta \in \mathbb{R}$.

(i) Derive Wald test, Rao score test a likelihood ratio test for testing $H_0 : \theta_X = \theta_0$ against the alternative $H_1 : \theta_X \neq \theta_0$.

# 8 Method of maximum likelihood - asymptotic tests with nuisance parameters

Let the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ be a random sample from the distribution with the density $f(x;\boldsymbol{\theta})$ (with respect to a $\sigma$-finite measure $\mu$), where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\mathsf{T}$ is an unknown parameter and $\boldsymbol{\theta}_X$ is its true value. Often we are interested in testing the null hypothesis $H_0 : \boldsymbol{\theta}_X \in \Theta_0$ against the alternative $H_1 : \boldsymbol{\theta} \in \Theta \setminus \Theta_0$, where $\Theta_0$ is subset of the parameter space $\Theta$. Likelihood ratio test for this situation can be written in the form

$$LR_n^* = 2\left(\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\widetilde{\boldsymbol{\theta}}_n)\right), \tag{6}$$

where $\widetilde{\boldsymbol{\theta}}_n$ is the maximal likelihood estimator under the null hypothesis, i.e.

$$\widetilde{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\Theta_0} \ell_n(\boldsymbol{\theta}).$$

Under the null hypothesis and regularity assumption it holds that the test statistic $LR_n^*$ has asymptotically $\chi^2$-distribution with degrees of freedom $\dim(\Theta) - \dim(\Theta_0)$.

In what follows we will treat the special case that we are interested in testing the first $q$ elements $(1 \leq q < p)$ of the vector $\boldsymbol{\theta}$. We denote this subvector as $\boldsymbol{\tau}$. The remaining $p - q$ elements will be denoted as $\boldsymbol{\psi}$ and we will call them nuisance parameters. Thus we can write $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\psi})$ and we want to test

$$H_0 : \boldsymbol{\tau}_X = \boldsymbol{\tau}_0 \text{ against the alternative } H_1 : \boldsymbol{\tau}_X \neq \boldsymbol{\tau}_0, \tag{7}$$

where $\boldsymbol{\psi}$ can be arbitrary.

Denote $\widehat{\boldsymbol{\tau}}_n$ the first $q$ components of the maximal likelihood estimator $\widehat{\boldsymbol{\theta}}_n$ and note that in this case one can write maximal likelihood estimator under the null hypothesis ($\widetilde{\boldsymbol{\theta}}_n$) in the form

$$\widetilde{\boldsymbol{\theta}}_n = (\boldsymbol{\tau}_0, \widetilde{\boldsymbol{\psi}}_n), \quad \text{where} \quad \widetilde{\boldsymbol{\psi}}_n = \arg\max_{\boldsymbol{\psi}} \ell_n(\boldsymbol{\tau}_0, \boldsymbol{\psi}).$$

Let $\boldsymbol{U}_{1n}(\boldsymbol{\tau}, \boldsymbol{\psi}) = \frac{\partial \ell_n(\boldsymbol{\tau},\boldsymbol{\psi})}{\partial \boldsymbol{\tau}}$ be the first $q$ components of the score function. Further denote $\widehat{\boldsymbol{J}}$ the estimator of the Fisher information matrix in a random vector $\boldsymbol{X}_i$ and assume that this estimator is consistent under the null hypothesis.

For testing the hypotheses (7) or one can use either the likelihood ratio test (6) or one of the following tests

$$W_n^* = n\left(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0\right)^\mathsf{T} \left[\widehat{\boldsymbol{J}}^{11}\right]^{-1} \left(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0\right), \qquad \text{(Wald test)},$$

$$R_n^* = \frac{1}{n} \boldsymbol{U}_{1n}^\mathsf{T}(\widetilde{\boldsymbol{\theta}}_n)\, \widehat{\boldsymbol{J}}^{11}\, \boldsymbol{U}_{1n}(\widetilde{\boldsymbol{\theta}}_n), \qquad \text{(Rao score test)},$$

where $\widehat{\boldsymbol{J}}^{11}$ is the upper left $(q, q)-$block of the matrix $\widehat{\boldsymbol{J}}^{-1}$ (i.e. of the inversion of the estimator Fisher information matrix). Each of the test statistics has under the null hypothesis (and under appropriate regularity assumptions) asymptotically $\chi^2$-distribution with $q$ degrees of freedom.

As the estimator of the Fisher information matrix in Wald test we usually use either

$$\widehat{\boldsymbol{J}} = \boldsymbol{J}(\widehat{\boldsymbol{\theta}}_n) \qquad \text{or} \qquad \widehat{\boldsymbol{J}} = -\frac{1}{n} \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\mathsf{T}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n}.$$

In Rao score test it is usually used either

$$\widehat{\boldsymbol{J}} = \boldsymbol{J}\big(\widetilde{\boldsymbol{\theta}}_n\big) \qquad \text{or} \qquad \widehat{\boldsymbol{J}} = -\frac{1}{n}\,\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\mathsf{T}}\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_n}$$

so that we can perform Rao score test without necessity to calculate the (full) maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$.

### Example 47. Gaussian distribution

Consider the random sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ from the Gaussian distribution $N(\mu, \sigma^2)$, where both parameters $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ are unknown. The corresponding density is of the form

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \qquad x \in \mathbb{R}.$$

(i) Derive likelihood ratio test, Rao score test and Wald test of the hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$.

### Example 48. Multinomial distribution

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent abd identically distributed random vectors from the multinomial distribution $\mathsf{M}(1; p_1, p_2, p_3, p_4)$, where

$$\mathsf{P}\big(\boldsymbol{X}_1 = (x_1, x_2, x_3, x_4)\big) = p_1^{x_1} \cdot p_2^{x_2} \cdot p_3^{x_3} \cdot p_4^{x_4},$$

where

$$x_i \in \{0, 1\}, \qquad 0 < p_i < 1, \qquad i = 1, 2, 3, 4$$

and it holds that

$$x_1 + x_2 + x_3 + x_4 = 1, \qquad p_1 + p_2 + p_3 + p_4 = 1.$$

(i) Derive the likelihood ratio test and Wald test of the hypothesis $H_0 : p_1 = \frac{1}{4}$ against the alternative $H_1 : p_1 \neq \frac{1}{4}$.

(ii) Derive the likelihood ratio test for the null hypothesis $H_0 : p_1 = p_2$ against the alternative $H_1 : p_1 \neq p_2$?

(iii) Derive the likelihood ratio test for the null hypothesis $H_0 : p_3 = 1.1\,p_1$ against the alternative $H_1 : p_3 \neq 1.1\,p_1$?

### Example 49. Multinomial distribution

The table below gives the number of lively born children in the Czech Republich in 2008 in different quarters of the year

| Quarter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Number | 28 737 | 30 871 | 31 915 | 28 047 |

With the help of the tests derived in Example 48 find the answer to the following questions.

(i) Can we say that the probability of a child being born in the first quarter is $\frac{1}{4}$?
(ii) Can we say that the probability of child being born in the first quarter is the same as in the second quarter?
(iii) Can we say that the probability of child being born in the third quarter is 1.1-time bigger than in the first quarter?

### Example 50. The simple linear model

Suppose that you observe independent and identically distributed random vectors $(X_1, Y_1)^\mathsf{T}, \ldots, (X_n, Y_n)^\mathsf{T}$ such that that the conditional distribution of $Y_i$ given $X_i$ is $\mathsf{N}(\beta_0 + \beta_1 X_i, \sigma^2)$ and $X_i$ has a distribution with the density $f_X(x)$ not depending on the unknown parameters $\beta_0$, $\beta_1$ a $\sigma^2$.

(i) Find likelihood ratio test null of the hypothesis $H_0 : \beta_1 = 0$ against the alternative that $H_1 : \beta_1 \neq 0$.

### Example 51. Model jednoduché logistické regrese

Suppose you observe independent and identically distributed random vectors $(X_1, Y_1)^\mathsf{T}, \ldots, (X_n, Y_n)^\mathsf{T}$, where

$$\mathsf{P}(Y_1 = 1 \,|\, X_1) = \frac{\exp\{\alpha + \beta\, X_1\}}{1 + \exp\{\alpha + \beta\, X_1\}}, \qquad \mathsf{P}(Y_1 = 0 \,|\, X_1) = \frac{1}{1 + \exp\{\alpha + \beta\, X_1\}},$$

and distribution $X_1$ does not depend on unknown parameters $\alpha$ a $\beta$.

(i) Derive a test of the null hypothesis $H_0 : \beta = 0$ against the alternative that $H_1 : \beta \neq 0$.
(ii) Calculate the $p$-value based on data in the table, where $X_i$ stands for the weight and $Y_i$ for the indicator of too high blood pressure. Calculate also the confidence interval for the parameter $\beta$.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_i$ | 70 | 85 | 76 | 59 | 92 | 102 | 65 | 87 | 73 | 102 |
| $Y_i$ | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |

### Example 52. Regression in exponential distribution

Let $(X_1, Y_1)^\top, \ldots, (X_n, Y_n)^\top$ be independent and identically distributed random vectors such that $Y_1$ given that $X_1 = x$ has an exponential distribution with the density

$$f_{Y|X}(y|x; \alpha, \beta) = \lambda(\alpha, \beta, x)\, \exp\big\{ -\lambda(\alpha, \beta, x)\, y \big\}\, \mathbb{I}\{y > 0\},$$

where $\lambda(\alpha, \beta, x) = \mathrm{e}^{\alpha + \beta\, x}$ a $\alpha$, $\beta$ are unknown parameters. Further assume that the distribution of $X_1$ does not depend on parameters $\alpha$ and $\beta$.

(i) Derive the likelihood ratio test, Rao score test and Wald test of the null hypothesis $\beta = 0$ against two-sided alternative that $\beta \neq 0$.

*For instance, you can think of* $Y$ *has a time to a breakdown of a given product and* $X$ *as the maximal temperature during the manufacturing of this product. Note that under the null hypothesis* $X$ *and* $Y$ *are independent.*

# 9 Results of some examples

### Example 1

(i) $\mathsf{E}\,(\,XY \mid X = x\,) = \frac{x(3x+2)}{3(2x+1)}$, for $x \in (0,1)$.

### Example 2

(i) $\mathsf{E}\big[\frac{Y}{X^2}|X\big] = 2\,X$.

(ii) $\mathsf{E}\,\frac{Y}{X^2} = 1$.

(iii) $\mathsf{E}\,Y = \frac{1}{4}$.

(iv) $\mathsf{var}(Y) = \frac{37}{28}$.

### Example 3

(i) $\mathsf{E}\big(Y \,\big|\, X = t\big) = t$ for $t \in (1,2)$ and $\mathsf{E}\big(Y \,\big|\, X\big) = X$.

(ii) $\mathsf{E}\Big(Y \,\Big|\, \log\big(\frac{X-1}{2-X}\big) = t\Big) = \frac{2\exp\{t\}+1}{\exp\{t\}+1}$ for $t \in (-\infty,\infty)$ and $\mathsf{E}\Big(Y \,\big|\, \log\big(\frac{X-1}{2-X}\big)\Big) = X$.

(iii) $\mathsf{E}\Big[\frac{Y}{X^6} \,\big|\, \log\big(\frac{X-1}{2-X}\big)\Big] = \frac{1}{X^5}$.

### Example 4

(i) $\mathsf{E}\big[Y|\exp\{X\}\big] = \frac{X^2+1}{2}$.

(ii) $\mathsf{E}\,Y = 1$.

(iii) $\mathsf{var}(Y) = 1$.

### Example 8

(i) $\boldsymbol{X}$ is sufficient.

(ii) $\big(|X_1|,\ldots,|X_n|\big)^{\mathsf{T}}$ is sufficient.

(iii) $\sum_{i=1}^n X_i$ is not sufficient.

(iv) $\sum_{i=1}^n |X_i|$ is not sufficient.

(v) $\sum_{i=1}^n X_i^2$ is sufficient.

(vi) $\frac{1}{n}\sum_{i=1}^n X_i^2$ is sufficient.

(vii) $\big(\frac{1}{n}\sum_{i=1}^{n-1} X_i^2, X_n^2\big)^{\mathsf{T}}$ is sufficient.

### Example 10

(i) $S(\boldsymbol{X}) = \big(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\big)^{\mathsf{T}}$

## Example 14

(i) $\big(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\big)^\mathsf{T}$

(ii) Statistic from (i) is not complete.

## Example 17

(i) $\Big(\sum_{i=1}^{n} \log(X_i), \sum_{i=1}^{n} \log(1 - X_i)\Big)^\mathsf{T}$

## Example 18

(ii) Consider the statistic $S_X^2 - S_Y^2$.

## Example 19

(ii) $\frac{1-\frac{1}{n}}{\left(\overline{X}_n+1-\frac{1}{n}\right)}$

(iii) Yes.

(iv) $\frac{\overline{X}_n\left(1-\frac{1}{n}\right)}{\left(\overline{X}_n+1-\frac{1}{n}\right)\left(\overline{X}_n+1-\frac{2}{n}\right)}$

## Example 21

(i) $\overline{X}_n$.

(ii) $\frac{n}{n-1}\,\overline{X}_n\big(1 - \overline{X}_n\big)$.

## Example 22

(i) $\overline{X}_n$.

(ii) $\big(1 - \frac{1}{n}\big)^{\sum_{i=1}^{n} X_i}$.

## Example 23

(i) It is sufficient to show that the estimator is unbiased (this is known) and that it is function of the complete sufficient statistic.

(ii) Similar as in (i), but here it is rather technical to show that the estimator is unbiased.

(iii) No, it cannot be as it is not a function of the complete sufficient statistic.

(iv) Similarly as in (i) and (ii).

(v) We need to find an unbiased estimator that is a function a complete sufficient statistic. A straightforward estimator would be $\left(\overline{X}_n\right)^2$. Try to calculate $\mathsf{E}\left(\overline{X}_n\right)^2$. Then find $a \in \mathbb{R}$ such that the estimator $W = \left(\overline{X}_n\right)^2 - a\,S_n^2$ is unbiased.

## Example 24

Viz Example 7.57 from Anděl: Základy matematické statistiky, 2007, MATFYZPRESS.

## Example 25

(i) $\widehat{\delta}_n = \min_{1\leq i\leq n} X_i - \frac{1}{n\,\lambda}$

## Example 26

(i) $\widehat{\lambda}_n = \frac{n-1}{\sum_{i=1}^{n} X_i}$

## Example 27

(i) Estimator $\widetilde{\theta}_n = 2\,\overline{X}_n$ is unbiased, but it is not the best unbiased estimator.
(ii) $\frac{n+1}{n}\max_{1\leq i\leq n} X_i$.

## Example 28

(i) $\boldsymbol{T} = \big(\sum_{i=1}^{n} X_{1i}, \ldots, \sum_{i=1}^{n} X_{(K-1)i}\big)^{\mathsf{T}}$
(ii) $\frac{1}{n(n-1)} \sum_{i=1}^{n} X_{1i} \sum_{i=1}^{n} X_{2i}$

## Example 29

(i) $\overline{X}_n$.
(ii) The maximal likelihood estimator is $\mathrm{e}^{-\overline{X}_n}$, and it holds that

$$\sqrt{n}\big(\mathrm{e}^{-\overline{X}_n} - \mathrm{e}^{-\lambda}\big) \xrightarrow{d} \mathsf{N}\big(0, \lambda\,\mathrm{e}^{-2\,\lambda}\big),$$

## Example 30

(i) $\widehat{\lambda}_n = \frac{1}{\overline{X}_n}$
(ii) $\sqrt{n}\big(\widehat{\lambda}_n - \lambda\big) \xrightarrow{\mathsf{d}} \mathsf{N}\big(0, \lambda^2\big)$

## Example 31

(i) $\widehat{p}_n = \frac{1}{1+\overline{X}_n}$ , $\sqrt{n}(\widehat{p}_n - p_X) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, p^2(1-p)\big)$
(ii) $\widehat{p}_n(1-\widehat{p}_n) = \frac{\overline{X}_n}{(1+\overline{X}_n)^2}$, $\sqrt{n}(\widehat{p}_n(1-\widehat{p}_n) - p(1-p)) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, (1-2p)^2p^2(1-p)\big)$

## Example 32

(i) The maximal likelihood estimator is any of the values from the interval

$$\Big(\max_{1\leq i\leq n} X_i - \tfrac{1}{2}, \min_{1\leq i\leq n} X_i + \tfrac{1}{2}\Big).$$

(ii) The estimator from (i) is consistent as $\max_{1\leq i\leq n} X_i \xrightarrow{\mathsf{P}} \theta + \frac{1}{2}$ and $\min_{1\leq i\leq n} X_i \xrightarrow{\mathsf{P}} \theta - \frac{1}{2}$ for $n \to \infty$.

## Example 33

(i) See Example 7.96 of the book of Anděl.

(ii) Note that the estimator is given only implicitly. Thus one needs to use the general result (1), which gives us that

$$\sqrt{n}\big(\widehat{\theta}_n - \theta\big) \xrightarrow{\mathsf{d}} \mathsf{N}(0,3).$$

## Example 34

(i) See Example 7.99 of the book of Anděl.

(ii) $J(\theta) = \frac{1}{\theta^2} + \mathsf{E}\, X^\theta \log^2(X) = \frac{1}{\theta^2}(1 + \int_0^\infty y^2 \log^2(y)\, \mathrm{e}^{-y} dy)$.

## Example 35

(i) The test has the critical region $\sum_{i=1}^n X_i \geq c$, where one can take $c$ as the $(1-\alpha)$-quantile of the distribution $\mathsf{Po}(n\lambda_0)$. The test does not depend on the choice of $\lambda_1$ from which one can conclude that the test is the most powerful test for testing $H_0 : \lambda_X = \lambda_0$ against $H_1 : \lambda_X > \lambda_0$.

(ii) In this situation the test is of the form $\sum_{i=1}^n X_i \leq c$.

## Example 36

(i) The test has the critical region $\sum_{i=1}^n X_i \geq c$, where one can take $c$ as the $(1-\alpha)$-quantile of the distribution $\mathsf{Bi}(n, p_0)$. The test does not depend on the choice of $p_1$ from which one can conclude that the test is the most powerful test for testing $H_0 : p_X = p_0$ against $H_1 : p_X > p_0$.

(ii) In this situation the test is of the form $\sum_{i=1}^n X_i \leq c$.

## Example 37

(i) The test has the critical region $\sum_{i=1}^n X_i \leq c$.

(ii) The test would have a critical region $\sum_{i=1}^n X_i \geq c$.

## Example 38

Put $Y_i = \log X_i$.

(i) $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n, \widehat{\sigma}_n^2)^\mathsf{T} = \big(\overline{Y}_n, \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2\big)^\mathsf{T}$.

(ii)

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\mu}_n \\ \widehat{\sigma}_n^2\end{pmatrix} - \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}\right) \xrightarrow[n\to\infty]{d} \mathsf{N}_2\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2, 0 \\ 0,\ 2\sigma^4\end{pmatrix}\right)$$

(iii) $(\overline{X}_n - \frac{u_{1-\alpha/2}\,\widehat{\sigma}_n}{\sqrt{n}}, \overline{X}_n + \frac{u_{1-\alpha/2}\,\widehat{\sigma}_n}{\sqrt{n}})$

## Example 39

(i) $\big(\min_{1\leq i\leq n} X_i, \max_{1\leq i\leq n} X_i\big)^\mathsf{T}$.

(ii) Estimator from (i) is consistent.

(iii) $\lim_{n\to\infty} \mathsf{P}\big(n\,(\widehat{b}_n - b) \leq x\big) = \exp\{\frac{x}{b-a}\}$ for $x < 0$. For $x \geq 0$ is this probability equal to 1. Thus $n\,(\widehat{b}_n - b) \xrightarrow[n\to\infty]{d} -Y$, where $Y$ has an exponential distribution with the parameter $\frac{1}{b-a}$.

## Example 40

(i) $\widehat{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^{\mathsf{T}}\right)^{-1} \sum_{i=1}^n \boldsymbol{X}_i Y_i$ and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \widehat{\boldsymbol{\beta}}_n^{\mathsf{T}} \boldsymbol{X}_i\right)^2$.

## Example 41

(i) $\sqrt{n}\,(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow[n\to\infty]{d} \mathsf{N}_p\big(\mathbf{0}_p, \big[\mathsf{E}\,\frac{\exp\{\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{X}_i\}}{(1+\exp\{\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{X}_i\})^2}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}\big]^{-1}\big)$;

(ii) $\big(\widehat{\beta}_{n1} \mp u_{1-\alpha/2}\sqrt{\frac{\widehat{J}^{11}}{n}}\big)$, where $\widehat{J}^{11}$ is the first diagonal element of the matrix

$$\left[\frac{1}{n}\sum_{i=1}^n \frac{\exp\{\widehat{\boldsymbol{\beta}}_n^{\mathsf{T}}\boldsymbol{X}_i\}}{(1+\exp\{\widehat{\boldsymbol{\beta}}_n^{\mathsf{T}}\boldsymbol{X}_i\})^2}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}\right]^{-1}.$$

Note that here we do not know the distribution of $\boldsymbol{X}_i$. Thus one cannot use $J(\widehat{\boldsymbol{\beta}}_n)$ as an estimate of the Fisher information matrix $J(\boldsymbol{\beta}) = \mathsf{E}\,\frac{\exp\{\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{X}_i\}}{(1+\exp\{\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{X}_i\})^2}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}$.

## Example 42

(i) MLE is $\widehat{\lambda}_n = \frac{1}{\overline{X}_n}$.

$$W_n = \frac{n\,(\widehat{\lambda}_n - \lambda_0)^2}{\widehat{\lambda}_n^2} = \left(\frac{\sqrt{n}\,(\widehat{\lambda}_n - \lambda_0)}{\widehat{\lambda}_n}\right)^2,$$

$$R_n = \left(\frac{\sqrt{n}\,(\widehat{\lambda}_n - \lambda_0)}{\lambda_0}\right)^2,$$

$$LR_n = 2\left[n\log\tfrac{\widehat{\lambda}_n}{\lambda_0} - \sum_{i=1}^n X_i\,(\widehat{\lambda}_n - \lambda_0)\right].$$

The null hypothesis is rejected if the value of the given test statistic is greater than (or equal to) $\chi_1^2(1-\alpha)$.

## Example 42

(i) MLE is $\widehat{p}_n = \frac{1}{1+\overline{X}_n}$.

$$W_n = \frac{n(\widehat{p}_n - p_0)^2}{\widehat{p}_n^2(1-\widehat{p}_n)}$$

$$R_n = \big(\tfrac{n}{p_0} - \tfrac{\sum_{i=1}^n X_i}{1-p_0}\big)^2\, n\, p_0^2(1-p_0),$$

$$LR_n = 2\left[n\,\log\tfrac{\widehat{p}_n}{p_0} + \sum_{i=1}^n X_i\log\tfrac{1-\widehat{p}_n}{1-p_0}\right].$$

## Example 44

(i) MLE: $\widehat{\mu}_n = \overline{X}_n$ and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

$$W_n = \frac{n\,(\widehat{\mu}_n - 0)^2}{\widehat{\sigma}_n^2} + \frac{n\,(\widehat{\sigma}_n^2 - 1)^2}{2\widehat{\sigma}_n^4} \geq \chi_2^2(1-\alpha),$$

$$R_n = \frac{\big(\sum_{i=1}^n X_i\big)^2}{n} + \frac{\big(\sum_{i=1}^n \big[X_i^2 - 1\big]\big)^2}{2n} \geq \chi_2^2(1-\alpha),$$

$$LR_n = -n\log\widehat{\sigma}_n^2 + \sum_{i=1}^n \big(X_i^2 - 1\big) \geq \chi_2^2(1-\alpha).$$

## Example 45

(i) $\widehat{\beta}_n = \frac{1}{\frac{1}{n}\sum_{i=1}^n Y_i X_i}$

(ii)

$$W_n = \frac{n\,(\widehat{\beta}_n - \beta_0)^2}{\widehat{\beta}_n^2}$$

$$R_n = \Big(\tfrac{n}{\beta_0} - \sum_{i=1}^n X_i Y_i\Big)^2 \beta_0^2$$

$$LR_n = 2\Big[n\log\tfrac{\widehat{\beta}_n}{\beta_0} - \sum_{i=1}^n X_i Y_i\,(\widehat{\beta}_n - \beta_0)\Big].$$

## Example 46

(i) For the asymptotic distribution of MLEE see Example 33. From this example we know that $J(\theta) = \frac{1}{3}$.

$$W_n = \frac{n\,(\widehat{\theta}_n - \theta_0)^2}{3}$$

$$R_n = \frac{\Big(n - \sum_{i=1}^n \frac{2e^{\theta_0 - X_i}}{1+e^{\theta_0 - X_i}}\Big)^2}{\frac{n}{3}}$$

$$LR_n = 2\,n\,(\widehat{\theta}_n - \theta_0) - 4\sum_{i=1}^n \log\Big(\tfrac{1+e^{\theta_0 - X_i}}{1+e^{\widehat{\theta}_n - X_i}}\Big)$$

It is worth noting that to calculate Rao score test we do not need to find $\widehat{\boldsymbol{\theta}}_n$ (which is given only implicitly as a root of a nonlinear equation). Thus we can perform Rao score test without special numerical software.

## Example 47

$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be MLE of $\sigma^2$ (without restrictions) and $\widetilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$ be the MLE estimator of $\sigma^2$ under $H_0$.

(i) $LR_n^* = n\log\frac{\widetilde{\sigma}_n^2}{\widehat{\sigma}_n^2}$, $W_n^* = \frac{n\,(\overline{X}_n - \mu_0)^2}{\widehat{\sigma}_n^2}$, $R_n^* = \frac{n\,(\overline{X}_n - \mu_0)^2}{\widetilde{\sigma}_n^2}$.
The critical region is always of the form $T_n \geq \chi_1^2(1-\alpha)$, where $T_n$ is one of the above test statistics.

## Example 48

Denote $Y_k = \sum_{i=1}^{n} X_{ik}$ pro $k \in \{1, \dots, 4\}$. Then MLE (without restrictions) is

$$\widehat{\boldsymbol{p}}_n = \left(\widehat{p}_{n1}, \dots, \widehat{p}_{n4}\right)^\mathsf{T} = \left(\tfrac{Y_1}{n}, \dots, \tfrac{Y_4}{n}\right)^\mathsf{T}.$$

The asymptotic distribution can be deduced directly from the central limit theorem (think why it is not possible to use the the general result about the asymptotic normality of MLE).

The likelihood ratio test is of the form

$$LR_n^* = 2 \sum_{k=1}^{4} X_k \log\left(\tfrac{\widehat{p}_{nk}}{\widetilde{p}_{nk}}\right) \geq \chi_1^2(1-\alpha).$$

(i) The estimate of $\boldsymbol{p}$ under the null hypothesis for the likelihood ration test is given by

$$\left(\widetilde{p}_{n1}, \dots, \widetilde{p}_{n4}\right)^\mathsf{T} = \left(\frac{1}{4}, \tfrac{3\,Y_2}{4\sum_{k=2}^{4} Y_k}, \tfrac{3\,Y_3}{4\sum_{k=2}^{4} Y_k}, \tfrac{3\,Y_4}{4\sum_{k=2}^{4} Y_k}\right)^\mathsf{T}.$$

Wald test has a critical region

$$\left| \frac{\sqrt{n}\left(\widehat{p}_{1n} - \frac{1}{4}\right)}{\sqrt{\widehat{p}_{1n}(1-\widehat{p}_{1n})}} \right| \geq u_{1-\alpha/2}$$

(ii) MLE of $\boldsymbol{p}$ under the null hypothesis

$$\left(\widetilde{p}_{n1}, \dots, \widetilde{p}_{n4}\right)^\mathsf{T} = \left(\tfrac{Y_1+Y_2}{2n}, \tfrac{Y_1+Y_2}{2n}, \widehat{p}_{n3}, \widehat{p}_{n4}\right)^\mathsf{T}.$$

(iii) MLE of $\boldsymbol{p}$ under the null hypothesis

$$\left(\widetilde{p}_{n1}, \dots, \widetilde{p}_{n4}\right)^\mathsf{T} = \left(\tfrac{Y_1+Y_3}{2.1\,n}, \widehat{p}_{n2}, \tfrac{1.1(Y_1+Y_3)}{2.1\,n}, \widehat{p}_{n4}\right)^\mathsf{T}.$$

## Example 50

MLE (without restrictions) $\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i-\overline{Y}_n)(X_i-\overline{X}_n)}{\sum_{i=1}^{n}(X_i-\overline{X}_n)^2}$, $\widehat{\beta}_0 = \overline{Y}_n - \widehat{\beta}_1\,\overline{X}_n$ and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1\,X_i)^2$.

MLE under the null hypothesis (i.e. $\beta_1 = 1$) is given $\widetilde{\beta}_0 = \overline{Y}_n$ and $\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{\beta}_0)^2$.

(i) $LR_n^* = n \log \frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} \geq \chi_2^2(1-\alpha)$,

## Example 51

MLE $\binom{\widehat{\alpha}_n}{\widehat{\beta}_n}$ withou restriction is given implicitly as a solution of the following likelihood equations

$$\sum_{i=1}^{n} \left[Y_i - \tfrac{\exp\{\alpha+\beta\,X_i\}}{1+\exp\{\alpha+\beta\,X_i\}}\right] = 0,$$

$$\sum_{i=1}^{n} X_i \left[Y_i - \tfrac{\exp\{\alpha+\beta\,X_i\}}{1+\exp\{\alpha+\beta\,X_i\}}\right] = 0.$$

MLE under the null hypothesis that $\beta = 0$ is given by $\binom{\widetilde{\alpha}_n}{0}$ where $\widetilde{\alpha}_n = \log\left(\frac{\overline{Y}_n}{1-\overline{Y}_n}\right)$.

As we do not know the distribution of $\boldsymbol{X}_i$ we estimate the Fisher information matrix as

$$\widehat{\boldsymbol{J}}(\alpha, \beta) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{\exp\{\alpha+\beta X_i\}}{(1+\exp\{\alpha+\beta X_i\})^2}, & \frac{1}{n}\sum_{i=1}^{n} \frac{X_i \exp\{\alpha+\beta X_i\}}{(1+\exp\{\alpha+\beta X_i\})^2} \\ \frac{1}{n}\sum_{i=1}^{n} \frac{X_i \exp\{\alpha+\beta X_i\}}{(1+\exp\{\alpha+\beta X_i\})^2}, & \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^2 \exp\{\alpha+\beta X_i\}}{(1+\exp\{\alpha+\beta X_i\})^2} \end{pmatrix}$$

(i) $W_n^* = \left(\frac{\sqrt{n}(\widehat{\beta}_n - 0)}{\sqrt{\widehat{J}^{22}(\widehat{\alpha}_n, \widehat{\beta}_n)}}\right)^2$, where $\widehat{J}^{22}(\widehat{\alpha}_n, \widehat{\beta}_n)$ is the second diagonal element of the inversion of the matrix $\widehat{\boldsymbol{J}}(\widehat{\alpha}_n, \widehat{\beta}_n)$.
$R_n^* = \frac{1}{n}\left(\sum_{i=1}^{n} X_i(Y_i - \frac{e^{\widetilde{\alpha}_n}}{1+e^{\widetilde{\alpha}_n}})\right)^2 \widehat{J}^{22}(\widetilde{\alpha}_n, 0)$, where $\widehat{J}^{22}(\widetilde{\alpha}_n, 0)$ is the second diagonal element of the inversion matrix $\widehat{\boldsymbol{J}}(\widetilde{\alpha}_n, 0)$.

| | test. stat. | p-value |
|---|---|---|
| $LR_n^*$ | 1.14 | 0.29 |
| $R_n^*$ | 1.08 | 0.30 |
| $W_n^*$ | 0.98 | 0.32 |

(ii) Asymptotic confidence interval for $\beta$ is given by

$$\left(\widehat{\beta}_n - u_{1-\alpha/2}\sqrt{\frac{\widehat{J}^{22}(\widehat{\alpha}_n, \widehat{\beta}_n)}{n}}, \widehat{\beta}_n + u_{1-\alpha/2}\sqrt{\frac{\widehat{J}^{22}(\widehat{\alpha}_n, \widehat{\beta}_n)}{n}}\right).$$

For the given data and $\alpha = 0.05$ the confidence interval is $(-0.055, 0.168)$.