UNITEXT 173

Francesca Gasperoni · Francesca Ieva
Anna Maria Paganoni

Exercise Book of Statistical Inference

Springer

UNITEXT

La Matematica per il 3+2

Volume 173

The **UNITEXT - La Matematica per il 3+2** series is designed for undergraduate and graduate academic courses, and also includes books addressed to PhD students in mathematics, presented at a sufficiently general and advanced level so that the student or scholar interested in a more specific theme would get the necessary background to explore it.

Originally released in Italian, the series now publishes textbooks in English addressed to students in mathematics worldwide.

Some of the most successful books in the series have evolved through several editions, adapting to the evolution of teaching curricula.

Submissions must include at least 3 sample chapters, a table of contents, and a preface outlining the aims and scope of the book, how the book fits in with the current literature, and which courses the book is suitable for.

For any further information, please contact the Editor at Springer: francesca.bonadei@springer.com

THE SERIES IS INDEXED IN SCOPUS

Francesca Gasperoni • Francesca Ieva •
Anna Maria Paganoni

Exercise Book of Statistical Inference

Springer

Francesca Gasperoni
Department of Mathematics
Politecnico di Milano
Milano, Italy

Francesca Ieva
Department of Mathematics
Politecnico di Milano
Milano, Italy

Anna Maria Paganoni
Department of Mathematics
Politecnico di Milano
Milano, Italy

ISSN 2038-5714 ISSN 2532-3318 (electronic)
UNITEXT
ISSN 2038-5722 ISSN 2038-5757 (electronic)
La Matematica per il 3+2
ISBN 978-3-031-86669-2 ISBN 978-3-031-86670-8 (eBook)
https://doi.org/10.1007/978-3-031-86670-8

This book is a translation of the original Italian edition “Eserciziario di Statistica Inferenziale” by Francesca Gasperoni et al., published by Springer-Verlag Italia S.r.l. in 2020. The translation was done with the help of an artificial intelligence machine translation tool. A subsequent human revision was done primarily in terms of content, so that the book will read stylistically differently from a conventional translation. Springer Nature works continuously to further the development of tools for the production of books and on the related technologies to support the authors.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Statistical inference is the discipline that underpins stochastic modelling, which is that area of Mathematics where uncertainty is part of the model and object of interest in the study. Calculus, linear algebra, and probability are among the main subjects on which the theory of statistical inference is based, with its main objective being the estimation of quantities of interest such as, the parametric and non-parametric laws of stochastic models, their relative asymptotic distributions, etc. In this framework, the study of advanced statistical models, such as, linear regression models, analysis of variance (ANOVA), and generalised regression models, is essential both in research and in business. Consequently, the implementation of algorithms in a suitable statistical software is proposed as a natural completion of the book.

This text was created with the aim of helping the students in the transition between the theoretical and methodological concepts of statistical inference and their software implementation. The first part of the text is mainly focused on exercises to be solved with *with pen and paper*, in order to apply notions derived from lemmas and theorems; while in the second part of the text we propose assignments, based on both the manual implementation of algorithms and the application of *built-in tools* for an effective analysis of datasets that are derived from real problems.

To optimise the understanding of the selected topics, and to accompany the reader in their study, the text is organised into chapters, that are composed of an introductory part, in which the theoretical foundations of statistical inference are recalled, and a second part that is composed of exercises, accompanied by a comprehensive solution on paper and, if appropriate, on software. In particular, for a thorough treatment of the theoretical part, refer to [3] and [5].

Regarding the computational solutions, the use of the statistical software `R` [6] (version 3.5.1) is proposed. This choice was guided by the fact that `R` is available for various operating systems (Unix, GNU/Linux, Mac OS X, Microsoft Windows) and can be downloaded for free from the website http://cran.r-project.org/. Moreover, in `R` there is a wide choice of libraries (packages) distributed and appropriately described on the Comprehensive R Archive Network (CRAN).

The text is organised into six main areas: a first area includes basic probability exercises (Chap. 1); a second area addresses the topic of point estimators (Chaps. 2, 3, and 4); a third area is focused on hypothesis testing and confidence intervals (Chaps. 5, 6, and 7); a fourth area focuses on the asymptotic properties of estimators (Chap. 8); and a fifth area is focused on multiple linear regression models, generalised regression and analysis of variance (Chaps. 9, 10, and 11). Regarding these three chapters, supplementary material is available online, containing the datasets needed to carry out some exercises, further insights and exercises. Finally, there is a last chapter, containing summary exercises, through which the student can gain a global view of the data analysis techniques illustrated in the book.

This text is written for students of undergraduate courses in Statistics, Mathematics, Engineering and for postgraduate courses in Data Science. Many of the exercises and laboratories proposed are derived from exercises and exam topics of the course Models and Methods for Statistical Inference taught in Mathematical Engineering at the Politecnico of Milan. We therefore thank the numerous colleagues and collaborators who have contributed, directly or indirectly, to the creation of the proposed material. In particular, an important contribution to the development of the Exercises and Laboratories must be recognised to Andrea Ghiglietti, Matteo Gregoratti, and Nicholas Tarabelloni.

Milan, Italy
January 2025

Francesca Gasperoni
Francesca Ieva
Anna Maria Paganoni

Contents

Acronyms

AIC	Akaike Inference Criterion
ARE	Asymptotic Relative Efficiency
BIC	Bayesian Inference Criterion
i.i.d.	independent and identically distributed
IC	Confidence Interval
K-R	Karlin-Rubin
l	log-Likelihood function
L	Likelihood function
L-S	Lehmann-Scheffé Theorem
LFGN	Strong Law of Large Numbers
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimator
MLR	Monotone Likelihood Ratio
MSE	Mean Square Error
N-P	Neyman-Pearson
R	Rejection region
TCL	Central Limit Theorem
UMP	Uniformly Most Powerful
UMVUE	Uniform Minimum Variance Unbiased Estimator
v.a.	random variable

Part I
Inferential Statistics

Chapter 1
Fundamentals of Probability and Statistics

1.1 Theory Recap

1.1.1 Expected Value, Variance and Covariance

Theorem 1.1 *Let X be a real r.v. with distribution function $F_X(x)$, let $Y = g(X)$, let $\mathcal{X} = \{x : f(x) > 0\}$ and let $\mathcal{Y} = \{y : f_Y(y) > 0\}$:*

- *If $g(\cdot)$ is an increasing function in $\mathcal{X}$, then $F_Y(y) = F_X(g^{-1}(y)) \quad \forall y \in \mathcal{Y}$.*
- *If $g(\cdot)$ is a decreasing function in $\mathcal{X}$ and X is a continuous r.v. then $F_Y(y) = 1 - F_X(g^{-1}(y)) \quad \forall y \in \mathcal{Y}$.*
- *Suppose that $f_X(x)$ is continuous in $\mathcal{X}$ and that $g^{-1}(\cdot)$ has a continuous derivative in $\mathcal{Y}$. Then the density of Y is as follows:*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}| & y \in \mathcal{Y}; \\ 0 & \textit{otherwise}. \end{cases}$$

Theorem 1.2 *Let X be a r.v. with continuous distribution function $F_X(x)$. Then the r.v. $Y = F_X(X)$ has law $Y \sim U(0, 1)$.*

Definition 1.1 (Mean) The expected value or mean of a r.v. $g(X)$ is defined as:

$$\mathbb{E}[g(X)] = \begin{cases} \int_{-\infty}^{+\infty} g(x) f_X(x)\, \mathrm{d}x & \text{if } X \text{ is a continuous r.v.;} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x)\, \mathrm{d}x & \text{if } X \text{ is a discrete r.v..} \end{cases}$$

Theorem 1.3 *Let X be a r.v.. Let a, b, c be scalars in $\mathbb{R}$. Then for any functions $g_1(x)$ and $g_2(x)$ for which the mean exists, the following hold:*

- $\mathbb{E}[ag_1(X) + bg_2(X) + c] = a\mathbb{E}[g_1(X)] + b\mathbb{E}[g_2(X)] + c$.
- *If $g_1(x) \geq 0 \quad \forall x$, then $\mathbb{E}[g_1(X)] \geq 0$.*

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_1

- *If* $g_1(x) \geq g_2(x) \quad \forall x$, *then* $\mathbb{E}[g_1(X)] \geq \mathbb{E}[g_2(X)]$.
- *If* $a \leq g_1(x) \leq b \quad \forall x$, *then* $a \leq \mathbb{E}[g_1(X)] \leq b$.

Definition 1.2 (Variance) The variance of a r.v. X is defined as:

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

and its square root is called the standard deviation.

Theorem 1.4 *Let X be a r.v. with finite variance. Let a, b, c be scalars in $\mathbb{R}$. Then the following holds:*

$$Var(aX + b) = a^2 Var(X).$$

Definition 1.3 (Covariance) Let X and Y be two r.v., then the covariance is defined as:

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Theorem 1.5 *Let X and Y be two r.v. with finite variance and a and b be two scalars. Then:*

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

Definition 1.4 (Correlation) Let X and Y be two r.v., then the correlation is defined as:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Theorem 1.6 *Let X and Y be two r.v. Then:*

- $\rho_{X,Y} \in [-1, 1]$.
- $|\rho_{X,Y}| = 1$ *if and only if there exists a number* $a \neq 0$ *and* b *such that* $\mathbb{P}\{Y = aX + b\} = 1$. *If* $a > 0$ *then* $\rho_{X,Y} = 1$, *if* $a < 0$ *then* $\rho_{X,Y} = -1$.

1.1.2 Joint and Marginal Laws

Theorem 1.7 *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a vector of r.v. with joint density $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then the marginal law of $X_1, \ldots, X_k$ is:*

$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} f_{\boldsymbol{X}}(\boldsymbol{x})\, dx_{k+1} dx_{k+2} \ldots dx_n \quad \text{if r.v. are continuous.}$$

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \sum_{(x_{k+1},\ldots,x_n)\in\mathbb{R}^{n-k}} f_{\boldsymbol{X}}(\boldsymbol{x}) \quad \textit{if r.v. are discrete}.$$

Definition 1.5 (Conditional Laws) Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a vector of r.v. with joint density $f_{\boldsymbol{X}}(\boldsymbol{x})$. $\forall(x_1,\ldots,x_k) \in \mathbb{R}^k$ such that $f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) > 0$, the conditional law of $\boldsymbol{X}$ given $(X_1,\ldots,X_k) = (x_1,\ldots,x_k)$ is a function of $(x_1,\ldots,x_k)$, $f_{X_1,\ldots X_n|X_1,\ldots,\ X_k}(x_1,\ldots x_n|x_1,\ldots,\ x_k)$, which is defined as:

$$f_{X_1,\ldots X_n|X_1,\ldots,\ X_k}(x_1,\ldots x_n|x_1,\ldots,\ x_k) = \frac{f_{X_{k+1},\ldots X_n}(x_{k+1},\ldots x_n)}{f_{X_1,\ldots X_k}(x_1,\ldots x_k)}.$$

Lemma 1.1 (Independence) *Let* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ *be a vector of r.v. with joint density* $f_{\boldsymbol{X}}(\boldsymbol{x})$. $X_1, X_2, \ldots X_n$ *are mutually independent r.v. if and only if:*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Theorem 1.8 (Independence) *If* X_1 *and* X_2 *are independent r.v. then* $Cov(X_1, X_2) = 0$.

1.1.3 Conditional Expected Values

Theorem 1.9 (Double Conditional Expected Value) *Let* X *and* Y *be two r.v., then:*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$$

if the expected values exist.

Theorem 1.10 (Conditional Variance) *Let* X *and* Y *be two r.v., then:*

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y])$$

if the expected values exist.

1.1.4 Convergences

Definition 1.6 (Almost Sure Convergence) A sequence of r.v. $X_1, X_2, \ldots$ converges almost surely to a r.v. X if $\forall \varepsilon > 0$ it holds:

$$\mathbb{P}\{\lim_{n \to +\infty} |X_n - X| < \varepsilon\} = 1$$

and it is denoted by $X_n \stackrel{q.c.}{\rightarrow} X$.

Definition 1.7 (Convergence in Probability) A sequence of r.v. $X_1, X_2, \ldots$ converges in probability to a r.v. X if $\forall \varepsilon > 0$ it holds:

$$\lim_{n \to \infty} \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0$$

or equivalently

$$\lim_{n \to \infty} \mathbb{P}\{|X_n - X| < \varepsilon\} = 1$$

and it is denoted by $X_n \stackrel{p}{\rightarrow} X$.

Theorem 1.11 (Strong Law of Large Numbers, SLLN) *Consider a sequence $X_1, X_2, \ldots$ of i.i.d. r.v., such that $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2 < +\infty$. Consider $\overline{X}_n = \sum_{i=1}^n X_i/n$. Then $\forall \varepsilon > 0$ it holds:*

$$\mathbb{P}\{\lim_{n \to +\infty} |\overline{X}_n - \mu| < \varepsilon\} = 1$$

that is, $\overline{X}_n$ converges almost surely to μ ($\overline{X}_n \stackrel{q.c.}{\rightarrow} \mu$).

Theorem 1.12 (Weak Law of Large Numbers) *Consider a sequence $X_1, X_2, \ldots$ of i.i.d. r.v., such that $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2 < +\infty$. Consider $\overline{X}_n = \sum_{i=1}^n X_i/n$. Then $\forall \varepsilon > 0$ it holds:*

$$\lim_{n \to +\infty} \mathbb{P}\{|\overline{X}_n - \mu| < \varepsilon\} = 1;$$

that is, $\overline{X}_n$ converges in probability to μ ($\overline{X}_n \stackrel{p}{\rightarrow} \mu$).

Theorem 1.13 *Consider a sequence of r.v. $X_1, X_2, \ldots$ that converges in probability to a r.v. X and let h be a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

Definition 1.8 (Convergence in Distribution) A sequence of r.v. $X_1, X_2, \ldots$ converges in distribution to a r.v. X if it holds:

$$\lim_{n \to +\infty} F_{X_n}(x) = F_X(x)$$

$\forall x$ where $F_X(x)$ is continuous and it is denoted $X_n \xrightarrow{\mathcal{L}} X$.

Theorem 1.14 *Consider a sequence of r.v. $X_1, X_2, \ldots$ that converges to a r.v. X:*

- *Almost sure convergence implies convergence in probability.*
- *Convergence in probability implies convergence in distribution.*
- *Convergence in distribution implies convergence in probability only if $X_1, X_2, \ldots$ converges to a constant.*

Theorem 1.15 (Slutsky's Theorem) *If $X_n \xrightarrow{\mathcal{L}} X$ and $Y_n \xrightarrow{p} a$, where a is constant then:*

$$X_n Y_n \xrightarrow{\mathcal{L}} aX.$$

$$X_n + Y_n \xrightarrow{\mathcal{L}} a + X.$$

Theorem 1.16 (Central Limit Theorem, CLT) *Consider a sequence $X_1, X_2, \ldots$ of i.i.d. r.v.s, such that $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2 < +\infty$. Consider $\overline{X}_n = \sum_{i=1}^{n} X_i/n$. Then it holds:*

$$\lim_{n\to+\infty} \mathbb{P}\left\{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq x\right\} = \phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{y^2/2}\, dy;$$

that is, $\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma} \xrightarrow{\mathcal{L}} Z \sim N(0, 1)$.

Theorem 1.17 (Delta Method) *Consider a sequence $X_1, X_2, \ldots$ of r.v.s, such that $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{\mathcal{L}} X \sim N(0, \sigma^2)$. Consider a specific function $g(\cdot)$ and a specific value θ. Suppose that $g'(\theta)$ exists and is non-zero. Then:*

$$\sqrt{n}(g(\overline{X}_n) - g(\theta)) \xrightarrow{\mathcal{L}} X \sim N(0, \sigma^2[g'(\theta)]^2).$$

If $g'(\theta) = 0$:

$$\sqrt{n}(g(\overline{X}_n) - g(\theta)) \xrightarrow{\mathcal{L}} X \sim \frac{\sigma^2}{2} g''(\theta)\chi^2(1).$$

For further study, refer to Chapters 1, 2 and 4 [3].

1.2 Exercises

Exercise 1.1 The joint law of the discrete random variables X and Y is partially described in Table 1.1.

Table 1.1 Joint law of variables X and Y

	Y = 2	Y = 4	
X = 0		0.1	0.3
X = 1	0.1		0.4
X = 2			
		0.6	

(a) Complete the table and state whether X and Y are independent.
(b) Calculate the law, the expected value and the conditional variance of Y given $X = 0$.
(c) Calculate the law, the expected value and the conditional variance of X given $Y = 2$.
(d) Calculate $\mathbb{E}[X|Y]$.

Exercise 1.2 Let (X, Y) be a continuous random vector with uniform distribution on the set

$$V = \Big\{(x, y) \in \mathbb{R}^2 \,:\, x \geq 0,\ y \geq 0,\ x^2 + y^2 \leq 9\Big\}.$$

(a) Write the density of (X, Y). Are the variables X and Y independent?
(b) Calculate $\mathbb{E}[X|Y]$.

Exercise 1.3 Calculate $\mathbb{E}[Y|X]$ for the pair of random variables (X, Y) with joint density

$$f(x, y) = \begin{cases} \frac{4}{5}\,(x + 3y)\,\mathrm{e}^{-x-2y}, & x, y > 0; \\ 0, & otherwise. \end{cases}$$

Exercise 1.4 The concentration X of a certain pollutant in a given volume of exhaust gas from an industrial process is uniformly distributed between 0 and 1 mg/m^3. A purification process has been developed that allows to reduce the concentration of that substance: if x is the concentration of pollutant in a given volume of gas subjected to purification, the concentration Y after purification is uniformly distributed between 0 and px mg/m^3, where $p \in (0, 1)$ is a given parameter.

(a) Determine the joint distribution of X and Y.
(b) Determine the distribution of Y.
(c) Are the two variables independent?
(d) If the concentration Y of pollutant after purification is known, what is the expected value for the corresponding concentration X before purification?

Exercise 1.5 During the drafting of a book, a preliminary version of the work is read by the author. Knowing that the number of errors on a page is a random variable with a Poisson distribution of parameter $\lambda = 3$, and that each error is discovered (in one reading) with probability $p = 0.7$, calculate:

(a) The law of the number of errors discovered on a page (e.g. the first one).
(b) The expected number of errors discovered on a page.
(c) The probability that two errors are discovered on the first page knowing that there are at most three.

Exercise 1.6 Let X and Y be two independent Bernoulli random variables with parameter p. Let $Z = I_{(X+Y=0)}$ be the indicator of the event $X + Y = 0$. Calculate $\mathbb{E}[X|Z]$ and $\mathbb{E}[Y|Z]$. Are these random variables still independent?

Exercise 1.7 Consider a random vector (X, Y) such that X has a uniform distribution over the interval $[0, 1]$ and, conditionally on $X = x$, Y has a Gaussian law with mean x and variance x^2.

(a) Explicitly write the conditional density $f_{(Y|X)}(y|x)$.
(b) Explicitly write the joint density $f_{(X,Y)}(x, y)$.
(c) Calculate $\mathbb{E}[Y|X]$.
(d) Calculate $\mathbb{E}[Y]$.
(e) Calculate $\mathrm{Var}[Y|X]$.
(f) Calculate $\mathrm{Var}[Y]$.

Exercise 1.8 Let (X, Y) be a continuous random vector with

$$f_Y(y) = \begin{cases} \dfrac{(1/2)^{1/2}\, y^{-1/2}\, \mathrm{e}^{-y/2}}{\Gamma(1/2)}, & y > 0; \\ 0, & y \leq 0. \end{cases}$$

$$f_{X|Y}(x|y) = (2\pi)^{-1/2}\, y^{1/2}\, \mathrm{e}^{-yx^2/2}, \quad x \in \mathbb{R}.$$

(a) Show that for every $y > 0$ there exists $\mathbb{E}[X|Y = y]$.
(b) Show that there exists $\mathbb{E}\big[\mathbb{E}[X|Y]\big]$ and calculate its value.
(c) Show that however $\mathbb{E}[X]$ does not exist.

Exercise 1.9 Consider $X_1, \ldots, X_n$ independent Bernoulli variables all with parameter p, where $n \geq 2$. Let Z be their sum and let $Y = X_1 + X_2 - X_1 X_2$ be the variable that indicates if there has been at least one success in the first two trials.

(a) Calculate $\mathbb{E}[X_1|Z]$ and $\mathbb{E}[X_2|Z]$ and their limits for $n \to \infty$.
(b) Determine the law of Y, calculate $\mathbb{E}[Y|Z]$ and its limit for $n \to \infty$.

Exercise 1.10 Let X_n be a sequence of independent random variables such that, for every i

$$\mathbb{P}(X_i > x) = \begin{cases} 1, & x \leq 1, \\ x^{-\lambda}, & x > 1. \end{cases}$$

where $\lambda > 1$

(a) Calculate the density of the r.v. X_i.
(b) Calculate the mean of the r.v. X_i.
(c) Determine the law of the r.v. $Y_i = \log(X_i)$.
(d) Study the convergence of the sequence of r.v. $\left\{(X_1 X_2 \dots X_n)^{1/n}\right\}$.

Exercise 1.11 Let $X_1, \dots, X_n, \dots$ be a sequence of independent and identically distributed r.v. with a uniform law in the interval $[0, \lambda]$, with $\lambda > 0$.

(a) Calculate, for each fixed n, the distribution function of the r.v. $T_n = n \min(X_1, \dots, X_n)$.
(b) Prove that the sequence of r.v. $T_1, \dots, T_n, \dots$ converges in law to a r.v. Y and identify the law of Y.

Exercise 1.12 Let (X_n) be a sequence of independent r.v. all with a Poisson law of parameter λ. What is the value, varying λ, of the limit

$$\lim_{n\to\infty} \mathbb{P}(X_1 + \cdots + X_n < n)?$$

Exercise 1.13 Let $\{X_n\}_{n\in\mathbb{N}^*}$ be a sequence of r.v. such that $X_n \sim \chi^2(n)$ for every $n \in \mathbb{N}^*$. Does the sequence X_n/n admit a limit? In what sense?

1.3 Solutions

1.1

(a) In the last column on the right we read the marginal law of X, $\mathbb{P}\{X = x\}$.
In the last row at the bottom we read the marginal law of Y, $\mathbb{P}\{Y = y\}$.
We can therefore complete the joint law, as reported in Table 1.2.

Table 1.2 Joint law of the variables X and Y

	Y = 2	Y = 4	
X = 0	0.2	0.1	0.3
X = 1	0.1	0.3	0.4
X = 2	0.1	0.2	0.3
	0.4	0.6	1

It is immediately apparent that they are not independent, as for example:

$$\mathbb{P}\{X = 0, Y = 2\} = 0.2 \neq \mathbb{P}\{X = 0\}\mathbb{P}\{Y = 2\} = 0.12.$$

(b)

$Y\|X = 0$	2	4	
	2/3	1/3	1

$$\mathbb{E}[Y|X = 0] = 2 \cdot 2/3 + 4 \cdot 1/3 = 8/3.$$

$$\text{Var}(Y|X = 0) = \mathbb{E}[Y^2|X = 0] - (\mathbb{E}[Y|X = 0])^2 = 4 \cdot 2/3 + 16 \cdot 1/3 - 64/9 = 8/9.$$

(c)

$X\|Y = 2$	0	1	2	
	1/2	1/4	1/4	1

$$\mathbb{E}[X|Y = 2] = 1 \cdot 1/4 + 2 \cdot 1/4 = 3/4.$$

$$\text{Var}(X|Y = 2) = \mathbb{E}[X^2|Y = 2] - (\mathbb{E}[X|Y = 2])^2 = 1 \cdot 1/4 + 4 \cdot 1/4 - 9/16 = 11/16.$$

(d)

$X\|Y = 4$	0	1	2	
	1/6	1/2	1/3	1

$$\mathbb{E}[X|Y = 4] = 1 \cdot 1/2 + 2 \cdot 1/3 = 7/6.$$

Therefore:

$$\mathbb{E}[X|Y] = \frac{3}{4} \cdot \mathbb{I}_{\{Y=2\}} + \frac{7}{6} \cdot \mathbb{I}_{\{Y=4\}}.$$

We note that $\mathbb{E}[X|Y]$ is a random variable function of Y.

1.2

(a) Given that the area of V is $\frac{9}{4}\pi$, the density of the vector (X, Y) is:

$$f_{X,Y}(x, y) = \frac{4}{9\pi}\mathbb{I}_V(x, y).$$

X and Y are not independent given that:

$$f_X(x) = \int_0^{\sqrt{9-x^2}} \frac{4}{9\pi}\,\mathrm{d}x = \frac{4}{9\pi}\sqrt{9-x^2}\,\mathbb{I}_{[0,3]}(x)$$

and by symmetry:

$$f_Y(y) = \frac{4}{9\pi}\sqrt{9-y^2}\,\mathbb{I}_{[0,3]}(y).$$

Therefore:

$$f_{(X,Y)}(x, y) \neq f_X(x) \cdot f_Y(y).$$

The support of (X, Y), V, is not factorisable.

(b)

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{9-y^2}}\,\mathbb{I}_{\left[0,\sqrt{9-y^2}\right]}(x) \quad \forall y : 0 \leq y \leq 3.$$

Therefore $X|Y$ has a uniform law on the interval $[0, \sqrt{9-y^2}]$ and therefore:

$$\mathbb{E}[X|Y] = \frac{\sqrt{9-y^2}}{2}.$$

1.3 We exploit the definition:

$$\mathbb{E}[Y|X = x] = \int_0^{+\infty} y \cdot f_{Y|X}(y|x)dy = \int_0^{+\infty} y \cdot \frac{f_{X,Y}(x, y)}{f_X(x)}dy.$$

Given that:

$$\begin{aligned} f_X(x) &= \int_0^{+\infty} f_{X,Y}(x, y)dy = \int_0^{+\infty} \frac{4}{5}\,(x + 3y)\,\mathrm{e}^{-x-2y}dy = \\ &= \frac{4}{5}x\mathrm{e}^{-x} \cdot \left[-\frac{\mathrm{e}^{-2y}}{2}\right]_0^{+\infty} + \frac{4}{5}3\mathrm{e}^{-x} \cdot \int_0^{+\infty} y\mathrm{e}^{-2y}dy = \\ &= \frac{4}{5}x\mathrm{e}^{-x}\frac{1}{2} + \frac{4}{5}3\mathrm{e}^{-x}\frac{1}{4} = \frac{1}{5}(2x + 3)\mathrm{e}^{-x}. \end{aligned}$$

We obtain:

$$f_{Y|X}(y|x) = \frac{\frac{4}{5}\,(x + 3y)\,\mathrm{e}^{-x-2y}}{\frac{1}{5}(2x + 3)\mathrm{e}^{-x}} = 4\left(\frac{x + 3y}{2x + 3}\right)e^{-2y}.$$

Substituting in the starting formula, we have:

$$\mathbb{E}[Y|X=x] = \int_0^{+\infty} y \cdot 4\Big(\frac{x+3y}{2x+3}\Big)e^{-2y}dy = \frac{4}{2x+3}\int_0^{+\infty}(xy\mathrm{e}^{-2y}+3y^2\mathrm{e}^{-2y})\,dy =$$
$$= \frac{4}{2x+3}(x\cdot 1/4 + 3\cdot 1/4) = \frac{x+3}{2x+3}.$$

Hence $\mathbb{E}[Y|X] = \frac{X+3}{2X+3}$.

1.4

(a) We know that: $X \sim \mathrm{U}_{[0,1]}$ and $Y|X=x \sim \mathrm{U}_{[0,px]}$. Hence:

$$f_{Y|X}(y|x) = \frac{1}{px}\,\mathbb{I}_{[0,px]}(y).$$

The joint law of (X, Y) will therefore be:

$$f_{X,Y}(x,y) = \frac{1}{px}\,\mathbb{I}_{[0,px]}(y)\cdot \mathbb{I}_{[0,1]}(x).$$

(b)

$$f_Y(y) = \int_{y/p}^{1}\frac{1}{px}\mathrm{d}x = -\frac{1}{p}\log\frac{y}{p}\,\mathbb{I}_{[0,p]}(y).$$

(c) We immediately observe that X and Y are not independent.
(d) We need to calculate $\mathbb{E}[X|Y]$.

$$f_{X|Y}(x|y) = \frac{\frac{1}{px}\cdot\mathbb{I}_{[0,px]}(y)\cdot\mathbb{I}_{[0,1]}(x)}{-\frac{1}{p}\log\left(\frac{y}{p}\right)\mathbb{I}_{[0,p]}(y)} =$$
$$= -\frac{1}{x}\frac{1}{\log\left(\frac{y}{p}\right)}\,\mathbb{I}_{[0,p]}(y)\cdot\mathbb{I}_{[y/p,1]}(x);$$

from which:

$$\mathbb{E}[X|Y=y] = \int_{y/p}^{1} -\frac{1}{\log\left(\frac{y}{p}\right)}\,\mathrm{d}x\cdot\mathbb{I}_{[0,p]}(y) =$$
$$= -\frac{1}{\log\left(\frac{y}{p}\right)}\cdot\left(1-\frac{y}{p}\right)\mathbb{I}_{[0,p]}(y).$$

$$\mathbb{E}[X|Y] = -\frac{1}{\log\left(\frac{Y}{p}\right)} \cdot \left(1 - \frac{Y}{p}\right) \mathbb{I}_{[0,p]}(Y).$$

1.5

(a) We define the following r.v. and derive their distributions:

- E = 'number of errors present on a page', $E \sim \mathcal{P}(\lambda)$.
- S = 'number of errors discovered on a page'.
- $S|E = n \sim Bin(n, p)$.

We calculate $\mathbb{P}\{S = k\}$, using the theorem of total probabilities:

$$\begin{aligned}
\mathbb{P}\{S = k\} &= \sum_{n=0}^{+\infty} \mathbb{P}\{S = k|E = n\} \cdot \mathbb{P}\{E = n\} \\
&= \sum_{n=k}^{+\infty} \binom{n}{k} p^k (1-p)^{(n-k)} \cdot \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \frac{p^k e^{-\lambda}}{k!} \cdot \sum_{n=k}^{+\infty} \frac{\lambda^{n-k+k}}{(n-k)!}(1-p)^{(n-k)} \\
&= \frac{(\lambda p)^k e^{-\lambda}}{k!} \cdot \sum_{n=k}^{+\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} \\
&= \frac{(\lambda p)^k e^{-\lambda}}{k!} \cdot e^{(\lambda - \lambda p)} = \frac{(\lambda p)^k e^{-\lambda p}}{k!} \quad k \geq 0.
\end{aligned}$$

We can therefore say that: $S \sim \mathcal{P}(\lambda p)$.

(b) $\mathbb{E}[S] = \lambda p$.

(c)

$$\begin{aligned}
\mathbb{P}\{S = 2|E \leq 3\} &= \frac{\mathbb{P}\{S = 2, E \leq 3\}}{\mathbb{P}\{E \leq 3\}} = \frac{\sum_{n=2}^{3} \mathbb{P}\{S = 2|E = n\}\mathbb{P}\{E = n\}}{\mathbb{P}\{E \leq 3\}} = \\
&= \frac{p^2 \cdot \frac{e^{-\lambda}\lambda^2}{2!} + \binom{3}{2} p^2 (1-p)^1 \cdot \frac{e^{-\lambda}\lambda^3}{3!}}{\frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^3}{3!}} = \\
&= \frac{0.1097 + 0.0988}{0.6472} = 0.3223.
\end{aligned}$$

1.6

(a) Note that $Z \sim Be((1-p)^2)$.

$$\mathbb{E}[X|Z=0] = \mathbb{P}\{X=1|Z=0\} = \frac{\mathbb{P}(X=1, Z=0)}{\mathbb{P}(Z=0)} = \frac{\mathbb{P}(X=1)}{\mathbb{P}(Z=0)} =$$
$$= \frac{p}{1-(1-p)^2} = \frac{p}{p(2-p)} = \frac{1}{2-p}.$$

$$\mathbb{E}[X|Z=1] = 0.$$

Hence:

$$\mathbb{E}[X|Z] = \frac{1}{2-p}\mathbb{I}_{\{0\}}(Z) = \mathbb{E}[Y|Z];$$

where the last equality is due to obvious reasons of symmetry.

Now $\mathbb{E}[X|Z] \not\perp\!\!\!\perp \mathbb{E}[Y|Z]$, in fact:

$$\mathbb{P}\left\{\mathbb{E}[X|Z] = \frac{1}{2-p}, \mathbb{E}[Y|Z] = \frac{1}{2-p}\right\} = \mathbb{P}\{Z=0\} = 2p-p^2 \neq (2p-p^2)^2.$$

1.7

$$X \sim U[0,1] \qquad Y|X=x \sim N(x, x^2).$$

(a)

$$f_{(Y|X)}(y|x) = \frac{1}{\sqrt{2\pi x^2}} e^{-\frac{(y-x)^2}{2x^2}}.$$

(b)

$$f_{(X,Y)}(x,y) = f_{(Y|X)}(y|x) f_X(x) = \frac{1}{\sqrt{2\pi x^2}} e^{-\frac{(y-x)^2}{2x^2}} \mathbb{I}_{[0,1]}(x).$$

(c)

$$\mathbb{E}[Y|X=x] = x \implies \mathbb{E}[Y|X] = X.$$

(d)

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[X] = 1/2.$$

(e)

$$\mathrm{Var}(Y|X) = X^2.$$

(f)

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(\mathbb{E}[Y|X]) + \mathbb{E}(\mathrm{Var}[Y|X]) = \\
&= \mathrm{Var}(X) + \mathbb{E}[X^2] = \\
&= 2 \cdot \mathrm{Var}(X) + (\mathbb{E}[X])^2 = \\
&= 2 \cdot \frac{1}{2} + \frac{1}{4} = \frac{5}{12}.
\end{aligned}$$

1.8

(a) $y > 0$. We calculate:

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} y^{1/2} \exp\left\{ -y \cdot x^2/2 \right\} \mathrm{d}x = 0$$

for obvious reasons of symmetry. Therefore: $\mathbb{E}[X|Y] = 0$.

(b) $\mathbb{E}[\mathbb{E}[X|Y]] = 0$.

(c)

$$\begin{aligned}
f_X(x) &= \int_0^{+\infty} f_{X|Y}(x|y) \cdot f_Y(y) = \\
&= \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} y^{1/2} \exp\{-yx^2/2\} \cdot \frac{1}{\sqrt{2}} \cdot \frac{y^{-1/2}}{\Gamma\left(\frac{1}{2}\right)} \exp\{-y/2\} = \\
&= \frac{1}{2\sqrt{\pi}\,\Gamma\left(\frac{1}{2}\right)} \cdot \int_0^{+\infty} \exp\{-y/2(1 + x^2)\}\,\mathrm{d}y = \\
&= \frac{1}{2\sqrt{\pi}\,\Gamma\left(\frac{1}{2}\right)} \cdot \frac{1}{1/2 \cdot (1 + x^2)} = \frac{1}{\pi(1 + x^2)}.
\end{aligned}$$

Therefore $X \sim Cauchy$ and $\mathbb{E}[X]$ does not exist.

1.9

(a)

$$\mathbb{E}[X_1|Z=k]=\frac{\mathbb{P}\{X_1=1,Z=k\}}{\mathbb{P}\{Z=k\}}=\frac{p\cdot\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}}=\frac{k}{n}.$$

Therefore:

$$\mathbb{E}[X_1|Z]=\mathbb{E}[X_2|Z]=\frac{Z}{n}=\frac{\sum Y_i}{n}\xrightarrow{q.c.}p$$

by the strong law of large numbers.

(b)

$$Y=X_1+X_2-X_1\cdot X_2.$$

Y can only take value 0 with probability $(1-p)^2$ or 1 with probability $1-(1-p)^2=2p-p^2$.
Therefore: $Y\sim Be(2p-p^2)$.

$$\begin{aligned}\mathbb{E}[Y|Z=k]&=\mathbb{E}[X_1|Z=k]+\mathbb{E}[X_2|Z=k]-\mathbb{E}[X_1X_2|Z=k]=\\&=\frac{2k}{n}-\frac{p^2\binom{n-2}{k-2}p^{k-2}(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}}=\\&=\frac{2k}{n}-\frac{k(k-1)}{n(n-1)}.\end{aligned}$$

Therefore:

$$\mathbb{E}[Y|Z]=\frac{2Z}{n}-\frac{Z(Z-1)}{n(n-1)}\xrightarrow{q.c.}2p-p^2;$$

always by the strong law of large numbers.

1.10

(a)

$$f_{X_i}(x)=\lambda\cdot x^{-(\lambda+1)}\,\mathbb{I}_{\{1,+\infty\}}(x);$$

which is obtained by deriving the distribution function $F_{X_i}(x)=1-\mathbb{P}\{X_i>x\}$.

(b)

$$\mathbb{E}[X]=\int_1^{+\infty}\lambda x^{-\lambda}\,\mathrm{d}x=\frac{\lambda}{\lambda-1}.$$

(c)

$$F_{Y_i}(y) = \mathbb{P}\{\log X_i \leq y\} = \mathbb{P}\{X_i \leq e^y\} = 1 - e^{-\lambda y} \quad \forall y > 0.$$

So $Y_i \sim \mathcal{E}(\lambda)$.

(d) Given that $\frac{1}{n}\sum log(X_i) \xrightarrow{a.s.} \frac{1}{\lambda}$ by the strong law of large numbers, then:

$$\left(\prod_{i=1}^{n} X_i\right)^{1/n} = \exp\left[\log\left(\prod_{i=1}^{n} X_i\right)^{1/n}\right] = \exp\left[\frac{\sum \log X_i}{n}\right] \xrightarrow{a.s.} e^{1/\lambda}.$$

1.11

(a)

$$\begin{aligned}
\mathbb{P}\{T_n \leq t\} &= \mathbb{P}\{n \min(X_1, \ldots, X_n) \leq t\} = \\
&= 1 - \mathbb{P}\{\min(X_1, \ldots, X_n) > t/n\} = \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}\{X_i > t/n\} = \\
&= 1 - \prod_{i=1}^{n} \left(1 - \frac{t}{n\lambda}\right) = \\
&= 1 - \left(1 - \frac{t}{n\lambda}\right)^n.
\end{aligned}$$

(b)

$$1 - \left(1 - \frac{t}{n\lambda}\right)^n \to 1 - e^{-t/\lambda}.$$

So $Y \sim \mathcal{E}(1/\lambda)$.

1.12 One can notice that:

$$\lim_{n\to\infty} \mathbb{P}(X_1 + \cdots + X_n < n) = \lim_{n\to\infty} \mathbb{P}(\overline{X}_n < 1).$$

Given that $\mathbb{E}[X_i] < \infty$ and $Var(X_i) < \infty$, we can apply the strong law of large numbers, which guarantees:

$$\overline{X}_n \xrightarrow{a.s.} \mathbb{E}[X_i] = \lambda.$$

The almost sure convergence implies convergence in distribution, therefore:

$$F_{\overline{X}_n}(t) \to F_\lambda(t).$$

Considering that λ is a constant, we can write $F_\lambda(t) = \mathbb{I}_{[\lambda,+\infty)}(t)$. So we conclude that:

$$\lim_{n\to\infty} \mathbb{P}(\overline{X}_n < 1) = \begin{cases} 1 & \text{if } \lambda < 1; \\ 0 & \text{if } \lambda > 1. \end{cases}$$

It remains to study the case $\lambda = 1$, as it is a point of discontinuity for $F_\lambda(t)$. If $\lambda = 1$, from the Central Limit Theorem we know that:

$$\lim_{n\to\infty} \sqrt{n} \cdot (\overline{X}_n - \mathbb{E}[X_i]) \xrightarrow{\mathcal{L}} N(0, \text{Var}(X_i)).$$

In this specific case, given that $\lambda = 1$, we can write:

$$\lim_{n\to\infty} \sqrt{n} \cdot (\overline{X}_n - 1) \xrightarrow{\mathcal{L}} N(0, 1).$$

Which translates to:

$$\mathbb{P}\{\overline{X}_n < 1\} = \mathbb{P}\{\sqrt{n} \cdot (\overline{X}_n - 1) < (1 - 1) \cdot \sqrt{n}\} = \mathbb{P}\{\sqrt{n} \cdot (\overline{X}_n - 1) < 0\}.$$

$$\mathbb{P}\{\sqrt{n} \cdot (\overline{X}_n - 1) < 0\} \to \mathbb{P}\{Z < 0\} = \Phi(0) = 1/2.$$

So we conclude that:

$$\lim_{n\to\infty} \mathbb{P}(\overline{X}_n < 1) = \begin{cases} 1 & \text{if } \lambda < 1; \\ 1/2 & \text{if } \lambda = 1; \\ 0 & \text{if } \lambda > 1. \end{cases}$$

1.13 Given $X_n \sim \chi^2(n)$ then there exists a sequence of i.i.d. r.v. $Y_1, Y_2, \ldots, Y_n$ such that $X_n = \sum Y_i$ and $Y_i \sim \chi^2(1)$, or $Y_i \sim \Gamma(1/2, 1/2)$.

We note that $\mathbb{E}[Y_i] = 1 < \infty$ and that $\text{Var}(Y_i) < \infty$.

We can therefore apply the Strong Law of Large Numbers and conclude that:

$$\frac{X_n}{n} = \frac{\sum Y_i}{n} \xrightarrow{a.s.} \mathbb{E}[Y_i] = 1.$$

Chapter 2
Sufficient, Minimal and Complete Statistics

2.1 Theory Reminders

Definition 2.1 (Statistic) Let $X_1, .., X_n$ be a sample of r.v. We define a statistic $T(\boldsymbol{X})$ as any function of the sample.

Definition 2.2 (Sufficient Statistics) A statistic $T(\boldsymbol{X})$ is sufficient for a parameter θ if the conditional distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X})$ does not depend on θ.

Theorem 2.1 (Factorisation) *Let $f(\boldsymbol{x}; \theta)$ be the joint distribution of a sample of r.v. $\boldsymbol{X}$. A statistic $T(\boldsymbol{X})$ is sufficient for the parameter θ if and only if there exist a function $g(t; \theta)$ and a function $h(\boldsymbol{x})$ such that $\forall\, \boldsymbol{x}$ and $\forall\, \theta$, the decomposition holds:*

$$f(\boldsymbol{x}; \theta) = g(T(\boldsymbol{x}); \theta)h(\boldsymbol{x}).$$

Theorem 2.2 *Let $X_1, .., X_n$ be a sample of i.i.d. r.v. such that $X_i \sim f(x; \boldsymbol{\theta})$. Let $f(x; \boldsymbol{\theta})$ belong to the exponential family, namely:*

$$f(x; \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right\} \qquad \boldsymbol{\theta} \in \mathbb{R}^d, \quad d \leq k.$$

Then $T(\boldsymbol{X}) = (\sum_{j=1}^{n} t_1(\boldsymbol{X}_j), \ldots, \sum_{j=1}^{n} t_k(\boldsymbol{X}_j))$ is a sufficient statistic for $\boldsymbol{\theta}$.

Definition 2.3 (Minimal Sufficient Statistics) A sufficient statistic $T(\boldsymbol{X})$ is said to be minimal for the parameter θ if for any sufficient statistic $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_2

Theorem 2.3 (L-S) *Let $f(\boldsymbol{x};\theta)$ be the joint density of the sample $\boldsymbol{X}$. Suppose there exists a function $T(\boldsymbol{X})$ such that for every pair of sample realisations $\boldsymbol{x}$ and $\mathbf{y}$ it holds:*

$$\frac{f(\boldsymbol{x};\theta)}{f(\boldsymbol{x};\theta)} \quad \text{independent of } \theta \quad \Leftrightarrow \quad T(\boldsymbol{x}) = T(\mathbf{y}).$$

Then $T(\boldsymbol{X})$ is a minimal sufficient statistic for θ.

Theorem 2.4 *Let $X_1, .., X_n$ be a sample of i.i.d. r.v. such that $X_i \sim f(x;\boldsymbol{\theta})$. Let $f(x;\boldsymbol{\theta})$ belong to the exponential family, namely:*

$$f(x;\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right\} \qquad \boldsymbol{\theta} \in \mathbb{R}^d, \quad d \le k.$$

If the image of $(w_1(\boldsymbol{\theta}), w_2(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta}))$ contains at least an open set of $\mathbb{R}^k$, then the statistic $T(\boldsymbol{X}) = (\sum_{j=1}^n t_1(\boldsymbol{X}_j), \ldots, \sum_{j=1}^n t_k(\boldsymbol{X}_j))$ is a sufficient and complete statistic for $\boldsymbol{\theta}$.

Definition 2.4 (Complete Statistics) Let $f(t;\theta)$ be a family of distributions for the statistic $T(\boldsymbol{X})$. This family of distributions is said to be complete if $\forall$ measurable g it holds:

$$\mathbb{E}_\theta[g(T)] = 0 \quad \forall\theta \quad \Rightarrow \quad \mathbb{P}_\theta\{g(T) = 0\} = 1 \quad \forall\theta.$$

Equivalently, the statistic $T(\boldsymbol{X})$ is said to be complete.

Theorem 2.5 *If a statistic $T(\boldsymbol{X})$ is sufficient and complete for θ then it is also minimal.*

For further study, refer to Chapter 6 [3].

2.2 Exercises

Exercise 2.1 Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$, with $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$. Let

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \qquad \text{and} \qquad S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

be the sample mean and variance. Prove that:

(a) $T(X_1, \ldots, X_n) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is a sufficient, minimal and complete statistic for (μ, σ^2).

(b) $T(X_1, \dots, X_n) = (\overline{X}, S^2)$ is a sufficient, complete and minimal statistic for (μ, σ^2).

Establish the law of the statistic considered at point (b).

Exercise 2.2 Given a random sample $X_1, \dots, X_n$ from a $N(\mu, 1)$, prove that $T = \sum_{i=1}^{n} X_i^2$ is not a sufficient statistic for aμ.

Exercise 2.3 Given a random sample X_1, X_2 from a $\mathcal{P}(\lambda)$, show that $T = X_1 + 2X_2$ is not a sufficient statistic for λ.

Exercise 2.4 Let $X_1, \dots, X_n$ be a random sample from a $U\big([0, \theta]\big)$, where $\theta > 0$. Show that $T = \max\{X_1, \dots, X_n\}$ is a sufficient, minimal and complete statistic for θ.

Exercise 2.5 Given a random sample $X_1, \dots, X_n$ from a $U\big([-\theta/2, \theta/2]\big)$, where $\theta > 0$, show that $T = (\min\{X_1, \dots, X_n\}, \max\{X_1, \dots, X_n\})$ is sufficient for θ. Find a minimal sufficient statistic.

Exercise 2.6 Given a random sample $X_1, \dots, X_n$ of discrete type, determine whether the statistic $T = (X_1, \dots, X_{n-1})$ is sufficient.

Exercise 2.7 Given a random sample drawn from a population with a beta law of parameters α and β, find a sufficient, minimal and complete statistic for (α, β).

Exercise 2.8 Let $X_1, \dots, X_n$ be a random sample from $f(x; \theta) = x \exp\{-x^2/2\theta\}/\theta$, $x > 0$, $\theta > 0$. Show that $\sum_i X_i^2$ is minimal sufficient for θ, but that $\sum_i X_i$ is not sufficient for θ.

Exercise 2.9 Let $X_1, ..., X_n$ be a random sample of real random variables having distribution F_θ continuous with density known except for the value of the parameter $\theta \in \Theta \subseteq \mathbb{R}$. Let $V = V(X_1, ..., X_n)$ be a statistic, $T = T(X_1, ..., X_n)$ a sufficient statistic and $U = U(X_1, ..., X_n)$ a complete statistic. Verify that:

(a) If W is a function of U, then W is a complete statistic.
(b) If T is a function of V, then V is a sufficient statistic.

Exercise 2.10 Let X be a discrete random variable that takes the values 0, 1, 2.

(a) Let $P(X = 0) = p$, $P(X = 1) = 3p$ and $P(X = 2) = 1 - 4p$ with $0 < p < 1/4$. Determine whether the family of distributions of X is complete.
(b) Let $P(X = 0) = p$, $P(X = 1) = p^2$ and $P(X = 2) = 1 - p - p^2$ with $0 < p < 1/2$. Determine whether the family of distributions of X is complete.

2.3 Solutions

2.1

(a) $X \sim N(\mu, \sigma^2)$ belongs to the exponential family, so we study $f_X(x)$.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

We have:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\mu}{2\sigma^2}\right\} = c(\mu, \sigma^2);$$

$$-\frac{1}{2\sigma^2}x^2 = w_1(\sigma^2)t_1(x); \qquad \frac{\mu}{\sigma^2}x = w_2(\mu, \sigma^2)t_2(x).$$

Therefore:

$$\boldsymbol{T}(\boldsymbol{x}) = (T_1(\boldsymbol{x}), T_2(\boldsymbol{x})) = \left(\sum_{j=1}^{n} x_j^2, \sum_{j=1}^{n} x_j\right)$$

is a bivariate sufficient statistic for (μ, σ^2).

Furthermore:

$$(w_1, w_2) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right) : \mathbb{R}^+ \text{ x } \mathbb{R} \to \mathbb{R}^- \text{ x } \mathbb{R}.$$

The image space is an open set of $\mathbb{R}^2$ then $\boldsymbol{T}(\boldsymbol{x})$ is a sufficient, complete and therefore also minimal statistic.

(b) Given that:

$$\overline{X}_n = \frac{\sum X_i}{n} \qquad S^2 = \frac{n}{n-1}\left(\frac{\sum X_i^2}{n} - \left(\frac{\sum X_i}{n}\right)^2\right);$$

$\boldsymbol{T}_1(\boldsymbol{x}) = \left(\overline{X}_n, S^2\right)$ is an invertible function of $\boldsymbol{T}_1(\boldsymbol{x}) = \left(\sum X_i, \sum X_i^2\right)$.
Therefore $\boldsymbol{T}_1(\boldsymbol{x})$ is a sufficient, complete and minimal statistic.

2.2 There are 2 ways to solve this exercise:

Method 1: Lehmann-Scheffé

We can use Lehmann-Scheffé, showing that there are two distinct realisations, **x** and **y**, such that:

$$T(\mathbf{x}) = T(\mathbf{y}).$$

Take for example $\mathbf{x} = (1, 1, \ldots, 1)$ and $\mathbf{y} = (-1, -1, \ldots, -1)$. We immediately notice that $T(\mathbf{x}) = T(\mathbf{y})$. Now evaluate:

$$\frac{f_{\mathbf{X}}(\mathbf{x};\mu)}{f_{\mathbf{X}}(\mathbf{y};\mu)} = \frac{(2\pi)^{-n/2}\exp\{-\sum(x_i-\mu)^2/2\}}{(2\pi)^{-n/2}\exp\{-\sum(y_i-\mu)^2/2\}} =$$
$$= \exp\left\{\sum[-(x_i-\mu)^2 + (y_i-\mu)^2]/2\right\} = \exp\{2\mu n\}.$$

Given that the ratio depends on μ, the statistic T($\mathbf{X}$) is not sufficient for μ.

Method 2: Sufficient and Minimal Statistic

The idea is to find a statistic W that is sufficient and minimal for μ and show that there is no function $g(\cdot;\mu)$ such that $W = g(T(\mathbf{X});\mu)$. This would prove that T($\mathbf{X}$) is not sufficient for μ.

Since the Normal belongs to the exponential family, it is immediately proven that $W = \sum X_i$ is a sufficient and complete (therefore also minimal) statistic for μ.
It is seen that there is no $g(\cdot;\mu)$ such that $\sum X_i = g(\sum X_i^2;\mu)$. Therefore $T(\mathbf{X})$ is not sufficient for μ.

2.3 Consider for example:

$$\mathbb{P}\{X_1 = 0, X_2 = 1 | T = 2\} = \frac{\mathbb{P}\{X_1 = 0, X_2 = 1\}}{T = 2} =$$
$$= \frac{\lambda e^{-2\lambda}}{\mathbb{P}\{X_1 = 0, X_2 = 1\} + \mathbb{P}\{X_1 = 2, X_2 = 0\}} =$$
$$= \frac{\lambda e^{-2\lambda}}{\lambda e^{-2\lambda} + \lambda^2 e^{-2\lambda}/2} =$$
$$= \frac{1}{1 + \frac{\lambda}{2}}.$$

Given that the law of $(X_1, X_2)|T$ depends on λ, T is not a sufficient statistic for λ.

2.4 $X \sim U[0, \theta]$ does not belong to the exponential family because the Spt(X) depends on θ.

$$f(\boldsymbol{x};\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{I}_{[0,\theta]}(x_i) = \frac{1}{\theta}\mathbb{I}_{[0,+\infty]}(x_{(1)})\mathbb{I}_{[0,\theta]}(x_{(n)}).$$

Therefore, by the factorisation criterion, $X_{(n)}$ is sufficient for θ.
Moreover, let g be such that $\mathbb{E}[g(T(X))] = 0$, $\forall\theta > 0$. Given that:

$$F_T(t) = \left(\frac{t}{\theta}\right)^n \mathbb{I}_{[0,\theta]}(t) + \mathbb{I}_{[\theta,+\infty]}(t);$$

then:

$$f_T(t) = \frac{n}{\theta^n} t^{n-1} \mathbb{I}_{[0,\theta]}(t).$$

We verify that T is a complete statistic by exploiting the definition.

$$\mathbb{E}[g(T)] = \int_0^\theta g(t) \frac{n}{\theta^n} t^{n-1} \, \mathrm{d}t = 0 \quad \forall\theta.$$

Deriving with respect to θ we obtain:

$$0 = \frac{1}{\theta^n} g(\theta)\theta^{n-1} + \underbrace{\left(\frac{\mathrm{d}\theta^{-n}}{\mathrm{d}\theta}\right) \int_0^\theta g(t) \frac{n}{\theta^n} t^{n-1} \, \mathrm{d}t}_{=\mathbb{E}[g(T)]=0} \quad \forall\theta.$$

Therefore, we can conclude that: $g(\theta) = 0 \, \forall\theta$. T is a complete and also minimal sufficient statistic for θ.

2.5

$$f(\boldsymbol{x}; \theta) = \frac{1}{\theta^n} \mathbb{I}_{\left[-\frac{\theta}{2}, \frac{\theta}{2}\right]}(X_{(1)}) \mathbb{I}_{\left[-\frac{\theta}{2}, \frac{\theta}{2}\right]}(X_{(n)}).$$

By the factorisation criterion $T = \left(X_{(1)}, X_{(n)}\right)$ is sufficient for θ.

We use Lehmann-Scheffé to find a minimal sufficient statistic:

$$\frac{f(\boldsymbol{x}; \theta)}{f(\boldsymbol{y}; \theta)} = \frac{\mathbb{I}_{\{2 \cdot \max\{|x_{(1)}|, |x_{(n)}|\}, +\infty\}}(\theta)}{\mathbb{I}_{\{2 \cdot \max\{|y_{(1)}|, |y_{(n)}|\}, +\infty\}}(\theta)}.$$

$$\Updownarrow$$

$$2 \cdot \max\left\{\left|x_{(1)}\right|, \left|x_{(n)}\right|\right\} = 2 \cdot \max\left\{\left|y_{(1)}\right|, \left|y_{(n)}\right|\right\}.$$

Therefore we can conclude that $2 \cdot \max\left\{\left|X_{(1)}\right|, \left|X_{(n)}\right|\right\}$ is a minimal sufficient statistic for θ.

2.6 It can be proven that the statistic $T = (X_1, \ldots, X_{n-1})$ is not sufficient using the definition:

$$\mathbb{P}\{\mathbf{X} = \mathbf{k} | T = t\} = \mathbb{P}\{X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k_n | X_1 = k_1, \ldots, X_{n-1} = k_{n-1}\} = \\ = \mathbb{P}\{X_n = k_n\}.$$

Indeed $\mathbb{P}\{X_n = k_n\}$ depends on the parameter of the distribution.

2.7 The distribution $Beta(\alpha, \beta)$ belongs to the exponential family.

$$f(x; \alpha, \beta) = \frac{1}{Beta(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{I}_{[0,1]}(x) =$$

$$= \frac{1}{Beta(\alpha, \beta)} \exp\left\{ \underbrace{(\alpha-1)}_{w_1(\alpha)} \underbrace{\log(x)}_{t_1(x)} + \underbrace{(\beta-1)}_{w_2(\beta)} \underbrace{\log(1-x)}_{t_2(x)} \right\} \mathbb{I}_{[0,1]}(x).$$

$$(w_1, w_2) : \mathbb{R}^+ \text{ x } \mathbb{R} \to [-1, +\infty] \text{ x } [-1, +\infty].$$

$[-1, +\infty] \text{ x } [-1, +\infty]$ contains an open set of $\mathbb{R}^2$.

This implies that the statistic:

$$\left(\sum log(X_i), \sum log(1 - X_i)\right) = \left(log\left(\prod X_i\right), log\left(\prod(1 - X_i)\right)\right)$$

is minimal and complete.

2.8 To show that $\sum_i X_i^2$ is minimal sufficient for θ, we just need to observe that this density belongs to the exponential family.

$$f_{\mathbf{X}}(\mathbf{X}; \theta) = \prod_{i=1}^{n} \frac{x_i \exp\{-x_i^2/2\theta\}}{\theta} = \frac{(\prod x_i) \exp\{-\sum x_i^2/2\theta\}}{\theta^n}.$$

We immediately recognise that $W(\mathbf{X}) = \sum x_i^2$ is a sufficient statistic for θ. Moreover, since $-1/2\theta : \mathbb{R} \to \mathbb{R}^-$ and $\mathbb{R}^-$ contains an open set of $\mathbb{R}$, then $W(\mathbf{X})$ is a complete sufficient statistic for θ (thus it is also minimal).

There are 2 ways to show that $T(\mathbf{X}) = \sum x_i$ is not sufficient for θ:

Method 1: Lehmann-Scheffé

We can use Lehmann-Scheffé, showing that there are two distinct realisations, **x** and **y**, such that:

$$T(\mathbf{x}) = T(\mathbf{y}) \not\Rightarrow \frac{f_{\mathbf{X}}(\mathbf{x}; \mu)}{f_{\mathbf{X}}(\mathbf{y}; \mu)} \text{ does not depend on } \theta.$$

Take for example $\mathbf{x} = (1, 1, \ldots, 1)$ and $\mathbf{y} = (1/n, 1/n, \ldots, 1/n, n - 1 + 1/n)$. We immediately notice that $T(\mathbf{x}) = T(\mathbf{y})$. Now evaluate:

$$\frac{f_{\mathbf{X}}(\mathbf{x}; \mu)}{f_{\mathbf{X}}(\mathbf{y}; \mu)} = \frac{\frac{(\prod x_i) \exp\{-\sum x_i^2/2\theta\}}{\theta^n}}{\frac{(\prod y_i) \exp\{-\sum y_i^2/2\theta\}}{\theta^n}} =$$

$$= \frac{\prod x_i}{\prod y_i} \cdot \exp\left\{ \sum \frac{-x_i^2 + y_i^2}{2\theta} \right\} =$$

$$= \frac{\prod x_i}{\prod y_i} \cdot \exp\left\{\sum \frac{n^3 - 2n^2 + 2n - 1}{2\theta}\right\}.$$

Since the ratio depends on θ, the statistic T(**X**) is not sufficient for θ.

Method 2: Minimal Sufficient Statistic
It is clear that there is no $g(\cdot;\theta)$ such that $\sum X_i^2 = g(\sum X_i;\theta)$. Therefore $T(\mathbf{X})$ is not sufficient for θ.

2.9

(a) W is a function of U, U is complete.

$$g \quad \text{s.t.} \quad \mathbb{E}[g(W)] = 0 \quad \forall\theta.$$

$$\Downarrow$$

$$\mathbb{E}[g(h(U))] = 0 \quad \forall\theta.$$

$$\Downarrow$$

$$\mathbb{P}\{g(h(U)) = g(W) = 0\} = 1.$$

$$\Downarrow$$

$$W \text{ is complete.}$$

(b)

$$f(\boldsymbol{x};\theta) = g(T(\boldsymbol{x};\theta)) \cdot h(\boldsymbol{x}) = g(l(V(;\theta))) \cdot h(\boldsymbol{x}) = r(V(\boldsymbol{x};\theta)) \cdot h(\boldsymbol{x}).$$

Therefore V is a sufficient statistic.

2.10

(a) $0 < p < \frac{1}{4}$, see Table 2.1.
If $\mathbb{E}[g(X)] = 0$ it means that:

$$p \cdot g(0) + 3p \cdot g(1) + (1 - 4p) \cdot g(2) = 0.$$

Table 2.1 Distribution of X, if $0 < p < \frac{1}{4}$

x	$f(x)$
0	p
1	$3p$
2	$1 - 4p$

Table 2.2 Distribution of X, if $0 < p < \frac{1}{2}$

x	$f(x)$
0	p
1	p^2
2	$1 - p - p^2$

Simply choose $g(0) = -3g(1)$ and $g(2) = 0$ because $\mathbb{E}[g(X)] = 0$ but $\mathbb{P}\{g(X) = 0\} \neq 0$. Therefore, it is not complete.

(b) $0 < p < \frac{1}{2}$, see Table 2.2.
Similarly, we have:

$$
\begin{aligned}
0 &= g(0)p + g(1)p^2 + g(2)(1 - p - p^2) \\
&= (g(1) - g(2))p^2 + (g(0) - g(2))p + g(2) \quad \forall p \in [0, 1/2].
\end{aligned}
$$

$$\Downarrow$$

$g(2) = 0$ and therefore $g(0) = g(1) = 0$ being the coefficients of degree 2 in p.

Then X is a complete statistic.

Chapter 3
Point Estimators

3.1 Theory Recap

Definition 3.1 (Point Estimators) A point estimator is any function $W(X_1, \ldots, X_n)$ of the sample $X_1, \ldots, X_n$. Every statistic is therefore a point estimator.

Definition 3.2 (Method of Moments) Let $X_1, \ldots, X_n$ be a sample of random variables with probability density $f(\boldsymbol{x}; \theta_1, \ldots, \theta_k)$. The estimators obtained with the method of moments can be derived from a system of k equations in which the first k moments of the sample $(m_1, \ldots, m_k)$ are equated with the first k moments of the population $(\mu_1, \ldots, \mu_k)$. Therefore, the following system must be solved with respect to θ:

$$\begin{cases} m_1 = \mu_1; & m_1 := \frac{1}{n}\sum_{i=1}^n X_i; \qquad \mu_1(\theta) := \mathbb{E}[X]; \\ m_2 = \mu_2; & m_2 := \frac{1}{n}\sum_{i=1}^n X_i^2; \qquad \mu_2(\theta) := \mathbb{E}[X^2]; \\ \vdots & \\ m_k = \mu_k; & m_k := \frac{1}{n}\sum_{i=1}^n X_i^k; \qquad \mu_k(\theta) := \mathbb{E}[X^k]. \end{cases}$$

Definition 3.3 (MLE) Let $X_1, \ldots, X_n$ be a sample of i.i.d. random variables with probability density $f_{X_i}(x; \theta_1, \ldots, \theta_k)$. The likelihood function is defined as follows:

$$L(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x; \theta_1, \ldots, \theta_k);$$

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_3

that is, it is the density of the sample seen as a function of θ and considering the sample realization as known. For a given sample realization $\boldsymbol{x}$, the maximum likelihood estimator, or MLE, is defined as follows:

$$\hat{\theta} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argsup}}\, L(\boldsymbol{\theta}; \boldsymbol{x});$$

where Θ is the parameter space. A method to find the MLE consists in studying the derivative of the log-likelihood, that is, the derivative of the logarithm of the likelihood.

Theorem 3.1 (Invariance Principle) *If $\hat{\theta}$ is MLE for θ, then $\tau(\hat{\theta})$ is MLE for $\tau(\theta)$, whatever $\tau(\cdot)$ is.*

Definition 3.4 (MSE) The mean squared error, or MSE, of an estimator T for the parameter θ is:

$$\mathbb{E}_\theta[(T-\theta)^2].$$

Definition 3.5 (Bias) The bias of an estimator T for a parameter θ is the difference between the expected value of T and the parameter θ.

$$Bias_\theta(T) = \mathbb{E}_\theta[T] - \theta.$$

An estimator is defined as unbiased if the bias is zero, that is, $\mathbb{E}_\theta[T] = \theta$.

We observe that the MSE can be expressed as follows:

$$\mathbb{E}_\theta[(T-\theta)^2] = Var_\theta(T) + (\mathbb{E}_\theta[T]-\theta)^2 = Var_\theta(T) + (Bias_\theta(T))^2.$$

3.2 Exercises

Exercise 3.1 Let $X_1, \dots, X_n$ be a random sample from a uniform law on the interval $[0, \theta]$, $\theta > 0$.

(a) Determine an estimator of θ using the method of moments.
(b) Is the found estimator unbiased?
(c) Is it sufficient?

Exercise 3.2 Let $X_1, \dots, X_n$ be a random sample from a uniform law on the interval $[a, b]$. Estimate a and b using the method of moments.

Exercise 3.3 Let $X_1, \dots, X_n$ be a sample of size n of independent random variables with density

$$f_X(x;\theta) = \theta x^{\theta-1} 1_{(0,1)}(x); \qquad \theta > 0.$$

(a) Calculate the maximum likelihood estimator $\hat{\theta}_n$ of θ.
(b) Determine and recognise the laws of $-\log X_k$ and of $-\sum_{k=1}^{n} \log X_k$.
(c) Is $\hat{\theta}_n$ biased?
(d) Calculate the mean squared error of $\hat{\theta}_n$.

Exercise 3.4 Let $X_1, \ldots, X_n$ be a random sample from $U\big([\theta - \frac{1}{2}, \theta + \frac{1}{2}]\big)$, $\theta \in \mathbb{R}$. Calculate the maximum likelihood estimator $\hat{\theta}_n$ of θ.

Exercise 3.5 Let $X_1, \ldots, X_n$ be a family of independent random variables all distributed according to an exponential law of parameter λ. Each X_i represents the instant of disintegration of a nucleus of a certain radioactive element. For each fixed $t \geq 0$ let:

- Y_i be the random variable that is 1 if the i-th nucleus is still alive at time t and 0 otherwise.
- V_n be the proportion of nuclei still alive at time t, of the n present at time 0.

(a) Find the law of Y_i and that of V_n.
(b) Verify if the Law of Large Numbers can be applied to V_n and state whether, and in what sense, the sequence V_n converges for $n \to \infty$ to a constant v. In this case, determine the constant v and express it in terms of the average lifetime τ of the generic radioactive nucleus.
(c) Assuming to observe the sample $Y_1, \ldots, Y_n$, propose an estimator of τ based on this sample.
(d) Assuming instead to observe the lifetimes $X_1, \ldots, X_n$, propose an estimator of τ based on this sample.

Exercise 3.6 Let X be a discrete random variable that can only take the values -2, 0, 2, respectively with probabilities:

$$p(-2) = \frac{1}{2} - \theta; \qquad p(0) = 2\theta; \qquad p(2) = \frac{1}{2} - \theta.$$

(a) For which θ is the function p a density?
(b) Let $X_1, \ldots, X_n$ be a sample of independent random variables with the same density p. Determine the expression of the function f such that the likelihood function is written as:

$$(2\theta)^{n - f(x_1, \ldots, x_n)} \left(\frac{1}{2} - \theta\right)^{f(x_1, \ldots, x_n)}.$$

(c) Calculate the maximum likelihood estimator for θ. Is it unbiased? Is it consistent?

Exercise 3.7 Let $X_1, \dots, X_{3n}$ be a sample of $3n$ independent r.v. of which:

$$X_1, \dots, X_{2n} \text{ of law } \mathcal{P}(\lambda);$$

$$X_{2n+1}, \dots, X_{3n} \text{ of law } \mathcal{P}(2\lambda);$$

where λ is an unknown parameter.

Determine the maximum likelihood estimator for λ and calculate its variance. Is it biased?

Exercise 3.8 Let $X_1, \dots, X_n$ be a random sample from a normal law $N(\mu, \sigma^2)$. Show that the maximum likelihood estimators of μ and σ^2 are:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n \,; \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Exercise 3.9 Consider the two independent random samples, $X_1, \dots, X_n$ from a population $N(\mu_1, \sigma^2)$ and $Y_1, \dots, Y_m$ from a population $N(\mu_2, \sigma^2)$, where the parameters μ_1, μ_2 and σ^2 are all unknown. Calculate the maximum likelihood estimator for $\theta = (\mu_1, \mu_2, \sigma^2)$.

Exercise 3.10 For $\theta \in [0, 1]$, let $f_X(x; \theta) = (\frac{\theta}{2})^{|x|}(1-\theta)^{1-|x|} I_{\{-1,0,1\}}(x)$, $x \in \mathbb{R}$, be the density of a random variable X.

(a) Is X a sufficient statistic? Is it a complete statistic?
(b) Is $|X|$ a sufficient statistic? Is it a complete statistic?
(c) Is either X or $|X|$ a minimal sufficient statistic?
Now consider the maximum likelihood estimator $T_1(X)$ and the estimator $T_2(X) = 2I_{\{1\}}(X)$.
(d) Calculate bias and mean squared error of T_1 and of T_2.
(e) Which estimator would you prefer between T_1 and T_2?

Exercise 3.11 Consider a discrete variable X described by the statistical model $f(x; \theta)$ where $\theta = 0, 1, 2$ (see Table 3.1).

(a) Determine the maximum likelihood estimator $\widehat{\theta}_1(X_1)$ of θ based on a single observation X_1.
(b) Calculate bias and mean squared error of $\widehat{\theta}_1$.
(c) Determine the maximum likelihood estimator $\widehat{\theta}_2(X_1, X_2)$ of θ based on the sample X_1, X_2.

Table 3.1 Density of X varying the parameter θ

x	0	1	2
$f(x; 0)$	1/2	1/2	0
$f(x; 1)$	1/3	1/3	1/3
$f(x; 2)$	1/4	1/4	1/2

(d) Calculate bias and mean squared error of $\widehat{\theta_2}$.
(e) Given the sample $x_1 = 0, x_2 = 2$, how would you estimate θ?

Exercise 3.12 The coefficient of variation

$$CV = \frac{\sigma}{|\mu|}$$

is an index introduced by Karl Pearson to study the relative variability of a distribution.

Suppose we have observed a sample of size n from a Normal distribution with unknown mean and variance and we only know the number T_n of sample observations that are greater than zero.

(a) Based solely on the information T_n, is it possible to provide an estimate of the coefficient of variation of the distribution?
(b) If $n = 1000$, for which values of T_{1000} will we estimate that the standard deviation of the distribution is less than 1/3 of the mean?

Exercise 3.13 Let $X_1, \ldots, X_n$ be the results of n measurements, independent and affected by random error, of the same unknown quantity μ, for which:

$$X_i = \mu + \epsilon_i \, ; \qquad\qquad \epsilon_1, \ldots, \epsilon_n \text{ i.i.d.}$$

In the case of error $\epsilon \sim N(0, \sigma^2)$:

(a) Determine the law of X_i.
(b) Show that the sample mean $\overline{X}_n$ is the maximum likelihood estimator for μ.

In the case of error $\epsilon \sim f(s; \sigma^2) = \frac{\mathrm{e}^{-2|s|/\sigma^2}}{\sigma^2}$:

(c) Determine the law of X_i.
(d) Show that the median $m(X_1, \ldots, X_n)$ is a maximum likelihood estimator for μ.

Exercise 3.14 Let $X_1, \ldots, X_n$ be a random sample from:

$$f(x; \theta) = \frac{\theta}{x^2} I_{[\theta, +\infty)}(x) \, ; \qquad \theta > 0.$$

(a) Prove that there is no moment estimator for θ.
(b) Determine the maximum likelihood estimator $\hat{\theta}$ for θ.
(c) Show that $\hat{\theta}$ is a minimal sufficient statistic.

Exercise 3.15 Let $X_1, \ldots, X_n$ be a random sample from a population with a uniform law on the interval $[\theta, 2\theta]$, where $\theta > 0$.

(a) Determine the moment estimator of θ.
(b) Determine the maximum likelihood estimator of θ.
(c) Determine a minimal sufficient statistic for θ.

Exercise 3.16 Consider the statistical model:

$$f_X(x;\alpha,\beta)=\frac{1}{\beta}\,\mathrm{e}^{-(x-\alpha)/\beta}\,I_{[\alpha,+\infty)}(x),\qquad \alpha\geq 0,\quad \beta>0.$$

(a) Calculate the mean $\mu(\alpha,\beta)$ and the variance $\sigma^2(\alpha,\beta)$ of a random variable with density $f_X(x;\alpha,\beta)$.
Let now $X_1,\ldots,X_n$ be a random sample from the density $f(x;\alpha,\beta)$.
(b) Calculate the maximum likelihood estimator $(\widehat{\alpha}_n,\widehat{\beta}_n)$ of (α,β) based on $X_1,\ldots,X_n$.
(c) Is the statistic $(\widehat{\alpha}_n,\widehat{\beta}_n)$ sufficient for (α,β)?
(d) Calculate the maximum likelihood estimator $\widehat{\mu}_n$ of μ based on $X_1,\ldots,X_n$.
(e) What is the mean square error of $\widehat{\mu}_n$?

3.3 Solutions

3.1

(a) Given that $\mathbb{E}[X_i]=\theta/2$, the estimator of θ calculated with the method of moments is:

$$\hat{\theta}_{MOM}=2\overline{X}_n.$$

(b) The estimator is unbiased, in fact:

$$\mathbb{E}[\hat{\theta}_{MOM}]=2\mathbb{E}[\overline{X}_n]=2\cdot\frac{\theta}{2}=\theta.$$

(c) $X_{(n)}$ is a minimal sufficient statistic for θ and since there is no function r such that $X_{(n)}=r(2\overline{X}_n)$, then $\hat{\theta}_{MOM}$ is not sufficient.

3.2

$$\mathbb{E}[X_i]=\frac{a+b}{2};$$

$$Var(X)=\frac{(b-a)^2}{12}\quad\Rightarrow\quad \mathbb{E}[X^2]=\frac{a^2+b^2-2ab+3a^2+3b^2+6ab}{12}$$
$$=\frac{1}{3}(a^2+b^2+ab);$$

therefore:

$$\begin{cases} a & = 2\mathbb{E}[X] - b; \\ 3\mathbb{E}[X^2] & = 4(\mathbb{E}[X])^2 + b^2 - 4b\mathbb{E}[X] + b^2 + 2b\mathbb{E}[X] - b^2; \end{cases}$$

from which we obtain:

$$b^2 - 2b\mathbb{E}[X] + (4(\mathbb{E}[X])^2 - 3\mathbb{E}[X^2]) = 0.$$

$$a = \mathbb{E}[X] - \sqrt{3\mathbb{E}[X^2] - 3(\mathbb{E}[X])^2}; \qquad b = \mathbb{E}[X] + \sqrt{3\mathbb{E}[X^2] - 3(\mathbb{E}[X])^2}.$$

From which we conclude that:

$$a = \overline{X}_n - \sqrt{\frac{3\sum_i (X_i - \overline{X}_n)^2}{n}}; \qquad b = \overline{X}_n + \sqrt{\frac{3\sum_i (X_i - \overline{X}_n)^2}{n}}.$$

The choice of signs is dictated by the fact that $a \leq b$.

3.3

(a) We calculate the likelihood and the log-likelihood:

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} \theta x_i^{\theta-1} \mathbb{I}_{[0,1]}(x_i) = \theta^n \, (x_i)^{\theta-1} \prod_{[0,1]} (x_i).$$

$$l(\theta; \boldsymbol{x}) = n \log(\theta) + (\theta - 1) \sum_{i=1}^{n} \log(x_i).$$

We derive the log-likelihood:

$$\frac{\partial l(\theta; \boldsymbol{x})}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log(x_i).$$

Therefore:

$$\frac{\partial l(\theta; \boldsymbol{x})}{\partial \theta} \geq 0 \quad \Longleftrightarrow \quad \frac{n}{\theta} \leq -\sum_{i=1}^{n} \log(x_i).$$

$$\hat{\theta}_{MLE} = -\frac{n}{\sum \log(x_i)}.$$

(b)

$$Y = -\log(X_i) \quad \Rightarrow \quad F_Y(t) = \mathbb{P}\{-\log(X_i) \le t\} = \mathbb{P}\{\log(X_i) \ge -t\}$$
$$= \mathbb{P}\{X \ge e^{-t}\}.$$

$$F_X(t) = \int_0^t \theta x^{\theta-1}\,\mathrm{d}x = t^{\theta}\mathbb{I}_{(0,1)}(t) + \mathbb{I}_{[1,+\infty)}(t).$$

$$F_Y(t) = 1 - e^{-t\theta} \qquad \Rightarrow \qquad -\log(X_i) \sim \mathcal{E}(\theta); \quad -\sum_{k=1}^{n}\log(X_k) \sim \Gamma(n,\theta).$$

(c)

$$\mathbb{E}[\hat{\theta}_{MLE}] = \mathbb{E}\left[-\frac{n}{\sum\log(X_i)}\right] = \frac{n\theta}{n-1}.$$

where the last equality is due to the fact that if $Y \sim \Gamma(n,\theta)$, then $\mathbb{E}\left[\frac{1}{Y}\right] = \frac{\theta}{n-1}$. Therefore $\hat{\theta}_{MLE}$ is biased.

(d)

$$MSE(\hat{\theta}_n) = Var(\hat{\theta}_n) + (bias)^2.$$
$$\mathbb{E}[\hat{\theta}_n^2] = \frac{n^2\theta^2}{(n-1)(n-2)}.$$
$$Var(\hat{\theta}_n) = \frac{n^2\theta^2}{(n-1)^2(n-2)}.$$
$$MSE(\hat{\theta}_n) = \frac{\theta^2(n+2)}{(n-1)(n-2)}.$$

3.4

$$L(\theta;\boldsymbol{x}) = \prod_{1}^{n}\mathbb{I}_{[\theta-1/2,\theta+1/2]}(x_i) = \mathbb{I}_{[X_{(1)}-1/2,X_{(n)}+1/2]}(\theta).$$

We can equally choose $\hat{\theta}_{MLE} = X_{(1)} - 1/2$ or $\hat{\theta}_{MLE} = X_{(n)} + 1/2$, or any point of the form $\hat{\theta}_{MLE} = \alpha(X_{(1)} - 1/2) + (1-\alpha)(X_{(n)} + 1/2)$, with $\alpha \in [0, 1]$.

3.5

(a)

$$Y_i = \mathbb{I}_{[X_i \ge t]} \qquad \Rightarrow \qquad Y_i \sim Be(e^{-\lambda t}).$$

V_n as the proportion of present nuclei is $\frac{\sum_{i=1}^{n} Y_i}{n}$, therefore $nV_n \sim Bi(n, e^{-\lambda t})$.

$$\mathbb{P}\{V_n = k/n\} = \binom{n}{k}(e^{-\lambda t})^k(1 - e^{-\lambda t})^{n-k}.$$

(b) By the Strong Law of Large Numbers:

$$V_n \stackrel{q.c.}{\to} \mathbb{E}[Y_i] = e^{-\lambda t} = e^{-t/\tau};$$

where $\tau = 1/\lambda$.

(c) $\bar{Y}_n$ is MLE estimator of $p = e^{-t/\tau}$ and also $\tau = -\frac{t}{\log p}$. By the principle of invariance $\widehat{\tau}_{MLE} = -\frac{t}{\log V_n}$.

(d) $\tau = 1/\lambda$ therefore the MLE estimator for the mean of exponentials is $\overline{X}_n$.

3.6

(a) It must be $0 \leq \theta \leq 1/2$. Also note that:

$$p(x) = \left(\frac{1}{2} - \theta\right)^{\frac{1}{2}|x|} \cdot (2\theta)^{1-\frac{1}{2}|x|}.$$

(b)

$$L(\theta; \boldsymbol{x}) = \left(\frac{1}{2} - \theta\right)^{\frac{1}{2}\sum_i |x_i|} \cdot (2\theta)^{1-\frac{1}{2}\sum_i |x_i|}.$$

therefore $f(\boldsymbol{x}) = \frac{1}{2}\sum_i |x_i|$. Note that $\frac{|X_i|}{2} \sim Be(1 - 2\theta)$.

(c) Therefore $(\widehat{1 - 2\theta}) = \frac{\sum |X_i|}{2n}$ and by the principle of invariance:

$$\hat{\theta}_{MLE} = \frac{2n - \sum_i |X_i|}{4n};$$

which is unbiased. It is also consistent by the Strong Law of Large Numbers.

3.7 We calculate the likelihood, the log-likelihood and we calculate the derivative of the latter with respect to the parameter.

$$L(\lambda; \boldsymbol{x}) = \frac{e^{-2n\lambda}\lambda^{\sum_{i=1}^{2n} x_i}}{x_1! \dots x_{2n}!} \cdot \frac{e^{-2n\lambda}(2\lambda)^{\sum_{i=2n}^{3n} x_i}}{x_{2n+1}! \dots x_{3n}!} \quad \propto \quad e^{-4n\lambda}\lambda^{\sum_{i=1}^{3n} x_i}.$$

$$l(\lambda; \boldsymbol{x}) \propto -4n\lambda + \sum_{i=1}^{3n} x_i \log(\lambda).$$

$$\frac{\partial l(\lambda; \boldsymbol{x})}{\partial \lambda} = -4n + \frac{\sum_{i=1}^{3n} x_i}{\lambda}.$$

From which we conclude that:

$$\widehat{\lambda}_{MLE} = \frac{\sum_{i=1}^{3n} X_i}{4n}.$$

$$Var(\widehat{\lambda}_{MLE}) = \frac{1}{16n^2}(2n\lambda + n2\lambda) = \frac{\lambda}{4n}.$$

Furthermore, the MLE estimator turns out to be unbiased, in fact:

$$\mathbb{E}[\widehat{\lambda}_{MLE}] = \frac{1}{4n}[2n\lambda + n2\lambda] = \lambda.$$

3.8 We write the likelihood and the log-likelihood:

$$L(\mu, \sigma; \boldsymbol{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2\}.$$

$$l(\mu, \sigma; \boldsymbol{x}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum (x_i - \mu)^2.$$

We derive the log-likelihood to find the MLEs:

$$\begin{cases} \frac{\partial l}{\partial \mu} & = \frac{1}{2\sigma^2} 2\sum_i (x_i - \mu) = \frac{\sum_i (x_i - \mu)}{\sigma^2}; \\ \frac{\partial l}{\partial \sigma^2} & = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (x_i - \mu)^2. \end{cases}$$

By setting the two equations of the system equal to 0, we obtain:

$$\begin{cases} \widehat{\mu}_{MLE} & = \overline{X}_n; \\ \widehat{\sigma}^2_{MLE} & = \frac{\sum_i (X_i - \overline{X}_n)^2}{n}. \end{cases}$$

It can be shown that the stationary point found is a maximum.

3.9 We calculate the likelihood and the log-likelihood:

$$L(\mu_1, \mu_2, \sigma^2; \boldsymbol{x}, \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{(n+m)/2}} \times \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_i (x_i - \mu_1)^2 + \sum_i (y_i - \mu_2)^2\right)\right\}.$$

$$l(\mu_1, \mu_2, \sigma^2; \boldsymbol{x}, \mathbf{y}) \propto -\frac{n+m}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left[\sum_i (x_i - \mu_1)^2 + \sum_i (y_i - \mu_2)^2\right].$$

By setting $\frac{\partial l}{\partial \mu_1} = \frac{\partial l}{\partial \mu_2} = 0$, we get $\widehat{\mu}_1 = \overline{X}_n$ and $\widehat{\mu}_2 = \overline{Y}_n$.

By setting $\frac{\partial l}{\partial \sigma^2} = 0$, we get:

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n+m}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\left(\sum_i (x_i - \mu_1)^2 + \sum_i (y_i - \mu_2)^2\right) = 0.$$

Therefore:

$$\widehat{\sigma}^2 = \frac{\sum_i (X_i - \widehat{\mu}_1)^2 + \sum_i (Y_i - \widehat{\mu}_2)^2}{n+m}.$$

3.10

(a) X is a sufficient statistic by the factorisation theorem, but it is not a complete statistic, in fact:

$$0 = \mathbb{E}[g(X)] = \frac{\theta}{2} g(-1) + (1-\theta) g(0) + \frac{\theta}{2} g(1) =$$
$$= g(0) + \theta \left(\frac{g(-1)}{2} + \frac{g(1)}{2} - g(0)\right).$$

Choosing $g(0) = 0$ and $g(-1) = g(1)$, we have $\mathbb{E}[g(X)] = 0$, therefore X is not a complete statistic for θ.

(b) $|X|$ is a sufficient statistic by the factorisation theorem, and it is also a complete statistic, in fact:

$$0 = \mathbb{E}[g(|X|)] = (1-\theta) g(0) + \theta g(1) =$$
$$= g(0) + \theta\,(g(1) - g(0)).$$

$\mathbb{E}[g(|X|)] = 0 \iff g(0) = g(1) = 0$, that is $\mathbb{P}(g(|X|) = 0) = 1$.

(c) $|X|$ is minimal sufficient, while X is not minimal sufficient because, if it were, it should be a function of $|X|$.

(d) We write the likelihood and the log-likelihood to find the MLE:

$$L(\theta; x) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|} =$$
$$= \begin{cases} L(\theta; 0) & = 1 - \theta \quad \Rightarrow \quad \hat{\theta} = 0; \\ L(\theta; \pm 1) & = \frac{\theta}{2} \quad \Rightarrow \quad \hat{\theta} = 1. \end{cases}$$

So $T_1(X) = |X|$.

Let's calculate the bias of the two estimators:

$$\mathbb{E}[T_1(X)] = \theta.$$
$$\mathbb{E}[T_2(X)] = 2\mathbb{P}\{X = 1\} = \theta.$$

Both are unbiased estimators for θ.
To calculate the MSE of the two estimators, it is therefore sufficient to calculate their variance:

$$Var(T_1(X)) = \theta(1-\theta).$$
$$Var(T_2(X)) = 4\mathbb{P}\{X = 1\} - \theta^2 = \theta(2-\theta).$$

(e) Among two unbiased estimators, the one with the lower variance is preferred, so T_1.

3.11

(a) We note that: $X_1 \in \{0, 1, 2\}$ and $\theta \in \{0, 1, 2\}$. We apply the definition of MLE:

$$\widehat{\theta}_1(X_1) = \underset{\theta\in\{0,1,2\}}{\text{argsup}}\ L(\theta; x_1) = \begin{cases} \text{argsup}_{\theta\in\{0,1,2\}}\{1/2, 1/3, 1/4\} & \text{if} \quad x_1 = 0; \\ \text{argsup}_{\theta\in\{0,1,2\}}\{1/2, 1/3, 1/4\} & \text{if} \quad x_1 = 1; \\ \text{argsup}_{\theta\in\{0,1,2\}}\{0, 1/3, 1/2\} & \text{if} \quad x_1 = 2. \end{cases}$$

So we get:

$$\widehat{\theta}_1(X_1) = \begin{cases} 0 & \text{if} \quad x_1 = 0; \\ 0 & \text{if} \quad x_1 = 1; \\ 2 & \text{if} \quad x_1 = 2. \end{cases}$$

(b) Calculate bias and mean squared error of $\widehat{\theta}_1$.

$$\mathbb{E}[\widehat{\theta}_1] = 2 \cdot \mathbb{P}\{X_1 = 2\} = \begin{cases} 2 \cdot 0 = 0 & \text{if} \quad \theta = 0; \\ 2 \cdot 1/3 = 2/3 & \text{if} \quad \theta = 1; \\ 2 \cdot 1/2 = 1 & \text{if} \quad \theta = 2. \end{cases}$$

$$\text{Bias}(\widehat{\theta}_1; \theta) = \begin{cases} 0 - 0 = 0 & \text{if} \quad \theta = 0; \\ 1 - 2/3 = 1/3 & \text{if} \quad \theta = 1; \\ 2 - 1 = 1 & \text{if} \quad \theta = 2. \end{cases}$$

$$\mathbb{E}[\widehat{\theta}_1^2] = 4 \cdot \mathbb{P}\{X_1 = 2\} = \begin{cases} 0 & \text{if} \quad \theta = 0; \\ 4/3 & \text{if} \quad \theta = 1; \\ 2 & \text{if} \quad \theta = 2. \end{cases}$$

$$Var(\widehat{\theta}_1) = \mathbb{E}[\widehat{\theta}_1^2] - (\mathbb{E}[\widehat{\theta}_1])^2 = \begin{cases} 0 - 0 = 0 & \text{if} \quad \theta = 0; \\ 4/3 - 4/9 = 8/9 & \text{if} \quad \theta = 1; \\ 2 - 1 = 1 & \text{if} \quad \theta = 2. \end{cases}$$

$$MSE[\widehat{\theta}_1; \theta] = Var(\widehat{\theta}_1) + (\text{Bias}(\widehat{\theta}_1; \theta))^2 = \begin{cases} 0 & \text{if} \quad \theta = 0; \\ 1 & \text{if} \quad \theta = 1; \\ 2 & \text{if} \quad \theta = 2. \end{cases}$$

(c) We construct Table 3.2:
We apply the definition of MLE:

$$\widehat{\theta}_2(X_1, X_2) = \underset{\theta \in \{0,1,2\}}{\text{argsup}} \; L(\theta; (x_1, x_2)) =$$

$$= \begin{cases} \text{argsup}_{\theta \in \{0,1,2\}}\{1/4, 1/9, 1/16\} & \text{if} \quad (x_1; x_2) = (0; 0); \\ \text{argsup}_{\theta \in \{0,1,2\}}\{1/2, 2/9, 1/8\} & \text{if} \quad (x_1; x_2) = (0; 1); \\ \text{argsup}_{\theta \in \{0,1,2\}}\{0, 2/9, 1/4\} & \text{if} \quad (x_1; x_2) = (0; 2); \\ \text{argsup}_{\theta \in \{0,1,2\}}\{1/4, 1/9, 1/16\} & \text{if} \quad (x_1; x_2) = (1; 1); \\ \text{argsup}_{\theta \in \{0,1,2\}}\{0, 2/9, 1/4\} & \text{if} \quad (x_1; x_2) = (1; 2); \\ \text{argsup}_{\theta \in \{0,1,2\}}\{0, 1/9, 1/4\} & \text{if} \quad (x_1; x_2) = (2; 2). \end{cases}$$

$$\widehat{\theta}_2(X_1, X_2) = \begin{cases} 0 & \text{if} \quad (x_1; x_2) = (0; 0); \\ 0 & \text{if} \quad (x_1; x_2) = (0; 1); \\ 2 & \text{if} \quad (x_1; x_2) = (0; 2); \\ 0 & \text{if} \quad (x_1; x_2) = (1; 1); \\ 2 & \text{if} \quad (x_1; x_2) = (1; 2); \\ 2 & \text{if} \quad (x_1; x_2) = (2; 2). \end{cases} =$$

Table 3.2 Joint density of X_1 and X_2 as the parameter varies

$(x_1; x_2)$	(0;0)	(0;1)	(0;2)	(1;1)	(1;2)	(2;2)
$f(x_1, x_2; 0)$	1/4	1/2	0	1/4	0	0
$f(x_1, x_2; 1)$	1/9	2/9	2/9	1/9	2/9	1/9
$f(x_1, x_2; 2)$	1/16	1/8	1/4	1/16	1/4	1/4

$$= \begin{cases} 0 & \text{if} \quad (x_1; x_2) \in \{(0;0); (0;1); (1;1)\}; \\ 2 & \text{if} \quad (x_1; x_2) \in \{(0;2); (1;2); (2;2)\}. \end{cases}$$

(d)

$$\mathbb{E}[\widehat{\theta}_2] = 2\mathbb{P}\{(X_1; X_2) \in \{(0;2); (1;2); (2;2)\}\} = \begin{cases} 2 \cdot 0 = 0 & \text{if} \quad \theta = 0; \\ 2 \cdot 5/9 = 10/9 & \text{if} \quad \theta = 1; \\ 2 \cdot 3/4 = 3/2 & \text{if} \quad \theta = 2. \end{cases}$$

$$\text{Bias}(\widehat{\theta}_2; \theta) = \begin{cases} 0 - 0 = 0 & \text{if} \quad \theta = 0; \\ 1 - 10/9 = -1/9 & \text{if} \quad \theta = 1; \\ 2 - 3/2 = 1/2 & \text{if} \quad \theta = 2. \end{cases}$$

$$\mathbb{E}[\widehat{\theta}_2^2] = 4\mathbb{P}\{(X_1; X_2) \in \{(0;2); (1;2); (2;2)\}\} = \begin{cases} 4 \cdot 0 = 0 & \text{if} \quad \theta = 0; \\ 4 \cdot 5/9 = 20/9 & \text{if} \quad \theta = 1; \\ 4 \cdot 3/4 = 3 & \text{if} \quad \theta = 2. \end{cases}$$

$$Var(\widehat{\theta}_2) = \mathbb{E}[\widehat{\theta}_2^2] - (\mathbb{E}[\widehat{\theta}_2])^2 = \begin{cases} 0 & \text{if} \quad \theta = 0; \\ 80/81 & \text{if} \quad \theta = 1; \\ 3/4 & \text{if} \quad \theta = 2. \end{cases}$$

$$MSE[\widehat{\theta}_2; \theta] = Var(\widehat{\theta}_2) + (\text{Bias}(\widehat{\theta}_2; \theta))^2 = \begin{cases} 0 & \text{if} \quad \theta = 0; \\ 1 & \text{if} \quad \theta = 1; \\ 1 & \text{if} \quad \theta = 2. \end{cases}$$

(e) $\widehat{\theta}_2 = 2$.

3.12

(a) We define the following r.v.:

$$Y_i = \mathbb{I}_{\{X_i > 0\}}; \qquad Y_i \sim Be(p).$$

$$T_n = \sum_1^n Y_i; \qquad T_n \sim Bin(n, p).$$

$$p = \mathbb{P}\{Y_i = 1\} = \mathbb{P}\{X_i > 0\} = \mathbb{P}\left\{\frac{X_i - \mu}{\sigma} > -\frac{\mu}{\sigma}\right\} = 1 - \phi\left(-\frac{\mu}{\sigma}\right) = \phi\left(\frac{\mu}{\sigma}\right).$$

We know that:

$$\widehat{p}_{MLE} = \bar{Y}_n = \frac{T_n}{n}.$$

By the invariance principle of the MLE we obtain:

$$\begin{aligned}\frac{T_n}{n} &= \phi\Big(\frac{\mu}{\sigma}\Big);\\ &\Longrightarrow \phi^{-1}\Big(\frac{T_n}{n}\Big) = \frac{\mu}{\sigma};\\ &\Longrightarrow |\phi^{-1}\Big(\frac{T_n}{n}\Big)| = \frac{|\mu|}{\sigma};\\ &\Longrightarrow |\phi^{-1}\Big(\frac{T_n}{n}\Big)|^{-1} = CV.\end{aligned}$$

(b)

$$|\phi^{-1}\Big(\frac{T_n}{n}\Big)|^{-1} = CV < 1/3 \Longrightarrow \phi^{-1}\Big(\frac{T_n}{n}\Big) > 3 \Longrightarrow \frac{T_n}{n} > \phi(3).$$
$$\frac{T_n}{n} > \phi(3) \Longrightarrow T_{1000} > 1000 \cdot 0.9987 = 998.7.$$

3.13

(a) $X_i \sim N(\mu, \sigma^2)$.
(b)

$$\begin{aligned}L(\mu; \mathbf{x}) &= \prod \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\Big\{-\frac{(x_i-\mu)^2}{2\sigma^2}\Big\} =\\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{-\sum \frac{(x_i-\mu)^2}{2\sigma^2}\Big\}.\\ l(\mu; \mathbf{x}) &= -\frac{n}{2}\log(2\pi\sigma^2) - \sum \frac{(x_i-\mu)^2}{2\sigma^2}.\\ \frac{\mathrm{d}l(\mu; \mathbf{x})}{\mathrm{d}\mu} &= \sum \frac{2(x_i-\mu)}{2\sigma^2} = \frac{\sum x_i - n\mu}{\sigma^2} = 0.\end{aligned}$$

Therefore:

$$\widehat{\mu}_{MLE} = \frac{\sum X_i}{n} = \overline{X}_n.$$

In the case of error $\epsilon \sim f(s; \sigma^2) = \frac{\mathrm{e}^{-2|s|/\sigma^2}}{\sigma^2}$.

(c)

$$f_X(x;\mu) = \frac{e^{-2|x-\mu|/\sigma^2}}{\sigma^2}.$$

(d) show that the median $m(X_1, \ldots, X_n)$ is a maximum likelihood estimator for μ.

$$L(\mu;\mathbf{x}) = \prod_1^n \frac{e^{-2|x_i-\mu|/\sigma^2}}{\sigma^2} = \frac{e^{-2\sum |x_i-\mu|/\sigma^2}}{\sigma^{2n}}.$$

$$m(x_1, .., x_n) = \underset{\mu\in\mathbb{R}}{\operatorname{arginf}} \sum |x_i - \mu|.$$

Then it can be inferred that $\widehat{\mu}_{MLE} = m(X_1, \ldots, X_n)$.

3.14

(a) It is immediately proven that the theoretical mean, $\mathbb{E}[X]$, does not exist, since $\int_\theta^\infty \frac{\theta}{x} dx$ diverges. Therefore, the method of moments is not applicable.
(b) We calculate $L(\theta;\mathbf{x})$:

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \frac{\theta}{x_i^2} I_{[\theta,+\infty)}(x_i) = \frac{\theta^n}{\prod x_i} I_{[0,x_{(1)}]}(\theta).$$

By drawing $L(\theta;\mathbf{x})$, it is immediately seen that the maximum is reached for $X_{(1)}$, therefore $\hat{\theta}_{MLE} = X_{(1)}$.
(c) To prove that $\hat{\theta}_{MLE}$ is a minimal sufficient statistic, we use the Lehmann-Scheffé theorem.
We denote $T(\mathbf{x}) = \hat{\theta}_{MLE}$ and verify that the hypotheses are respected:

$\Rightarrow$

$$\text{given } \mathbf{x} \text{ and } \mathbf{y} \text{ different, } \frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} \text{ does not depend on } \theta \Rightarrow T(\mathbf{x}) = T(\mathbf{y}).$$

Proof
We know that the following quantity does not depend on θ by hypothesis:

$$\frac{\frac{\theta^n}{\prod x_i} I_{[0,x_{(1)}]}(\theta)}{\frac{\theta^n}{\prod y_i} I_{[0,y_{(1)}]}(\theta)}.$$

Then it must hold: $I_{[0,x_{(1)}]}(\theta) = I_{[0,y_{(1)}]}(\theta)$, therefore $x_{(1)} = y_{(1)}$, that is $T(\mathbf{x}) = T(\mathbf{y})$.

$\Leftarrow$

$$\text{given } \mathbf{x} \text{ and } \mathbf{y} \text{ different, } T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} \text{ does not depend on } \theta.$$

Proof
We know by hypothesis that $x_{(1)} = y_{(1)}$. Then the following ratio does not depend on θ:

$$\frac{\frac{\theta^n}{\prod x_i} I_{[0,x_{(1)}]}(\theta)}{\frac{\theta^n}{\prod y_i} I_{[0,y_{(1)}]}(\theta)}.$$

Since the L-S hypotheses hold, we can say that $T(\mathbf{X})$ is a minimal sufficient statistic for θ.

3.15

(a)

$$\mathbb{E}[X_i] = \frac{3\theta}{2} = \frac{\sum X_i}{n} \implies \hat{\theta}_{MOM} = \frac{2}{3}\overline{X}_n.$$

(b)

$$L(\theta;\mathbf{x}) = \prod \frac{1}{\theta}\mathbb{I}_{[\theta,2\theta]}(x_i) = \frac{1}{\theta^n}\prod \mathbb{I}_{[\theta,2\theta]}(x_i) = \frac{1}{\theta^n}\mathbb{I}_{[X_{(n)}/2,X_{(1)}]}(\theta).$$

Since $L(\theta;\mathbf{x})$ is monotonically decreasing, it can be concluded that $\hat{\theta}_{MLE} = X_{(n)}/2$.

(c) From the factorisation theorem we can identify in the $L(\theta;\mathbf{x})$, $g(T(\mathbf{x});\theta) = \frac{1}{\theta^n}\mathbb{I}_{[X_{(n)}/2,X_{(1)}]}(\theta)$ and $h(\mathbf{x}) = 1$. Therefore $T(\mathbf{x}) = (X_{(1)}, X_{(n)})$ is a sufficient statistic.
To prove that it is complete we use the L-S theorem. We verify the hypotheses:

$\Rightarrow$

$$\text{given different } \mathbf{x} \text{ and } \mathbf{y}, \frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} \text{ does not depend on } \theta \Rightarrow T(\mathbf{x}) = T(\mathbf{y}).$$

Proof
We know that the following quantity does not depend on θ by hypothesis:

$$\frac{\frac{1}{\theta^n}\mathbb{I}_{[x_{(n)}/2,x_{(1)}]}(\theta)}{\frac{1}{\theta^n}\mathbb{I}_{[y_{(n)}/2,y_{(1)}]}(\theta)}.$$

Then it must hold: $\mathbb{I}_{[x_{(n)}/2, x_{(1)}]}(\theta) = \mathbb{I}_{[y_{(n)}/2, y_{(1)}]}(\theta)$, thus $x_{(n)} = y_{(n)}$ and $x_{(1)} = y_{(1)}$, i.e. $T(\mathbf{x}) = T(\mathbf{y})$.

$\Leftarrow$

$$\text{given different } \mathbf{x} \text{ and } \mathbf{y},\ T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} \text{ does not depend on } \theta.$$

Proof

We know by hypothesis that $x_{(n)} = y_{(n)}$ and $x_{(1)} = y_{(1)}$. Then the following ratio does not depend on θ:

$$\frac{\frac{1}{\theta^n}\mathbb{I}_{[x_{(n)}/2, x_{(1)}]}(\theta)}{\frac{1}{\theta^n}\mathbb{I}_{[y_{(n)}/2, y_{(1)}]}(\theta)}.$$

Since the hypotheses of the theorem are satisfied we can say that $T(\mathbf{X})$ is a sufficient and complete statistic.

3.16

(a) We note that X is a shifted exponential: $X = \alpha + W$, where $W \sim \mathcal{E}(1/\beta)$.

$$\mathbb{E}[X] = \mathbb{E}[\alpha + W] = \alpha + \beta.$$

$$Var(X) = Var(\alpha + W) = \alpha^2 + Var(W) = \alpha^2 + \beta^2.$$

Let now $X_1, \ldots, X_n$ be a random sample from the density $f(x; \alpha, \beta)$.

(b)

$$L(\alpha, \beta; \boldsymbol{x}) = \prod_1^n \frac{1}{\beta} \mathrm{e}^{-(x_i-\alpha)/\beta}\, I_{[\alpha,+\infty)}(x_i) = \frac{1}{\beta^n} \mathrm{e}^{-\sum(x_i-\alpha)/\beta}\, I_{[\alpha,+\infty)}(x_{(1)}) =$$

$$= \frac{1}{\beta^n} \mathrm{e}^{-\sum(x_i-\alpha)/\beta}\, I_{[0,x_{(1)}]}(\alpha).$$

We evaluate α and β separately. I fix β and see that α varies as $\exp\{\alpha\}\mathbb{I}_{[0,x_{(1)}]}(\alpha)$, which is a monotonically increasing function. Therefore:

$$\widehat{\alpha}_{MLE} = X_{(1)}.$$

To evaluate $\widehat{\beta}_{MLE}$ we calculate and derive the log-likelihood:

$$l(\alpha, \beta; \boldsymbol{x}) = -n \cdot \log(\beta) - \frac{\sum(x_i - \alpha)}{\beta}.$$

$$\frac{\partial l(\alpha, \beta; \boldsymbol{x})}{\partial \beta} = -\frac{n}{\beta} + \frac{\sum(x_i - \alpha)}{\beta^2} = 0.$$

$$\widehat{\beta}_{MLE} = \overline{X} - \frac{\alpha}{n} = \overline{X} - \frac{X_{(1)}}{n}.$$

(c) The statistic $(\widehat{\alpha}_n, \widehat{\beta}_n)$ is sufficient for (α, β) by the factorisation theorem.
(d) By the principle of maximum likelihood:

$$\widehat{\mu}_n = \overline{X} - \frac{X_{(1)}}{n} + X_{(1)} = \overline{X} + \frac{(n-1)}{n} X_{(1)}.$$

(e)

$$\mathbb{E}[\widehat{\mu}_n] = \mathbb{E}\Big[\overline{X} + \frac{(n-1)}{n} X_{(1)}\Big] = \alpha + \beta + \frac{n-1}{n}\Big(\alpha + \frac{\beta}{n}\Big).$$

$$Var[\widehat{\mu}_n] = Var\Big(\overline{X} + \frac{(n-1)}{n} X_{(1)}\Big) = \frac{\alpha^2 + \beta^2}{n} + \frac{(n-1)^2}{n^2}\Big(\alpha^2 + \frac{\beta^2}{n^2}\Big).$$

$$\begin{aligned} MSE(\widehat{\mu}_n) &= Var[\widehat{\mu}_n] - (\mathbb{E}[\widehat{\mu}_n] - \mu)^2 = \\ &= \frac{\alpha^2 + \beta^2}{n} + \frac{(n-1)^2}{n^2}\Big(\alpha^2 + \frac{\beta^2}{n^2}\Big) - \Big(\frac{n-1}{n}\Big(\alpha + \frac{\beta}{n}\Big)\Big)^2 = \\ &= \frac{\alpha^2 + \beta^2}{n} - \frac{(n-1)^2}{n^2} \frac{2\alpha\beta}{n}. \end{aligned}$$

Chapter 4
Uniform Minimum Variance Unbiased Estimators (UMVUEs)

4.1 Theory Recap

Definition 4.1 (UMVUE) An estimator T^* is said to be an unbiased estimator of uniformly minimum variance, UMVUE, for $\tau(\theta)$ if it satisfies $\mathbb{E}_\theta[T^*] = \tau(\theta)\ \forall\theta$ and if $\forall$ unbiased estimator T for $\tau(\theta)$, it holds:

$$Var_\theta(T^*) \leq Var_\theta(T) \qquad \forall\theta.$$

Theorem 4.1 (Cramér-Rao Inequality) *Let $X_1, \ldots, X_n$ be a sample of r.v. with density $f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ and let $T(\boldsymbol{X}) = T(X_1, \ldots, X_n)$ be any estimator that satisfies:*

$$\frac{d}{d\theta}\mathbb{E}_\theta[T(\boldsymbol{X})] = \int_\chi \frac{\partial}{\partial\theta}[T(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x};\theta)]\, d\boldsymbol{x}$$

and

$$Var_\theta(T(\boldsymbol{X})) < \infty.$$

Then:

$$Var_\theta(T(\boldsymbol{X})) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[T(\boldsymbol{X})]\right)^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\right)^2\right]} = \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[T(\boldsymbol{X})]\right)^2}{I_n(\theta)};$$

where $I_n(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\right)^2\right]$ is called Fisher information.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_4

In the case where $X_1, \ldots, X_n$ are i.i.d. r.v., $I_n(\theta) = nI_1(\theta)$.

Lemma 4.1 *If $f_X(x;\theta)$ satisfies:*

$$\frac{d}{d\theta}\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f_X(x;\theta)\right] = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f_X(x;\theta)\right) f_X(x;\theta)\right] dx;$$

then:

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} f_X(x;\theta)\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta^2}\log f_X(x;\theta)\right].$$

It is important to note that the hypothesis of this Lemma is always satisfied by densities belonging to the exponential family.

Theorem 4.2 (Rao-Blackwell) *Let W be any unbiased estimator for $\tau(\theta)$ and let T be a sufficient statistic for θ. We define $\phi(T) = \mathbb{E}[W|T]$. Then $\mathbb{E}_\theta[\phi(T)] = \tau(\theta)$ and $Var_\theta(\phi(T)) \leq Var_\theta(W)$ $\forall\theta$, that is, $\phi(T)$ is a uniformly better unbiased estimator than W for $\tau(\theta)$.*

Theorem 4.3 *Let T be a sufficient and complete statistic for θ and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is UMVUE for $\mathbb{E}[\phi(T)]$.*

Theorem 4.4 (Uniqueness of UMVUE) *If W is UMVUE for $\tau(\theta)$, then W is unique.*

4.2 Exercises

Exercise 4.1 Consider the statistical model given by the exponential laws $\mathcal{E}(\nu)$, $\nu > 0$, (family of laws regular according to Fréchet, Cramér and Rao) and let $X_1, \ldots, X_n$ be a random sample drawn from a population described by such a model.

(a) Calculate the lower limit for the variance of an unbiased estimator of $\mathbb{E}_\nu[X] = 1/\nu$ based on the sample.
(b) Show that $\overline{X}_n$ is a UMVUE for $\mathbb{E}_\nu[X] = 1/\nu$.
(c) Starting from the statistic $\min\{X_1, \ldots, X_n\}$ construct another correct estimator for $\mathbb{E}_\nu[X] = 1/\nu$ and calculate its mean square error.
(d) Compare the two estimators.

Exercise 4.2 Let $X_1, \ldots, X_n$ be a random sample from a uniform law on the interval $[0, \theta]$, $\theta > 0$.

(a) Determine the maximum likelihood estimator of θ and calculate its bias.
(b) Deduce from (a) a correct estimator for θ and calculate its mean square error.
(c) Is the statistic found in (b) a UMVUE for θ?

Table 4.1 Estimators for the parameters μ, σ and σ^2

Parameter	UMVUE
μ	$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
σ^2	$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$
σ	$\sqrt{\frac{n-1}{2}}\frac{\Gamma[(n-1)/2]}{\Gamma[n/2]}S_n$

Exercise 4.3 Let $X_1, \ldots, X_n$ be a random sample from a normal law $N(\mu, \sigma^2)$ with unknown parameters. Show that the UMVUEs for the following parameters are precisely the estimators indicated in Table 4.1.

Exercise 4.4 Given a random sample $X_1, \ldots, X_n$ ($n \geq 1$) drawn from a population $B(p)$, $p \in [0, 1]$, find the UMVUE for p and p^2.

Exercise 4.5 Given a random sample $X_1, ..., X_n$ from a distribution $\mathcal{N}(\mu, 1)$, we want to estimate $\tau(\mu) = \mu^2$.

(a) Find T_n the maximum likelihood estimator for μ^2.
(b) Find $\widehat{\tau}_n$, the uniformly minimum variance unbiased estimator for μ^2.
(c) Calculate the variance of $\widehat{\tau}_n$. (It may be useful to remember that the fourth moment of a random variable $Y \sim \mathcal{N}(m, s^2)$ is $\mathbb{E}[Y^4] = m^4 + 6m^2s^2 + 3s^4$.)
(d) Show that the variance of $\widehat{\tau}_n$ is strictly greater than the Cramér–Rao limit.

Exercise 4.6 Given a random sample $X_1, \ldots, X_n, n \geq 2$, drawn from a population $N(\mu, \sigma^2)$, find the estimator of σ^2 of the form αS^2 with *minimum* mean square error.

Exercise 4.7 Given $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$, consider the estimator of $\tau(\lambda) = \mathbb{P}_\lambda(X = 0) = \mathrm{e}^{-\lambda}$ defined by $T = I_{\{0\}}(X)$.

(a) Show that T is the UMVUE of $\mathrm{e}^{-\lambda}$.
(b) Show that the mean square error of T does not reach the lower limit of Cram'er-Rao.

Exercise 4.8 Let $X_1, \ldots, X_n$ be a sample of rank n of independent random variables with density:

$$f_X(x; \theta) = \theta x^{\theta-1} 1_{(0,1)}(x), \qquad \theta > 0.$$

(a) Find the maximum likelihood estimator $\hat{\theta}_n$ of θ and calculate its bias.
(b) Deduce from (a) a corrected estimator for θ and calculate its mean square error.
(c) Does it satisfy the Cramér-Rao inequality?
(d) Is it the UMVUE?

Exercise 4.9 Let $(X_1, ..., X_n)$ be a random sample drawn from a distribution with density:

$$f(x;\theta) = \theta(1+x)^{-(1+\theta)} I_{(0,+\infty)}(x), \quad x \in \mathbb{R}, \quad \theta > 0.$$

(a) In the case where $\theta > 1$, estimate θ with the method of moments.
(b) Find, if they exist, the maximum likelihood estimators of θ and of $1/\theta$.
(c) Find, if it exists, a sufficient and complete statistic and determine its distribution.
(d) Find, if they exist, the UMVUE of θ and of $1/\theta$.
(e) Determine the Cramér-Rao lower limit for unbiased estimators of $1/\theta$. Compare this quantity with the mean square error of the UMVUE for $1/\theta$.

Exercise 4.10 Let $(X_1, ..., X_n)$ be a random sample drawn from a Poisson distribution with parameter $\lambda > 0$. Let $\tau(\lambda) = e^{-\lambda}(1+\lambda)$.

(a) Find a maximum likelihood estimator for $\tau(\lambda)$.
(b) Find an unbiased estimator of $\tau(\lambda)$.
(c) Find the UMVUE of $\tau(\lambda)$.

Exercise 4.11 Let $X_1, ..., X_n$ be a random sample from a $\Gamma(2, 1/\theta)$ with $\theta > 0$. Therefore, we have:

$$f(x;\theta) = \theta^{-2}\, x\, \mathrm{e}^{-x/\theta} I_{(0,+\infty)}(x).$$

(a) Determine a sufficient and complete statistic for θ.
(b) Determine the maximum likelihood estimator $\hat{\theta}_n$ for θ.
(c) Show that $\hat{\theta}_n$ coincides with the estimator $\bar{\theta}_n$ obtained by the method of moments.
(d) What is the law of $\hat{\theta}_n$?
(e) Is $\hat{\theta}_n$ biased?
(f) Is $\hat{\theta}_n$ UMVUE?
(g) Determine the maximum likelihood estimator $\widehat{\sigma}_n^2$ for the variance of X_1.

Exercise 4.12 Let X be a random variable with values in $(0, \infty)$ such that $\log(X)$ has a $N(\mu, 1)$ distribution with μ as an unknown real parameter. In other words, X has a log-normal distribution. For $n \geq 1$, let $X_1, ..., X_n$ be a random sample from the distribution of X.

(a) Calculate the mean θ of X.
(b) Determine the maximum likelihood estimator $T_n = T_n(X_1, ..., X_n)$ for θ.
(c) Calculate the bias of T_n for estimating θ.
(d) Starting from T_n, determine an estimator W_n that is UMVUE for θ.
(e) Calculate the Fisher information $I(\theta)$.

N.B. It may be helpful to remember the moment generating function of a $N(\mu, \sigma^2)$:

$$m(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}) \qquad \forall t \in \mathbb{R}.$$

Exercise 4.13 Let $X_1, \ldots, X_n$ be a random sample of size $n \geq 3$ drawn from a Bernoulli population with parameter $p \in [0, 1]$. Let T be the product of the first three observations, that is

$$T = X_1 X_2 X_3.$$

(a) Show that T is a correct estimator of p^3.
(b) Calculate the mean square error of T and compare it with the Cramér-Rao lower bound for correct estimators of p^3 based on a sample of size $n \geq 3$.
(c) Starting from T, find the UMVUE for p^3 based on a sample of size $n \geq 3$.

Exercise 4.14 Given a random sample $X_1, \ldots, X_n$ from a Bernoulli distribution $B(p)$, consider the statistic:

$$T(X_1, \ldots, X_n) = \begin{cases} 1, & \text{if } X_1 = 1,\ X_2 = 0; \\ 0, & \text{otherwise.} \end{cases}$$

(a) Verify that $T(X_1, \ldots, X_n)$ is an unbiased estimator of the variance σ^2 of the distribution.
(b) Do you find the estimates provided by $T(X_1, \ldots, X_n)$ interesting?
(c) Starting from $T(X_1, \ldots, X_n)$, construct the UMVUE $V(X_1, \ldots, X_n)$ for σ^2.

4.3 Solutions

4.1

(a) Let T be a generic unbiased estimator for $1/\nu$. We calculate the Cramér-Rao bound. We need to calculate $I_n(\nu) = nI_1(\nu)$.

$$\begin{aligned} I_1(\nu) &= \mathbb{E}\left[\left(\frac{\partial}{\partial \nu} \log f_X(x; \nu)\right)^2\right] = \\ &= \mathbb{E}\left[\left(\frac{\partial(\log \nu - \nu X)}{\partial \nu}\right)^2\right] = \\ &= \mathbb{E}[(1/\nu - X)^2] = Var(X) = 1/\nu^2. \end{aligned}$$

The Cramér-Rao inequality states that:

$$Var(T) \geq \frac{(\frac{d}{d\nu}(1/\nu))^2}{I_n(\nu)} = \frac{(-\frac{1}{\nu^2})^2}{\frac{n}{\nu^2}} = \frac{1}{n\nu^2}.$$

(b) $\overline{X}_n$ is an unbiased estimator of $\mathbb{E}[X] = \frac{1}{\nu}$.
Furthermore, $Var(\overline{X}_n) = 1/(n\nu^2)$ which reaches the Cramér-Rao bound, therefore $\overline{X}_n$ is UMVUE for $\mathbb{E}[X] = 1/\nu$.

(c) We define the following random variable W:

$$W = \min\{X_1, \ldots, X_n\}.$$

We calculate the law of W:

$$\mathbb{P}\{W \geq t\} = (\mathbb{P}\{X_i \geq t\})^n = e^{-\nu t n} \quad \Rightarrow \quad W \sim \mathcal{E}(n\nu).$$

Therefore:

$$\mathbb{E}[W] = \frac{1}{n\nu};$$

so nW is an unbiased estimator for $1/\nu$.

$$MSE(nW) = Var(nW) = n^2 \cdot \frac{1}{n^2\nu^2} = \frac{1}{\nu^2}.$$

(d)

$$MSE(nW) = \frac{1}{\nu^2} > \frac{1}{n\nu^2} = MSE(\overline{X}_n) \qquad \forall \nu.$$

This implies that $\overline{X}_n$ is the best estimator.

4.2

(a)

$$L(\theta; \mathbf{x}) = \prod \frac{1}{\theta}\mathbb{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\prod \mathbb{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\mathbb{I}_{[X_{(n)},+\infty]}(\theta).$$

$$\hat{\theta}_{MLE} = X_{(n)}.$$

We calculate its distribution and bias:

$$F_{X_{(n)}}(t) = \left(\frac{t}{\theta}\right)^n \qquad \Rightarrow \qquad f_{X_{(n)}}(t) = \frac{n}{\theta^n}t^{n-1}\mathbb{I}_{[0,\theta]}(t).$$

$$\mathbb{E}[\hat{\theta}_{MLE}] = \int_0^\theta \frac{n}{\theta^n} t^n \, \mathrm{d}t = \frac{n}{n+1}\theta.$$

$$Bias = \mathbb{E}[\hat{\theta}_{MLE}] - \theta = -\frac{1}{n+1}\theta.$$

(b) Hence:

$$\left(\frac{n+1}{n}\right)\hat{\theta}_{MLE} = \left(\frac{n+1}{n}\right) X_{(n)}$$

is an unbiased estimator for θ.

$$MSE\left[\left(\frac{n+1}{n}\right)\hat{\theta}_{MLE}\right] = \left(\frac{n+1}{n}\right)^2 Var(X_{(n)}) =$$

$$= \left(\frac{n+1}{n}\right)^2 \cdot \frac{n}{(n+1)^2(n+2)} \cdot \theta = \frac{\theta^2}{n(n+2)}.$$

(c) $\left(\frac{n+1}{n}\right)\hat{\theta}_{MLE}$ is a minimal and complete sufficient statistic (see Chap. 2, Exercise 2.4), therefore it is UMVUE for θ.

4.3

(a)

$$\mathbb{E}[\overline{X}_n] = \mu.$$

$\overline{X}_n$ is an unbiased estimator for μ. And being a function of a complete and minimal sufficient statistic, it is UMVUE of μ.

(b)

$$\frac{n-1}{\sigma^2}S_n^2 \sim \chi^2(n-1) = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right).$$

$$\mathbb{E}\left[\frac{n-1}{\sigma^2}S_n^2\right] = \frac{n-1}{\sigma^2}\mathbb{E}[S_n^2] = n-1 \implies \mathbb{E}[S_n^2] = \sigma^2.$$

S_n^2 is an unbiased estimator for σ^2. And being a function of a complete and minimal sufficient statistic, it is UMVUE of σ^2.

(c)

$$\mathbb{E}\Big[\sqrt{\frac{n-1}{\sigma^2}S_n^2}\Big] = \sqrt{\frac{n-1}{\sigma^2}}\mathbb{E}\Big[S_n\Big] = \int_0^{+\infty} t^{1/2} t^{\frac{n-1}{2}-1} e^{-t/2} (1/2)^{\frac{n-1}{2}} \frac{1}{\Gamma(\frac{n-1}{2})}\,dt =$$

$$= \int_0^{+\infty} t^{\frac{n}{2}-1} e^{-t/2} (1/2)^{\frac{n-1}{2}} \frac{1}{\Gamma(\frac{n-1}{2})}\,dt =$$

$$= \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}(1/2)^{-\frac{1}{2}} \int_0^{+\infty} t^{\frac{n}{2}-1} e^{-t/2} (1/2)^{\frac{n}{2}} \frac{1}{\Gamma(\frac{n}{2})}\,dt =$$

$$= \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sqrt{2}.$$

$$\sqrt{\frac{n-1}{\sigma^2}}\mathbb{E}\Big[S_n\Big] = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sqrt{2}.$$

$$\mathbb{E}\Big[S_n\Big] = \frac{\sigma}{\sqrt{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sqrt{2}.$$

Therefore:

$$\sqrt{\frac{n-1}{2}}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}S_n$$

is unbiased for σ. And being a function of a complete and minimal sufficient statistic, it is UMVUE of σ.

4.4 $\overline{X}_n$ is an unbiased estimator for p and is a minimal and complete sufficient statistic for p and therefore is UMVUE for p.

$$\mathbb{E}[(\overline{X}_n)^2] = \frac{p(1-p)}{n} + p^2 = \frac{p}{n} + p^2\left(1 - \frac{1}{n}\right).$$

Hence:

$$T = \left[\overline{X}_n^2 - \frac{\overline{X}_n}{n}\right] \cdot \frac{n}{n-1}$$

is an unbiased estimator for p^2 and a function of $\overline{X}_n$. Therefore, T is UMVUE for p^2.

4.5

(a) $\overline{X}_n$ is MLE for μ. Therefore, by the invariance principle $T_n = \overline{X}_n^2$ is MLE for μ^2.

(b)

$$\mathbb{E}[\overline{X}_n^2] = \frac{1}{n} + \mu^2 \qquad \Rightarrow \qquad \widehat{\tau}_n = \overline{X}_n^2 - \frac{1}{n}$$

is UMVUE for μ^2.

(c)

$$Var(\widehat{\tau}_n) = Var(\overline{X}_n^2) = \mathbb{E}[\overline{X}_n^4] - \left(\mathbb{E}[\overline{X}_n^2]\right)^2 =$$
$$= \mu^4 + 6\frac{\mu^2}{n} + \frac{3}{n^2} - \mu^4 - \frac{1}{n^2} - \frac{2\mu^2}{n} = \frac{4}{n}\mu^2 + \frac{2}{n^2}.$$

To calculate the variance of the estimator we have taken into account that: $\overline{X}_n \sim N(\mu, 1/n)$.

(d)

$$I_1 = \mathbb{E}_\mu\left[\left(\frac{\partial}{\partial\mu}\log f_X(x;\mu)\right)^2\right] = \mathbb{E}_\mu\left[\left(\frac{\partial}{\partial\mu}[-(x\mu)^2/2]\right)\right] = \mathbb{E}[(X-\mu)^2] = 1.$$

So the Cramér-Rao limit is $\frac{4\mu^2}{n}$ and the following inequality holds:

$$\frac{(\tau'(\mu))^2}{n-1} = \frac{4\mu^2}{n} < \frac{4}{n}\mu^2 + \frac{2}{n^2} = Var(\widehat{\tau}_n) \qquad \forall n.$$

4.6

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2).$$

$$MSE(\alpha S^2) = Var(\alpha S^2) + \left(\mathbb{E}[\alpha S^2] - \sigma^2\right)^2 =$$
$$= \alpha^2\frac{2\sigma^4}{n-1} + \left((\alpha-1)\sigma^2\right)^2 = \sigma^4\left(\frac{2\alpha^2}{n-1} + \alpha^2 + 1 - 2\alpha\right).$$

$$\frac{\partial\left[\sigma^4\left(\frac{2\alpha^2}{n-1} + \alpha^2 + 1 - 2\alpha\right)\right]}{\partial\alpha} = 2\alpha\left(\frac{2}{n-1} + 1\right) - 2 \leq 0.$$

$$\alpha \leq \frac{n-1}{n+1}.$$

So $MSE(\alpha S^2)$ is minimum for $T = \frac{n-1}{n+1} S^2$.

4.7

(a)

$$\tau(\lambda) = e^{-\lambda}; \qquad T = \mathbb{I}_{\{0\}}(X); \qquad T \sim Be(e^{-\lambda}).$$

$\tau(\lambda)$ is unbiased.

$$\mathbb{E}[T|X] = T \qquad \Rightarrow \qquad T \text{ is UMVUE for } e^{-\lambda}.$$

(b)

$$MSE(\tau) = Var(\tau) = e^{-\lambda}(1 - e^{-\lambda}).$$

$$I_1(\lambda) = \mathbb{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}\log f_X(x;\lambda)\right)^2\right] = \mathbb{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}[-\lambda + X\log\lambda]\right)^2\right] =$$
$$= \mathbb{E}_\lambda\left[\left(-1 + \frac{X}{\lambda}\right)^2\right] = \frac{1}{\lambda^2}Var(X) = \frac{1}{\lambda}.$$

The Cramér-Rao limit is:

$$\frac{(\tau'(\lambda))^2}{n/\lambda} = \frac{(-e^{-\lambda})^2}{1/\lambda} = \lambda e^{-2\lambda}.$$

So we can conclude that:

$$e^{-\lambda}(1-e^{-\lambda}) \geq \lambda e^{-2\lambda} \quad \iff \quad (1-e^{-\lambda}) \geq \lambda e^{-\lambda} \quad \iff \quad (e^{\lambda}-1) \geq \lambda.$$

$\forall\lambda > 0$ $MSE(\tau(\lambda))$ is greater than the Cramér-Rao limit.

4.8

(a)

$$L(\theta;\mathbf{x}) = \prod \theta x_i^{\theta-1}\mathbb{I}_{(0,1)}(x_i) = \theta^n\left(\prod x_i\right)^{\theta-1}\prod \mathbb{I}_{(0,1)}(x_i).$$

$$l(\theta;\mathbf{x}) = n\log(\theta) + (\theta-1)\sum\log(x_i) + \sum\log\mathbb{I}_{(0,1)}(x_i).$$

$$\frac{d\,l(\theta;\mathbf{x})}{d\theta} = n/\theta + \sum \log(x_i) = 0 \implies \hat{\theta}_{MLE} = -\frac{n}{\sum \log x_i}.$$

To calculate its bias, I investigate the distribution of $Y_i = -logX_i$.

$$\mathbb{P}\{Y_i \le y\} = \mathbb{P}\{-\log X_i \le y\} = \mathbb{P}\{X_i \ge \exp(-y)\} = 1 - \mathbb{P}\{X_i < \exp(-y)\} =$$
$$= 1 - \int_0^{e^{-y}} \theta x^{\theta-1} 1_{(0,1)}(x)\,dx = 1 - x^\theta \Big|_0^{e^{-y}} = 1 - e^{-\theta y}.$$

So $F_Y(y;\theta) = (1 - e^{-\theta y})\mathbb{I}_{[0,+\infty]}(y)$. We recognise the exponential distribution, $Y_i = -log(X_i) \sim \mathcal{E}(\theta)$. Due to the link between exponential and gamma, we can say: $Y_n = -\sum \log X_i \sim \Gamma(n,\theta)$.

$$\mathbb{E}[\hat{\theta}_{MLE}] = \mathbb{E}\Big[-\frac{n}{\sum \log x_i}\Big] = n\mathbb{E}\Big[-\frac{1}{\sum \log x_i}\Big] = n\mathbb{E}\Big[-\frac{1}{Y_n}\Big].$$

So we calculate:

$$\mathbb{E}\Big[-\frac{1}{Y_n}\Big] = \int_0^{+\infty} \frac{1}{y}\frac{\theta^n y^{n-1} e^{-\theta y}}{\Gamma(n)}\,dy = \frac{\Gamma(n-1)}{\Gamma(n)}\theta \int_0^{+\infty} \frac{\theta^{n-1} y^{n-2} e^{-\theta y}}{\Gamma(n-1)}\,dy =$$
$$= \frac{\Gamma(n-1)}{\Gamma(n)}\theta = \frac{(n-2)!}{(n-1)!}\theta = \frac{\theta}{(n-1)}.$$

So $\mathbb{E}[\hat{\theta}_{MLE}] = \frac{n\theta}{(n-1)}$.

(b)

$$T = \frac{n-1}{n}\hat{\theta}_{MLE} = -\frac{(n-1)}{\sum \log(X_i)}.$$

T is unbiased for θ.

$$MSE(T) = Var(T) = \frac{\theta^2}{(n-2)}.$$

(c) We calculate $I_n(\theta) = nI_1(\theta)$.

$$I_1(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}(\log\theta + (\theta-1)\log X)\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{\theta} + \log X\right)^2\right]$$
$$= Var(-\log X) = \frac{1}{\theta^2}.$$

Hence, the Cramér-Rao bound is:

$$\frac{\theta^2}{n} < \frac{\theta^2}{(n-2)} = Var(T).$$

(d) Given that $\sum \log X_i$ is a minimal and complete sufficient statistic for θ, T is UMVUE for θ.

4.9

(a)

$$X_1, \ldots, X_n \sim f_X(x;\theta) = \theta(1+x)^{-(1+\theta)} \mathbb{I}_{[0,+\infty)}(x) \qquad x \in \mathbb{R}, \quad \theta > 0.$$

Let $\theta > 1$, we apply the method of moments:

$$\begin{aligned}
\mathbb{E}_\theta[X] &= \int_0^{+\infty} x\theta(1+x)^{-(1+\theta)}\,\mathrm{d}x \overset{x+1=t}{=} \\
&= \int_1^{+\infty} \theta(t-1)t^{-(1+\theta)}\,\mathrm{d}t = \theta\int_1^{+\infty} t^{-\theta}\,\mathrm{d}t + \theta\int_1^{+\infty} t^{-(1+\theta)}\,\mathrm{d}t = \\
&= \theta\left[\frac{t^{-\theta+1}}{-\theta+1}\right]_1^{+\infty} - \theta\left[\frac{t^{-\theta}}{-\theta}\right]_1^{+\infty} = \\
&= \theta\left[-\frac{1}{1-\theta}\right] - \theta\left[\frac{1}{\theta}\right] = \\
&= -1 - \frac{\theta}{1-\theta} = \frac{-1+\theta-\theta}{1-\theta} = -\frac{1}{1-\theta} = \\
&= \frac{1}{\theta-1}.
\end{aligned}$$

Then, by the method of moments, we obtain:

$$\overline{X}_n = \frac{1}{\theta-1} \qquad \Rightarrow \qquad \hat{\theta}_{MOM} = 1 + \frac{1}{\overline{X}_n}.$$

(b) We study the likelihood of the sample:

$$L(\theta;\boldsymbol{x}) = \theta^n \left(\prod_i (1+x_i)\right)^{-(1+\theta)}.$$

$$l(\theta;\boldsymbol{x}) = n\log\theta - (1+\theta)\sum_i \log(1+x_i).$$

$$\frac{\partial l(\theta;\boldsymbol{x})}{\partial\theta} = \frac{n}{\theta} - \sum_i \log(1+x_i) \geq 0 \qquad \Longleftrightarrow \qquad \theta \leq \frac{n}{\sum \log(1+x_i)}.$$

Then:

$$\hat{\theta}_{MLE} = \frac{n}{\sum \log(1+X_i)};$$

and by the invariance principle:

$$\widehat{\left(\frac{1}{\theta}\right)}_{MLE} = \frac{\sum \log(1+X_i)}{n}.$$

(c)

$$f_X(x;\theta) = \theta \exp\{-(1+\theta)\log(1+x)\}\mathbb{I}_{[0,+\infty)}(x)$$

belongs to the exponential family, therefore $T = \sum_i \log(1+X_i)$ is sufficient by the factorisation criterion.
Moreover, given that:

$$w : \theta \to -(1+\theta) \qquad w : \mathbb{R}^+ \to (-\infty, 1) \supset \text{ open in } \mathbb{R};$$

T is also complete.
$Y = \log(1+X)$. We calculate the law of Y:

$$F_Y(t) = \mathbb{P}\{Y \leq t\} = \mathbb{P}\{\log(1+X) \leq t\} = \mathbb{P}\{(1+X) \leq e^t\} = \mathbb{P}\{X \leq e^t - 1\} =$$
$$= \int_0^{e^t-1} \theta(1+x)^{-(1+\theta)}\,\mathrm{d}x \overset{x+1=t}{=}$$
$$= \int_1^{e^t} \theta t^{-(1+\theta)}\,\mathrm{d}t = \left.\frac{\theta t^{-\theta}}{-\theta}\right|_1^{e^t} = 1 - e^{-t\theta}.$$

then $Y \sim \mathcal{E}(\theta)$, therefore $T = \sum_i \log(1+X_i) \sim \Gamma(n,\theta)$.

(d) We observe:

$$\mathbb{E}\left[\frac{T}{n}\right] = \frac{1}{n}\sum \mathbb{E}\left[\log(1+X_i)\right] = \frac{1}{\theta} \qquad \Rightarrow \qquad \frac{T}{n} \quad \text{is UMVUE for } 1/\theta.$$

$$\mathbb{E}\left[\frac{1}{T}\right] = \frac{\theta}{n-1} \qquad \Rightarrow \qquad \tilde{T} = \frac{n-1}{\sum \log(1+X_i)} \qquad \Rightarrow \qquad \tilde{T} \quad \text{is UMVUE for } \theta.$$

The equality $\mathbb{E}\left[\frac{1}{T}\right] = \frac{\theta}{n-1}$ is obtained through the properties of the gamma (see Chap. 3, Exercise 3.3).

(e) We calculate the Cramér-Rao lower bound, keeping in mind that $I_n(\theta) = nI_1(\theta)$.

$$I_1(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\right)^2\right] = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\right)^2\right] =$$
$$= \mathbb{E}_\theta\left[\left(\frac{1}{\theta} - \log(1+X)\right)^2\right] = Var(Y) \overset{Y\sim\mathcal{E}(\theta)}{=} \frac{1}{\theta^2}.$$

The Cramér-Rao bound is therefore:

$$\frac{(\tau'(\theta))^2}{n/\theta^2} = \frac{(-\frac{1}{\theta^2})^2}{n\frac{1}{\theta^2}} = \frac{1}{n\theta^2}.$$

Let's calculate the $MSE(\tilde{T})$.

$$MSE(\tilde{T}) = Var\left(\frac{\sum\log(1+X_i)}{n}\right) = \frac{1}{n\theta^2}.$$

Therefore, $\tilde{T}$ reaches the Cramér-Rao limit.

4.10

(a)

$$\tau(\lambda) = e^{-\lambda}(1+\lambda).$$
$$\widehat{\lambda}_{MLE} = \overline{X}_n.$$

Therefore, by the principle of invariance:

$$\widehat{\tau}(\lambda)_{MLE} = e^{-\overline{X}_n}(1+\overline{X}_n).$$

(b)

$$e^{-\lambda}(1+\lambda) = \mathbb{P}\{X \le 1\}.$$

If we introduce the r.v. $Y_i = \mathbb{I}_{[0,1]}(X_i)$, $Y_i \sim Be(e^{-\lambda}(1+\lambda))$, we observe that:

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{[0,1]}(X_i) \qquad \text{is an unbiased estimator for } \tau(\lambda).$$

(c) Let's calculate the UMVUE for $\tau(\lambda)$:

$$T = \mathbb{E}_\lambda\left[\overline{Y}_n \middle| \sum_{i=1}^n X_i\right].$$

$$\mathbb{E}_\lambda\left[Y_1 \middle| \sum_{i=1}^n X_i = k\right] =$$

$$= \mathbb{P}_\lambda\left\{Y = 1 \middle| \sum_{i=1}^n X_i = k\right\} = \mathbb{P}_\lambda\left\{X_1 \leq 1 \middle| \sum_{i=1}^n X_i = k\right\} =$$

$$= \mathbb{P}_\lambda\left\{X_1 = 0 \middle| \sum_{i=1}^n X_i = k\right\} + \mathbb{P}_\lambda\left\{X_1 = 1 \middle| \sum_{i=1}^n X_i = k\right\} =$$

$$= \frac{\mathbb{P}_\lambda\left\{X_1 = 0, \sum_{i=1}^n X_i = k\right\}}{\mathbb{P}_\lambda\left\{\sum_{i=1}^n X_i = k\right\}} + \frac{\mathbb{P}_\lambda\left\{X_1 = 1, \sum_{i=1}^n X_i = k\right\}}{\mathbb{P}_\lambda\left\{\sum_{i=1}^n X_i = k\right\}} =$$

$$= \frac{\mathbb{P}_\lambda\{X_1 = 0\}\,\mathbb{P}_\lambda\left\{\sum_{i=2}^n X_i = k\right\} + \mathbb{P}_\lambda\{X_1 = 1\}\,\mathbb{P}_\lambda\left\{\sum_{i=2}^n X_i = k-1\right\}}{\mathbb{P}_\lambda\left\{\sum_{i=1}^n X_i = k\right\}} =$$

$$= \left(e^{-\lambda} e^{-(n-1)\lambda} \frac{((n-1)\lambda)^k}{k!} + \lambda e^{-\lambda} e^{-(n-1)\lambda} \frac{((n-1)\lambda)^{k-1}}{(k-1)!}\right) \cdot \frac{k!}{e^{-n\lambda}(n\lambda)^k} =$$

$$= \left(\frac{n-1}{n}\right)^k + k\frac{(n-1)^k}{n^k} = \left(\frac{n-1}{n}\right)^k \cdot \left(1 + \frac{k}{n-1}\right).$$

Then:

$$\mathbb{E}_\lambda\left[X_1 \middle| \sum_{i=1}^n X_i\right] = \left(\frac{n-1}{n}\right)^{\sum_i X_i} \cdot \left(1 + \frac{\sum_i X_i}{n-1}\right)$$

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

is UMVUE for $\tau(\lambda)$.

4.11

(a)

$$X_1, \ldots, X_n \sim \Gamma(2, 1/\theta) \quad \theta > 0; \quad \Rightarrow \quad f_X(x;\theta) = \left(\frac{1}{\theta}\right)^2 x e^{-x/\theta} \mathbb{I}_{[0,+\infty)}(x).$$

$f_X(x;\theta)$ belongs to the exponential family, therefore $\sum_{i=1}^n X_i$ is a sufficient statistic. Since $w : \mathbb{R}^+ \to (-\infty, 0)$ contains an open set of $\mathbb{R}$, $\sum_{i=1}^n X_i$ is a sufficient and complete statistic, therefore also minimal.

(b) Let's calculate the likelihood:

$$L(\theta; \boldsymbol{x}) = \frac{1}{\theta^{2n}} \prod_{i=1} x_i e^{-\frac{\sum x_i}{\theta}} \prod_i \mathbb{I}_{[0,+\infty)}(x_i).$$

$$l(\theta; \boldsymbol{x}) \propto -2n \log\theta - \frac{\sum X_i}{\theta}.$$

$$\frac{\partial l(\theta; \boldsymbol{x})}{\partial\theta} = -\frac{2n}{\theta} + \frac{\sum x_i}{\theta^2} \geq 0 \qquad \theta \leq \frac{\sum x_i}{2n}.$$

Therefore, $\hat{\theta}_{MLE} = \frac{\overline{X}_n}{2}$.

(c) Let's calculate the mean of X:

$$\mathbb{E}[X] = \frac{1}{\theta} \qquad \Rightarrow \qquad \hat{\theta}_{MOM} = \frac{\overline{X}_n}{2}.$$

(d) Given that:

$$X_i \sim \Gamma(2, 1/\theta) \;\; \Rightarrow \;\; \sum X_i \sim \Gamma(2n, 1/\theta) \;\; \Rightarrow \;\; \hat{\theta}_n = \frac{\sum X_i}{2n} \sim \Gamma(2n, 2n/\theta).$$

(e) Given that $\mathbb{E}\left[\hat{\theta}_n\right] = \theta$, then $\hat{\theta}_n$ is an unbiased estimator for θ.

(f) $\hat{\theta}_n$ is UMVUE as it is an unbiased estimator and a function of a sufficient and complete statistic.

(g)

$$Var(X_i) = 2\theta^2 \qquad \Rightarrow \qquad \widehat{\sigma}^2_{MLE} = 2\frac{\overline{X}_n^2}{4} = \frac{\overline{X}_n^2}{2}$$

by the invariance principle.

4.12

(a)

$$\theta = \mathbb{E}[X] = \mathbb{E}[e^Y] = m_Y(1) = e^{\mu+1/2}.$$

(b)

$$f_X(x;\mu) = \frac{1}{\sqrt{2\pi}}\frac{1}{x}\exp\left\{-\frac{1}{2}(\log x-\mu)^2\right\}\mathbb{I}_{[0,+\infty)}(x).$$

$$L(\mu;\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}}\cdot\left(\prod_i x_i\right)\exp\left\{-\frac{1}{2}\sum_i(\log x_i-\mu)^2\right\}\prod_i \mathbb{I}_{[0,+\infty)}(x_i).$$

$$l(\mu;\boldsymbol{x}) \propto -\frac{1}{2}\sum_i(\log x_i-\mu)^2.$$

$$\frac{\partial l(\mu;\boldsymbol{x})}{\partial\mu} = \frac{1}{2}2\sum_i(\log x_i-\mu)$$

$$= \sum_i(\log x_i) - n\mu \geq 0 \qquad \Longleftrightarrow \qquad \mu \leq \frac{\sum\log x_i}{n}.$$

By the invariance principle of MLEs:

$$T_n = \hat{\theta}_{MLE} = \exp\left\{\frac{1}{2}+\frac{\sum\log X_i}{n}\right\}.$$

(c)

$$\mathbb{E}[\hat{\theta}_{MLE}] = \mathbb{E}\left[\exp\left\{\frac{1}{2}+\frac{\sum\log X_i}{n}\right\}\right] =$$

$$= \exp\left\{\frac{1}{2}\right\}\left(\exp\left\{\frac{\mu}{n}+\frac{1}{2n^2}\right\}\right)^n = \exp\left\{\frac{1}{2}\right\}\left(\exp\left\{\mu+\frac{1}{2n}\right\}\right) =$$

$$= \exp\left\{\mu+\frac{1}{2}\right\}\exp\left\{\frac{1}{2n}\right\}.$$

(d) The UMVUE W_n is simply obtained by correcting the estimator T_n:

$$W_n = \mathrm{e}^{-1/(2n)}T_n.$$

Indeed, by the Lehmann-Scheffé theorem, W_n is a function of a sufficient and complete statistic for θ, so it is the unique UMVUE of θ.

(e) Let $Y = \log X \sim N(\mu, 1)$. The definition of $I(\theta)$ for an i.i.d. sample is as follows:

$$I(\theta) = n \cdot \mathbb{E}\Big[\Big(\frac{\partial \log f_Y(y)}{\partial \theta}\Big)^2\Big].$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(y-\mu)^2\}.$$

$$\begin{aligned}\log f_Y(\log x) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}(y-\mu)^2 = \\ &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}(y - \log\theta + \frac{1}{2})^2.\end{aligned}$$

$$\frac{\partial \log f_Y(y)}{\partial \theta} = \frac{1}{\theta}(y - \log\theta + \frac{1}{2}).$$

$$\begin{aligned}I(\theta) &= n \cdot \mathbb{E}\Big[\frac{1}{\theta^2}(Y - \log\theta + \frac{1}{2})^2\Big] = n \cdot \frac{1}{\theta^2}\mathbb{E}\Big[(Y - \log\theta + \frac{1}{2})^2\Big] = \\ &= n \cdot \frac{1}{\theta^2}\mathbb{E}\Big[(Y-\mu)^2\Big] = \frac{n}{\theta^2} Var(Y) = \frac{n}{\theta^2}.\end{aligned}$$

4.13

(a)

$$X_1, \ldots, X_n \sim Be(p); \quad T = X_1 X_2 X_3; \quad T \sim Be(p^3).$$

$$\mathbb{E}[T] = p^3.$$

(b) We calculate the $MSE(T)$:

$$MSE(T) = Var(T) = p^3(1-p^3) = p^3(1-p)(1+p+p^2).$$

We calculate the Cramér-Rao limit:

$$\begin{aligned}I_1(p) &= \mathbb{E}_p\left[\left(\frac{\partial(\log f_X(x;p))}{\partial p}\right)^2\right] \\ &= \mathbb{E}_p\left[\left(\frac{\partial(x\log p + (1-x)\log(1-p))}{\partial p}\right)^2\right] =\end{aligned}$$

$$= \mathbb{E}_p\left[\left(\frac{x}{p} - \frac{1-x}{1-p}\right)^2\right] = \mathbb{E}_p\left[\frac{(x-px-p+px)^2}{(p(1-p))^2}\right] =$$

$$= \mathbb{E}_p\left[\frac{(x-p)^2}{(p(1-p))^2}\right] = \frac{1}{p(1-p)}.$$

The Cramér-Rao limit is therefore:

$$\frac{(3p^2)^2}{n\frac{1}{p(1-p)}} = \frac{9p^4p(1-p)}{n} = \frac{9p^5(1-p)}{n}.$$

$$MSE(T) = p^3(1-p^3) = p^3(1-p)(1+p+p^2) \geq \frac{9}{n}p^5(1-p)$$

$$(1+p+p^2) \geq \left(\frac{9}{n} - 1\right)p^2 \quad (9/n \leq 3).$$

(c) I know that $\sum_i X_i$ is a sufficient and complete statistic for p.
$T = X_1X_2X_3$ is correct for p^3.
Then $\mathbb{E}[T \big| \sum_i X_i]$ is UMVUE for p^3.

$$\mathbb{E}[T \Big| \sum_i X_i = k] = \mathbb{P}\left\{T = 1 \Big| \sum_i X_i = k\right\} = \frac{p^3\binom{n-3}{k-3}p^{k-3}(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} =$$

$$= \frac{(n-3)!k!}{(k-3)!n!} = \frac{k(k-1)(k-2)}{n(n-1)(n-2)}.$$

Then the UMVUE estimator is defined as follows:

$$\begin{cases} 0 & \text{if } \sum X_i \leq 2; \\ \frac{\sum X_i(\sum X_i - 1)(\sum X_i - 2)}{n(n-1)(n-2)} & \text{if } \sum X_i > 2. \end{cases}$$

4.14

(a)

$$\mathbb{E}[T] = \mathbb{P}\{X_1 = 1,\ X_2 = 0\} = p(1-p).$$

(b) No, since Var(T) and therefore MSE(T) do not depend on n.

(c) I exploit the Lehmann-Scheffé theorem, considering that $\sum X_i$ is a sufficient and complete statistic for p:

$$\begin{aligned}V(X_1, \dots, X_n) = \mathbb{E}[T|\sum X_i = k] = 1 \cdot \mathbb{P}\{X_1 = 1,\ X_2 = 0|\sum X_i = k\} = \\ = \frac{\mathbb{P}\{X_1 = 1,\ X_2 = 0, \sum_3^n X_i = k-1\}}{\mathbb{P}\{\sum_1^n X_i = k\}} = \\ = \frac{p(1-p)\binom{n-2}{k-1}p^{k-1}(1-p)^{n-2-k+1}}{\binom{n}{k}p^k(1-p)^{n-k}} = \\ = \frac{\binom{n-2}{k-1}}{\binom{n}{k}} = \frac{k(n-k)}{n(n-1)}.\end{aligned}$$

Therefore:

$$V(X_1, \dots, X_n) = \frac{\sum X_i(n - \sum X_i)}{n(n-1)} = \frac{n\overline{X}(1-\overline{X})}{n-1}.$$

Chapter 5
Likelihood Ratio Test

5.1 Theory Recap

Definition 5.1 (Errors in Hypothesis Testing) Consider the following hypothesis test:

$$H_0 : \theta \in \Theta_0 \qquad \text{vs} \qquad H_0 : \theta \in \Theta_0^c.$$

We then define:

- Type I error: H_0 is true, i.e. $\theta \in \Theta_0$, and we decide to reject H_0.
- Type II error: H_0 is false, i.e. $\theta \in \Theta_0^c$, and we decide to accept H_0.

See Table 5.1.

We define the Rejection Region, R. Then:

$$\mathbb{P}_\theta\{\boldsymbol{X} \in R\} = \begin{cases} \text{probability of committing Type I error} & \text{if} \quad \theta \in \Theta_0; \\ 1 - \text{probability of committing Type II error} & \text{if} \quad \theta \in \Theta_0^c. \end{cases}$$

Definition 5.2 (Power of the Test) The power function of a hypothesis test with rejection region R is a function of θ defined as follows:

$$\beta(\theta) = \mathbb{P}_\theta\{\boldsymbol{X} \in R\}.$$

Definition 5.3 (Size of the Test) The size α of a test with power function $\beta(\theta)$ is defined as follows:

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha;$$

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_5

Table 5.1 Errors in hypothesis testing

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct	Type I error
	H_1	Type II error	Correct

where $\alpha \in [0, 1]$.

Definition 5.4 (Level of the Test) The level α of a test with power function $\beta(\theta)$ is defined as follows:

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha;$$

where $\alpha \in [0, 1]$.

Definition 5.5 (Unbiased Test) A test with power function $\beta(\theta)$ is unbiased if:

$$\beta(\theta') \geq \beta(\theta'') \qquad \forall \theta' \in \Theta_0^c, \quad \theta'' \in \Theta_0.$$

Definition 5.6 (Likelihood Ratio Test, LRT) Consider the following test:

$$H_0 : \theta \in \Theta_0 \qquad \text{vs} \qquad H_1 : \theta \in \Theta_0^c.$$

The test statistic based on the likelihood ratio is defined as follows:

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\Theta_0} L(\theta; \boldsymbol{x})}{\sup_{\Theta} L(\theta; \boldsymbol{x})}.$$

The likelihood ratio test, LRT, is any test whose rejection region has the following form $\{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$, where $c \in (0, 1)$.

5.2 Exercises

Exercise 5.1 Given a random sample $X_1, \ldots, X_5$ from a law $B(p)$, with p unknown and $0 \leq p \leq 1$, we want to test the null hypothesis $H_0 : p = 1/2$ against the alternative hypothesis $H_1 : p \neq 1/2$. We intend to use a critical region of the type

$$R = \left\{ \left| \overline{x}_5 - \frac{1}{2} \right| > c \right\}.$$

(a) Find the values of c that give a test of size $\alpha = 10\%$.
(b) Find the values of c that give a test of level $\alpha = 10\%$.

Exercise 5.2 A sample of size 1 is extracted from a population $\mathcal{P}(\lambda)$. To verify $H_0 : \lambda = 1$ against $H_1 : \lambda = 2$, consider the critical region $R = \{x > 3\}$. Find the probabilities of Type I and Type II errors and the power of the test against $\lambda = 2$.

Exercise 5.3 Consider the statistical model given by the exponential laws $\mathcal{E}(\nu)$, $\nu > 0$, and let $X_1, \ldots, X_n$ be a random sample drawn from a population described by this model. Find the tests of size α based on the likelihood ratio for:

(a) $\nu = \nu_0$ against $\nu \neq \nu_0$.
(b) $\nu \leq \nu_0$ against $\nu > \nu_0$.

Exercise 5.4 Given $X \sim Bi(n, p)$, with n known and p unknown in $[0, 1]$:

(a) Find a level α test based on the likelihood ratio for $H_0 : p \leq p_0$ against $H_1 : p > p_0$.
(b) Explicitly write the rejection region in the case $n = 5$, $p_0 = 0.3$, $\alpha = 0.03$.

Exercise 5.5 Given a random sample $X_1, \ldots, X_n$, $n \geq 2$, drawn from a population $N(\mu, \sigma^2)$ with μ and σ both unknown, find the tests based on the likelihood ratio for $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$.

Exercise 5.6 Let $X_1, \ldots, X_n$ be a random sample from a uniform law on $\{1, \ldots, N\}$, where $N \in \mathbb{N}$. Find tests based on the likelihood ratio, also determining the level α, for:

(a) $N \leq N_0$ against $N > N_0$.
(b) $N = N_0$ against $N \neq N_0$.

Exercise 5.7 Let $X_1, \ldots, X_n$ be a random sample from a uniform law on the interval $[0, \theta]$, $\theta > 0$. Find the tests based on the likelihood ratio, also determining the level α, for:

(a) $\theta \leq \theta_0$ against $\theta > \theta_0$,
(b) $\theta = \theta_0$ against $\theta \neq \theta_0$.
(c) In the case $\theta_0 = 1$, find the minimum sample size n for which the test of size $\alpha = 5\%$ found in (b) has a power of at least 0.8 against $\theta = 3/2$.

Exercise 5.8 Let X be a sample of unit size from a distribution with density:

$$f(x; \theta) = \frac{2}{\theta^2}(\theta - x) I_{(0,\theta)}(x)$$

with $\theta \in (0, \infty)$. Consider the hypothesis test:

$$H_0 : \theta = 1 \quad \text{vs.} \quad H_1 : \theta > 1.$$

(a) Let δ_0 be the test with critical region:

$$R_0 = \{X > 1\}.$$

Calculate its level and its power function.

(b) Repeat the reasoning in the previous point in the case:

$$H_0 : \theta \leq 1 \quad \text{vs.} \quad H_1 : \theta > 1.$$

Exercise 5.9 Let $X_1, \ldots, X_n$ be a random sample from

$$f_X(x; \theta) = \frac{\theta}{x^2} I_{[\theta,+\infty)}(x), \qquad \theta > 0.$$

(a) Determine the critical region of the size α test based on the likelihood ratio for:

$$H_0 : \theta \leq 1 \qquad \text{vs} \qquad H_1 : \theta > 1.$$

(b) Calculate the power function of the test found in (a) and draw its graph.

(c) How large must n be if the test found in (a) of size $\alpha = 0.04$ is to have power 1 against $\theta = 3$?

Exercise 5.10 Let $X_1, \ldots, X_n$ be a random sample from a Normal population (θ, σ^2). Consider the test:

$$H_0 : \theta \leq \theta_0 \qquad \text{vs} \qquad H_1 : \theta > \theta_0.$$

(a) Assuming σ^2 is known, show that the test for which H_0 is rejected if

$$\overline{X} > \theta_0 + z_{1-\alpha}\sqrt{\sigma^2/n}$$

has size α. Also show that this test is equivalent to the one obtained from the likelihood ratios.

(b) Assuming σ^2 is unknown, show that the test that rejects H_0 if

$$\overline{X} > \theta_0 + t_{n-1,1-\alpha}\sqrt{S^2/n}$$

has size α. Also show that this test is equivalent to the one obtained from the likelihood ratios.

5.3 Solutions

5.1

(a)

$$\alpha = \mathbb{P}_{\frac{1}{2}}\left\{\left|\overline{X}_5 - \frac{1}{2}\right| > c\right\} = \mathbb{P}_{\frac{1}{2}}\left\{\frac{\sum X_i}{5} - \frac{1}{2} > c\right\} + \mathbb{P}_{\frac{1}{2}}\left\{\frac{\sum X_i}{5} - \frac{1}{2} < -c\right\} =$$

$$= \mathbb{P}_{\frac{1}{2}}\left\{\sum X_i > \left(\frac{1}{2} + c\right) \cdot 5\right\} + \mathbb{P}_{\frac{1}{2}}\left\{\sum X_i < \left(\frac{1}{2} - c\right) \cdot 5\right\}.$$

We know that, under H_0, $\sum X_i \sim Bin(5, 1/2)$. We define $k = \left(\frac{1}{2} + c\right) \cdot 5$ and $\tilde{k} = \left(\frac{1}{2} - c\right) \cdot 5$ and investigate how these values vary with c ($c \in [0, 1]$). We immediately observe that $2.5 \leq k \leq 7.5$ and $-2.5 \leq \tilde{k} \leq 2.5$. For a complete study, see Table 5.2.

In Table 5.3 the possible values of α are reported. We immediately observe that the test never reaches a *size* of 10%.

(b) From Table 5.3 it is immediately clear that to have a test of *level* α = 10%, we must choose $c \geq 3/10$.

5.2

$$R = \{x : X > 3\}.$$

$$\mathbb{P}\{\text{Type I error}\} = \mathbb{P}_{\lambda=1}\{X > 3\} = 1 - \mathbb{P}_{\lambda=1}\{X \leq 3\} =$$

$$= 1 - e^{-1}\left(1 + 1 + \frac{1}{2} + \frac{1}{6}\right) = 1 - e^{-1}\left(\frac{16}{6}\right) = 1 - \frac{8}{3e} = 0.019.$$

Table 5.2 Possible values of k, c and $\tilde{k}$

k	c	$\tilde{k}$
$2.5 \leq k \leq 3$	$0 \leq c \leq 1/10$	$2 \leq \tilde{k} \leq 2.5$
$3 \leq k \leq 4$	$1/10 \leq c \leq 3/10$	$1 \leq \tilde{k} \leq 2$
$4 \leq k \leq 5$	$3/10 \leq c \leq 1/2$	$0 \leq \tilde{k} \leq 1$
$k > 5$	$1/2 < c \leq 1$	$\tilde{k} > 0$

Table 5.3 Possible values of c and corresponding values of $\mathbb{P}\{\boldsymbol{x} \in R\}$

c	$\mathbb{P}_{\frac{1}{2}}\left\{\sum X_i > k\right\}$	$\mathbb{P}_{\frac{1}{2}}\left\{\sum X_i < \tilde{k}\right\}$	Total
$0 \leq c \leq 1/10$	1/2	1/2	1
$1/10 \leq c \leq 3/10$	6/32	6/32	3/8 = 0.375
$3/10 \leq c \leq 1/2$	1/32	1/32	1/16 = 0.0625
$c > 1/2$	0	0	0

$\mathbb{P}\{\text{Type II error}\} = \mathbb{P}_{\lambda=2}\{X \leq 3\} =$

$$= e^{-2}\left(1 + 2 + \frac{4}{2} + \frac{8}{6}\right) = e^{-2}\left(\frac{19}{3}\right) = 0.86.$$

The power against $\lambda = 2$ is $1 - \mathbb{P}_{\lambda=2}\{X > 3\} = 0.14$.

5.3

(a)

$$H_0 : \nu = \nu_0 \qquad \text{vs} \qquad H_1 : \nu \neq \nu_0.$$

We calculate $L(\nu; \mathbf{x})$ and apply the LRT:

$$L(\nu; \boldsymbol{x}) = \nu^n \exp\left\{ -\nu \sum x_i \right\} \prod \mathbb{I}_{[0,+\infty]}(x_i).$$

$$\lambda(\boldsymbol{x}) = \frac{L(\nu_0; \boldsymbol{x})}{\sup_{\nu>0} L(\nu; \boldsymbol{x})} = \frac{\nu_0^n \exp\left\{ -\nu_0 \sum x_i \right\} \prod \mathbb{I}_{[0,+\infty]}(x_i)}{\sup_{\nu>0} \nu^n \exp\left\{ -\nu \sum x_i \right\} \prod \mathbb{I}_{[0,+\infty]}(x_i)}.$$

The sup of the denominator corresponds to the $L(\nu; \mathbf{x})$ evaluated in correspondence with $\hat{\nu}_{MLE} = 1/\overline{X}_n$.

$$\begin{aligned}
\lambda(\boldsymbol{x}) &= \frac{\nu_0^n \exp\left\{ -\nu_0 \sum x_i \right\} \prod \mathbb{I}_{[0,+\infty]}(x_i)}{\left(\frac{1}{\overline{X}_n}\right)^n \exp\left\{ -\left(\frac{1}{\overline{X}_n}\right) \sum x_i \right\} \prod \mathbb{I}_{[0,+\infty]}(x_i)} = \\
&= \left(\nu_0 \overline{X}_n\right)^n \exp\left\{ n - \nu_0 \sum x_i \right\} = \\
&= \left(\nu_0 \overline{X}_n \exp\left\{ 1 - \nu_0 \overline{X}_n \right\}\right)^n.
\end{aligned}$$

We then define the critical region as: $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$ and $c \in [0, 1]$. In extreme cases, we find trivial results:

$$c = 0 \implies R = \emptyset \implies \text{never reject} \implies \mathbb{P}\{\text{Type II error}\} = 1.$$

$$c = 1 \implies R = \mathbb{R}^n \implies \text{always reject} \implies \mathbb{P}\{\text{Type I error}\} = 1.$$

We then focus on $c \in (0, 1)$.

$$\begin{aligned}
R &= \left\{ \boldsymbol{x} : \left(\nu_0 \overline{X}_n \exp\left\{ 1 - \nu_0 \overline{X}_n \right\}\right)^n \leq c \right\} = \\
&= \left\{ \boldsymbol{x} : \nu_0 \overline{X}_n \exp\left\{ 1 - \nu_0 \overline{X}_n \right\} \leq c^{1/n} = k \right\}.
\end{aligned}$$

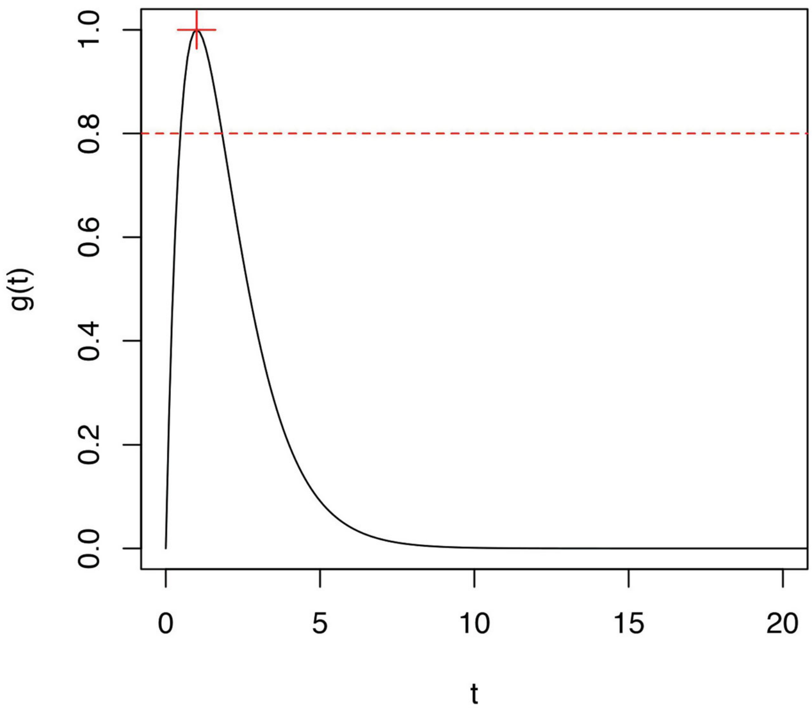

Fig. 5.1 Representation of $\lambda(\boldsymbol{x})$, where $g(t) = t \cdot (1 - e^{t})$ and the dashed horizontal line represents k. The maximum of $g(t)$ is 1 and is reached at $t = 1$ (the maximum is identified with a cross)

Refer to Fig. 5.1, where $t = \nu_0 \overline{X}_n$. We can say that: $R = \left\{\boldsymbol{x} : \nu_0 \overline{X}_n \leq \bar{t}_1\right\} \cup \left\{\boldsymbol{x} : \nu_0 \overline{X}_n \geq \bar{t}_2\right\}$. To define $\bar{t}_1$ and $\bar{t}_2$, we set the test level to α:

$$\alpha = \mathbb{P}_{\nu_0}\{\text{Type I error}\} = \mathbb{P}_{\nu_0}\{\boldsymbol{x} \in R\} = \mathbb{P}_{\nu_0}\{\nu_0 \overline{X}_n \leq \bar{t}_1\} + \mathbb{P}_{\nu_0}\{\nu_0 \overline{X}_n \geq \bar{t}_2\}.$$

We study the distribution of $\nu_0 \overline{X}_n$:

$$X_i \sim \mathcal{E}(\nu_0) = \Gamma(1, \nu_0) \implies \sum X_i \sim \Gamma(n, \nu_0) \implies \nu_0 \overline{X}_n \sim \Gamma(n, n).$$

Then, a possible choice is given by: $\bar{t}_1 = \gamma_{\alpha/2}(n, n)$ and $\bar{t}_2 = \gamma_{1-\alpha/2}(n, n)$.

(b)

$$H_0 : \nu \leq \nu_0 \qquad \text{vs} \qquad H_1 : \nu > \nu_0.$$

We carry out a procedure similar to that of point (a).

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{0<\nu\le\nu_0} L(\nu_0;\boldsymbol{x})}{\sup\limits_{\nu>0} L(\nu;\boldsymbol{x})} = \frac{\sup\limits_{0<\nu\le\nu_0} \nu^n \exp\left\{-\nu\sum x_i\right\}}{\sup\limits_{\nu>0} \nu^n \exp\left\{-\nu\sum x_i\right\}}$$
$$= \frac{\sup\limits_{0<\nu\le\nu_0} \nu^n \exp\left\{-\nu\sum x_i\right\}}{\left(\frac{1}{\overline{X}_n}\right)^n \exp\left\{-\left(\frac{1}{\overline{X}_n}\right)\sum x_i\right\}}.$$

We study the derivative of the numerator to see where (and if) the *sup* is reached:

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\nu^n \exp\left\{-\nu\sum x_i\right\} \ge 0$$
$$\nu^{n-1}\exp\left\{-\nu\sum x_i\right\}\left(n-\nu\sum x_i\right) \ge 0$$
$$\nu \le 1/\overline{X}_n.$$

The numerator has a maximum at $\hat{\nu}_{MLE} = 1/\overline{X}_n$.
Therefore, we need to distinguish two cases, based on whether $\hat{\nu}_{MLE} = 1/\overline{X}_n$ falls within the interval $(0, \nu_0]$:

$$\lambda(\boldsymbol{x}) = \begin{cases} \dfrac{\left(\frac{1}{\overline{X}_n}\right)^n \exp\left\{-\left(\frac{1}{\overline{X}_n}\right)\sum x_i\right\}}{\left(\frac{1}{\overline{X}_n}\right)^n \exp\left\{-\left(\frac{1}{\overline{X}_n}\right)\sum x_i\right\}} = 1, & \text{if} \quad \nu_0 \ge 1/\overline{X}_n; \\ \dfrac{\nu_0^n \exp\left\{-\nu_0\sum x_i\right\}}{\left(\frac{1}{\overline{X}_n}\right)^n \exp\left\{-\left(\frac{1}{\overline{X}_n}\right)\sum x_i\right\}} = \left(\nu_0\overline{X}_n \exp\left\{1-\nu_0\overline{X}_n\right\}\right)^n, & \text{if} \quad \nu_0 < 1/\overline{X}_n. \end{cases}$$

$$\begin{aligned} R &= \left\{\boldsymbol{x} : 1 \le c, \quad \nu_0\overline{X}_n \ge 1\right\} \\ &\quad \cup \left\{\boldsymbol{x} : \left(\nu_0\overline{X}_n \exp\left\{1-\nu_0\overline{X}_n\right\}\right)^n \le c, \quad \nu_0\overline{X}_n < 1\right\} \\ &= \emptyset \cup \left\{\boldsymbol{x} : \nu_0\overline{X}_n \exp\left\{1-\nu_0\overline{X}_n\right\} \le k, \quad \nu_0\overline{X}_n < 1\right\} = \\ &= \left\{\boldsymbol{x} : \nu_0\overline{X}_n \le \bar{t}_1\right\}. \end{aligned}$$

$\overline{X}_n$ has a $\Gamma(n, n\nu)$ distribution, so $\nu\overline{X}_n \sim \Gamma(n, n)$ and:

$$\alpha = \sup_{0<\nu\leq\nu_0} \mathbb{P}\left\{\nu_0\overline{X}_n \leq \bar{t}_1\right\} = \sup_{0<\nu\leq\nu_0} \mathbb{P}\left\{\nu\overline{X}_n \leq \frac{\nu}{\nu_0}\bar{t}_1\right\} = \mathbb{P}_{\nu_0}\left\{\nu_0\overline{X}_n \leq \frac{\nu}{\nu_0}\bar{t}_1\right\}.$$

The last equality is due to the fact that the sup is reached for $\nu = \nu_0$. Then $\frac{\nu}{\nu_0}\bar{t}_1 = \gamma_\alpha(n, n)$.

5.4

(a) We calculate $L(p; x)$ and apply the LRT:

$$L(p; x) = \binom{n}{x} p^k (1-p)^{n-x}.$$

$$\lambda(x) = \frac{\sup_{0\leq p\leq p_0} L(p; x)}{\sup_{p\in[0,1]} L(p; x)} = \frac{\sup_{0\leq p\leq p_0} \binom{n}{x} p^x (1-p)^{n-x}}{\sup_{p\in[0,1]} \binom{n}{x} p^x (1-p)^{n-x}}.$$

The sup of the denominator corresponds to $L(p; \mathbf{x})$ evaluated at $\hat{p}_{MLE} = x/n$ (see Fig. 5.2).

$$\lambda(x) = \frac{\sup_{0\leq p\leq p_0} p^x (1-p)^{n-x}}{\left(\frac{x}{n}\right)^x (1-\frac{x}{n})^{n-x}} = \begin{cases} 1, & \text{if } x/n \leq p_0 \leq 1; \\ (\frac{n\cdot p_0}{x})^x (\frac{n(1-p_0)}{n-x})^{n-x}, & \text{if } 0 \leq p_0 < x/n. \end{cases}$$

$$\begin{aligned} R = &\left\{x : 1 \leq c, \quad x/n \leq p_0 \leq 1\right\} \\ &\cup \left\{\boldsymbol{x} : \left(\frac{n\cdot p_0}{x}\right)^x \left(\frac{n(1-p_0)}{n-x}\right)^{n-x} \leq c, \quad 0 \leq p_0 < x/n\right\} = \\ = &\emptyset \cup \left\{x : x\log\left(\frac{n\cdot p_0}{x}\right)\right. \\ &\left. + (n-x)\log\left(\frac{n(1-p_0)}{n-x}\right) \leq \log c, \quad 0 \leq p_0 < x/n\right\} = \\ = &\left\{x : x\log(n\cdot p_0) - x\log x + (n-x)\log[n(1-p_0)]\right. \\ &\left. - (n-x)\log(n-x) \leq \log c, \quad 0 \leq p_0 < x/n\right\}. \end{aligned}$$

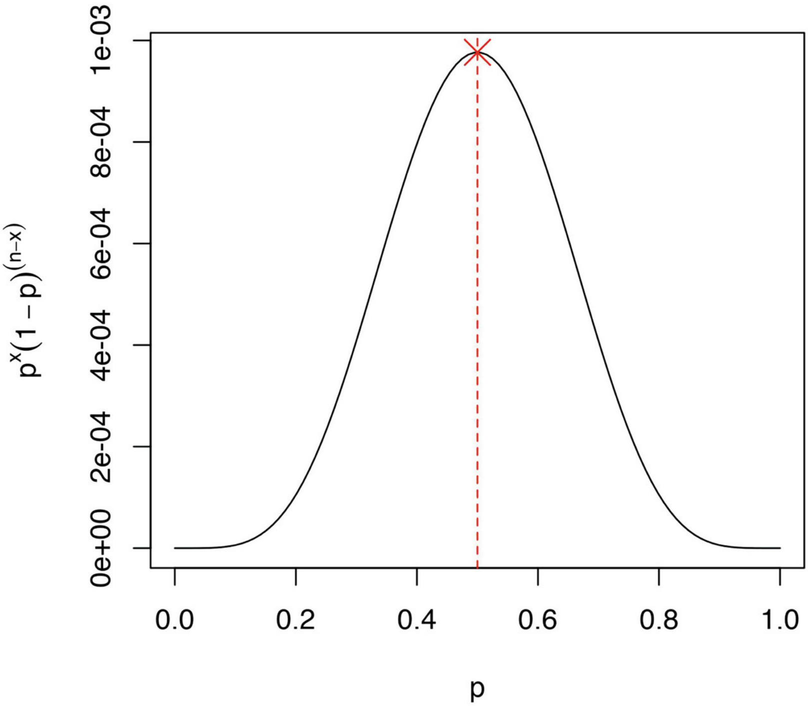

Fig. 5.2 Representation of $L(p; x)$. The maximum of $L(p; x)$ is reached for $p = \frac{x}{n}$ (in this case 0.5, since we arbitrarily chose $x = 5$ and $n = 10$)

We try to express in function of x. We therefore study the derivative of $f(x) = x \log(n \cdot p_0) - x \log x + (n - x) \log[n(1 - p_0)] - (n - x) \log(n - x)$.

$$\begin{aligned} f'(x) &= \log(n \cdot p_0) - \log(x) - 1 - \log[n(1 - p_0)] + \log(n - x) + 1 = \\ &= \log\left(\frac{n \cdot p_0}{n(1 - p_0)}\right) + \log\left(\frac{n - x}{x}\right) = \\ &= \log\left(\frac{p_0}{1 - p_0}\right) + \log\left(\frac{n - x}{x}\right) = \\ &= \log\left(\frac{p_0}{1 - p_0}\frac{n - x}{x}\right) = \log\left(\frac{p_0}{x}\frac{n - x}{1 - p_0}\right) \leq 0 \quad \text{for } p_0 \leq \frac{x}{n}. \end{aligned}$$

So we conclude that:

$$R = \left\{x : x \geq \tilde{c}\right\}.$$

That is, if we record a high number of successes, we reject H_0.

To make explicit $\tilde{c}$, requiring that the test has level α:

$$\alpha = \sup_{0 \le p \le p_0} \mathbb{P}\{X \ge \tilde{c}\} = \sup_{0 \le p \le p_0} 1 - \mathbb{P}\{X < \tilde{c}\}$$

$$= \sup_{0 \le p \le p_0} 1 - \sum_{x=0}^{\tilde{c}-1} \binom{n}{x} p^x (1-p)^{(n-x)}$$

It can be shown that the sup is realised for $p = p_0$ and consequently calculate numerically $\tilde{c}$.

(b)

$$\tilde{c} = 0 \implies \binom{5}{0} 0.3^0 (1-0.3)^{(5-0)} = 1.$$

$$\tilde{c} = 1 \implies \binom{5}{1} 0.3^1 (1-0.3)^{(5-1)} = 0.8319.$$

$$\tilde{c} = 2 \implies \binom{5}{2} 0.3^2 (1-0.3)^{(5-2)} = 0.3087.$$

$$\tilde{c} = 3 \implies \binom{5}{3} 0.3^3 (1-0.3)^{(5-3)} = 0.1631.$$

$$\tilde{c} = 4 \implies \binom{5}{4} 0.3^4 (1-0.3)^{(5-4)} = 0.0308.$$

$$\tilde{c} = 5 \implies \binom{5}{5} 0.3^5 (1-0.3)^{(5-5)} = 0.0024.$$

The test never reaches size $\alpha = 0.03$, but level $\alpha = 0.03$ yes: $R = \{x > 4\}$.

5.5 Let's calculate the LRT statistic, knowing that S_0^2 is MLE for σ^2:

$$\lambda(\boldsymbol{x}) = \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_i (x_i - \overline{X})^2\right\}}{(2\pi S_0^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2S_0^2} \sum_i (x_i - \overline{X})^2\right\}} =$$

$$= \left(\frac{S_0^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\left(\frac{\sum_i (x_i - \overline{X})^2}{\sigma_0^2} - n\right)\right\} =$$

$$= \left(\frac{S_0^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left\{\frac{n}{2}\left(1 - \frac{S_0^2}{\sigma_0^2}\right)\right\}.$$

We obtain as rejection region (or critical region) R:

$$R = \{\lambda \le c\} \iff \left\{ \frac{S_0^2}{\sigma_0^2} \exp\left\{1 - \frac{S_0^2}{\sigma_0^2}\right\} \le \sqrt[n]{c} = k \right\}.$$

Let $t = \frac{S_0^2}{\sigma_0^2}$. We therefore need to study the function $g(t) = t\exp(1-t)$ (see Fig. 5.1).

$$g'(t) = \exp\{1-t\} - t\exp\{1-t\} = \exp\{1-t\}(1-t);$$

therefore $g(t)$ is monotonically increasing in [0, 1] and monotonically decreasing in $[1, +\infty)$.

$$R = \left\{ \frac{S_0^2}{\sigma_0^2} < s_1 \right\} \cup \left\{ \frac{S_0^2}{\sigma_0^2} > s_2 \right\}.$$

Given that under H_0: $\frac{S_0^2}{\sigma_0^2} = \frac{n-1}{n}\frac{S^2}{\sigma_0^2} \sim \frac{1}{n}\chi^2(n-1)$, we can write R as:

$$R = \left\{ (n-1)\frac{S^2}{\sigma_0^2} < \chi^2_{\alpha/2}(n-1) \right\} \cup \left\{ (n-1)\frac{S^2}{\sigma_0^2} > \chi^2_{1-\alpha/2}(n-1) \right\}.$$

5.6

(a) $N \le N_0$ against $N > N_0$.

$$L(N; \boldsymbol{x}) = \prod_1^n \frac{1}{N}\mathbb{I}_{\{1,..,N\}}(x_i) = \frac{1}{N^n}\mathbb{I}_{\{x_{(n)},+\infty\}}(N).$$

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{N \le N_0} L(N; \boldsymbol{x})}{\sup\limits_{N \in \{1,+\infty)} L(N; \boldsymbol{x})} = \frac{\sup\limits_{N \le N_0} \frac{1}{N^n}\mathbb{I}_{\{x_{(n)},+\infty\}}(N)}{\sup\limits_{N \in \{1,+\infty)} \frac{1}{N^n}\mathbb{I}_{\{x_{(n)},+\infty\}}(N)} = \begin{cases} 0, & \text{if } x_{(n)} > N_0; \\ 1, & \text{if } x_{(n)} \le N_0. \end{cases}$$

We set up the R, focusing on $c \in (0, 1)$.

$$R = \left\{ \boldsymbol{x} : 1 \le c, \quad x_{(n)} \le N_0 \right\} \cup \left\{ \boldsymbol{x} : 0 \le c, \quad x_{(n)} > N_0 \right\} =$$
$$= \emptyset \cup \left\{ x_{(n)} > N_0 \right\}.$$

We choose a R of level α:

$$\alpha = \sup_{N \le N_0} \mathbb{P}\Big\{X_{(n)} > N_0\Big\} = 0.$$

(b) $N = N_0$ against $N \neq N_0$.

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{N=N_0} L(N; \boldsymbol{x})}{\sup\limits_{N \in \{1,+\infty)} L(N; \boldsymbol{x})} = \frac{\sup\limits_{N=N_0} \frac{1}{N^n}\mathbb{I}_{\{x_{(n)},+\infty\}}(N)}{\sup\limits_{N \in \{1,+\infty)} \frac{1}{N^n}\mathbb{I}_{\{x_{(n)},+\infty\}}(N)} =$$

$$= \begin{cases} 0, & \text{if } x_{(n)} > N_0; \\ (x_{(n)}/N_0)^n, & \text{if } x_{(n)} \le N_0. \end{cases}$$

We set the R, focusing on $c \in (0, 1)$:

$$R = \Big\{\boldsymbol{x} : (x_{(n)}/N_0)^n \le c, \quad x_{(n)} \le N_0\Big\} \cup \Big\{\boldsymbol{x} : 0 \le c, \quad x_{(n)} > N_0\Big\} =$$

$$= \Big\{x_{(n)} \le c^{1/n} N_0\Big\} \cup \Big\{x_{(n)} > N_0\Big\}.$$

We choose an R of level α:

$$\alpha = \sup_{N=N_0} \mathbb{P}\Big\{X_{(n)} \le c^{1/n} N_0\Big\} + \mathbb{P}\Big\{X_{(n)} > N_0\Big\} = \left(\frac{\lfloor c^{1/n} N_0 \rfloor}{N_0}\right)^n;$$

where $\lfloor a \rfloor$, $a \in \mathbb{R}$, indicates the lower integer part of a.

5.7

(a) We calculate $L(\theta; \mathbf{x})$ (see Fig. 5.3):

$$L(\theta; \boldsymbol{x}) = \prod_1^n \frac{1}{\theta}\mathbb{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\mathbb{I}_{[x_{(n)},+\infty]}(\theta).$$

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{0 \le \theta \le \theta_0} L(\theta; \boldsymbol{x})}{\sup\limits_{\theta \in [0,+\infty)} L(\theta; \boldsymbol{x})} = \frac{\sup\limits_{0 \le \theta \le \theta_0} \frac{1}{\theta^n}\mathbb{I}_{[x_{(n)},+\infty]}(\theta)}{\sup\limits_{\theta \in [0,+\infty)} \frac{1}{\theta^n}\mathbb{I}_{[x_{(n)},+\infty]}(\theta)}.$$

The sup of the denominator corresponds to $L(\theta; \mathbf{x})$ evaluated at $\hat{\theta}_{MLE} = X_{(n)}$.

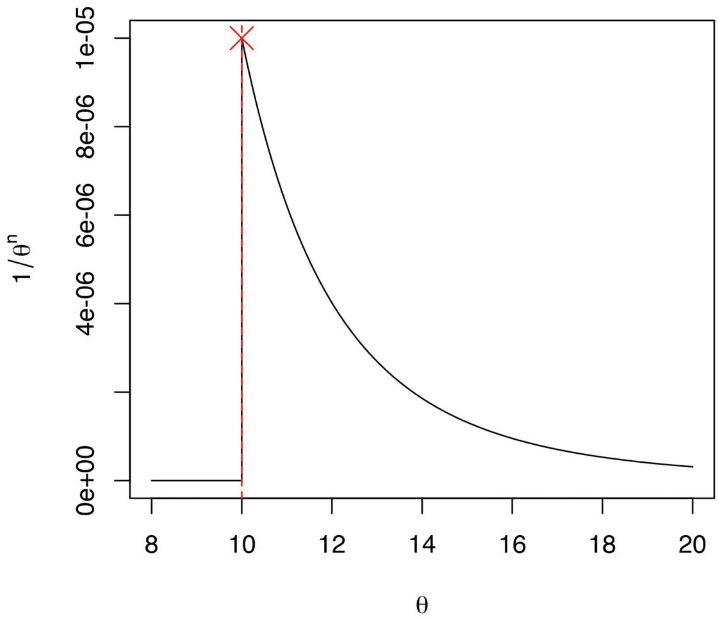

Fig. 5.3 Representation of $L(\theta; \boldsymbol{x})$. The maximum of $L(\theta; \boldsymbol{x})$, highlighted in the graph with a cross, is reached at $\theta = X_{(n)}$ (in this case equal to 10)

$$\lambda(\boldsymbol{x}) = \frac{\sup_{0 \le \theta \le \theta_0} \frac{1}{\theta^n} \mathbb{I}_{[x_{(n)}, +\infty]}(\theta)}{\frac{1}{x_{(n)}^n}} = \begin{cases} 1, & \text{if } x_{(n)} \le \theta_0; \\ 0, & \text{if } x_{(n)} > \theta_0. \end{cases}$$

We set the R, focusing on $c \in (0, 1)$:

$$R = \left\{ \boldsymbol{x} : 1 \le c, \quad x_{(n)} \le \theta_0 \right\} \cup \left\{ \boldsymbol{x} : 0 \le c, \quad x_{(n)} > \theta_0 \right\} = \\ = \emptyset \cup \left\{ x_{(n)} > \theta_0 \right\}.$$

We choose an R of level α:

$$\alpha = \sup_{0 \le \theta \le \theta_0} \mathbb{P}\left\{ X_{(n)} > \theta_0 \right\} = \sup_{0 \le \theta \le \theta_0} 1 - \mathbb{P}\left\{ X_{(n)} \le \theta_0 \right\} = \\ = \sup_{0 \le \theta \le \theta_0} 1 - \left(\mathbb{P}\left\{ X_1 \le \theta_0 \right\} \right)^n = \sup_{0 \le \theta \le \theta_0} 1 - (\theta_0/\theta)^n = 0.$$

(b) We set up the LRT:

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta=\theta_0} L(\theta;\boldsymbol{x})}{\sup_{\theta\in[0,+\infty)} L(\theta;\boldsymbol{x})} = \frac{\frac{1}{\theta_0^n}\mathbb{I}_{[x_{(n)},+\infty]}(\theta_0)}{\sup_{\theta\in[0,+\infty)} \frac{1}{\theta^n}\mathbb{I}_{[x_{(n)},+\infty]}(\theta)}.$$

The sup of the denominator corresponds to $L(\theta; \mathbf{x})$ evaluated at $\hat{\theta}_{MLE} = X_{(n)}$.

$$\lambda(\boldsymbol{x}) = \frac{\frac{1}{\theta_0^n}\mathbb{I}_{[x_{(n)},+\infty]}(\theta_0)}{\frac{1}{x_{(n)}^n}} = \begin{cases} (x_{(n)}/\theta_0)^n, & \text{if } x_{(n)} \le \theta_0; \\ 0, & \text{if } x_{(n)} > \theta_0. \end{cases}$$

We set the R, focusing on $c \in (0, 1)$:

$$R = \Big\{\boldsymbol{x} : (x_{(n)}/\theta_0)^n \le c, \quad x_{(n)} \le \theta_0\Big\} \cup \Big\{\boldsymbol{x} : 0 \le c, \quad x_{(n)} > \theta_0\Big\} =$$
$$= \Big\{x_{(n)} \le \theta_0\sqrt[n]{c}\Big\} \cup \Big\{x_{(n)} > \theta_0\Big\}.$$

We choose a R of level α:

$$\alpha = \sup_{\theta=\theta_0} \mathbb{P}\Big\{X_{(n)} \le \theta_0\sqrt[n]{c}\Big\} + \mathbb{P}\Big\{X_{(n)} > \theta_0\Big\}$$
$$= \mathbb{P}\Big\{X_{(n)} \le \theta_0\sqrt[n]{c}\Big\} + 1 - \mathbb{P}\Big\{X_{(n)} \le \theta_0\Big\} =$$
$$= \left(\mathbb{P}\Big\{X_1 \le \theta_0\sqrt[n]{c}\Big\}\right)^n + 1 - \left(\mathbb{P}\Big\{X_1 \le \theta_0\Big\}\right)^n$$
$$= (\theta_0\sqrt[n]{c}/\theta_0)^n + 1 - (\theta_0/\theta_0)^n = c.$$

So: $R_\alpha = \Big\{x_{(n)} \le \theta_0\sqrt[n]{\alpha}\Big\} \cup \Big\{x_{(n)} > \theta_0\Big\}$.

(c)

$$\beta(\theta) = \mathbb{P}\{\boldsymbol{x} \in R_\alpha\} = \mathbb{P}\{X_{(n)} \le \theta_0\sqrt[n]{\alpha}\} + \mathbb{P}\{X_{(n)} > \theta_0\} =$$
$$= (\theta_0\sqrt[n]{\alpha}/\theta)^n + 1 - (\theta_0/\theta)^n.$$

Substituting $\alpha = 5\%$ and $\theta_0 = 1$, we obtain the following power function:

$$\beta(\theta) = (\sqrt[n]{0.05}/\theta)^n + 1 - (1/\theta)^n.$$

We evaluate the function at $\theta = 3/2$ and require it to exceed 80%.

$$\beta(3/2) = 0.05 \cdot (2/3)^n + 1 - (2/3)^n \ge 0.8$$

$$-0.95 \cdot (2/3)^n \geq -0.2$$

$$(2/3)^n \leq 0.211$$

$$n \geq \frac{\log(0.211)}{\log(2/3)} = 3.84 \implies n \geq 4.$$

5.8

(a)

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}\{X \in R\} = \sup_{\theta=1} \mathbb{P}\{X > 1\} = \sup_{\theta=1} 1 - \mathbb{P}\{X \leq 1\} = 1 - \mathbb{P}_1\{X \leq 1\} =$$
$$= 1 - \int_0^1 2(1-x)\,\mathrm{dx} = 1 - (2x - x^2)\Big|_0^1 = 1 - 2 + 1 = 0.$$

$$\beta(\theta) = \mathbb{P}\{X \in R\} = \sup_{\theta=1} \mathbb{P}\{X > 1\} = \int_1^\theta \frac{2}{\theta^2}(\theta - x)\,\mathrm{dx} = \left(\frac{2}{\theta}x - \frac{2}{\theta^2}\frac{x^2}{2}\right)\Big|_0^\theta =$$
$$= 2 - \frac{2}{\theta} - 1 + \frac{1}{\theta^2} = 1 - \frac{2}{\theta} + \frac{1}{\theta^2} = \left(1 - \frac{1}{\theta}\right)^2 = \left(\frac{\theta - 1}{\theta}\right)^2.$$

The power function $\beta(\theta)$ is represented in Fig. 5.4.

(b)

$$H_0 : \theta \leq 1 \quad \text{vs.} \quad H_1 : \theta > 1.$$

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}\{X \in R\} = \sup_{\theta \leq 1} \mathbb{P}\{X > 1\} = \sup_{\theta \leq 1} 1 - \mathbb{P}\{X \leq 1\} =$$
$$= \sup_{\theta \leq 1} 1 - \int_0^1 \frac{2}{\theta^2}(\theta - x) I_{(0,\theta)}(x)\,\mathrm{dx} = 1 - \int_0^\theta \frac{2}{\theta^2}(\theta - x)\,\mathrm{dx} = 0.$$

$$\beta(\theta) = \mathbb{P}\{X \in R\} = \mathbb{P}\{X > 1\} = 1 - \mathbb{P}\{X \leq 1\} =$$
$$= 1 - \int_0^1 \frac{2}{\theta^2}(\theta - x) I_{(0,\theta)}(x)\,\mathrm{dx} =$$
$$= \begin{cases} 1 - \int_0^\theta \frac{2}{\theta^2}(\theta - x)\,\mathrm{dx} = 1 - 1 = 0 & \text{if} \quad \theta < 1; \\ 1 - \int_0^1 \frac{2}{\theta^2}(\theta - x)\,\mathrm{dx} = 1 - \frac{2}{\theta} + \frac{1}{\theta^2} & \text{if} \quad \theta \geq 1. \end{cases}$$

The power function $\beta(\theta)$ is represented in Fig. 5.5.

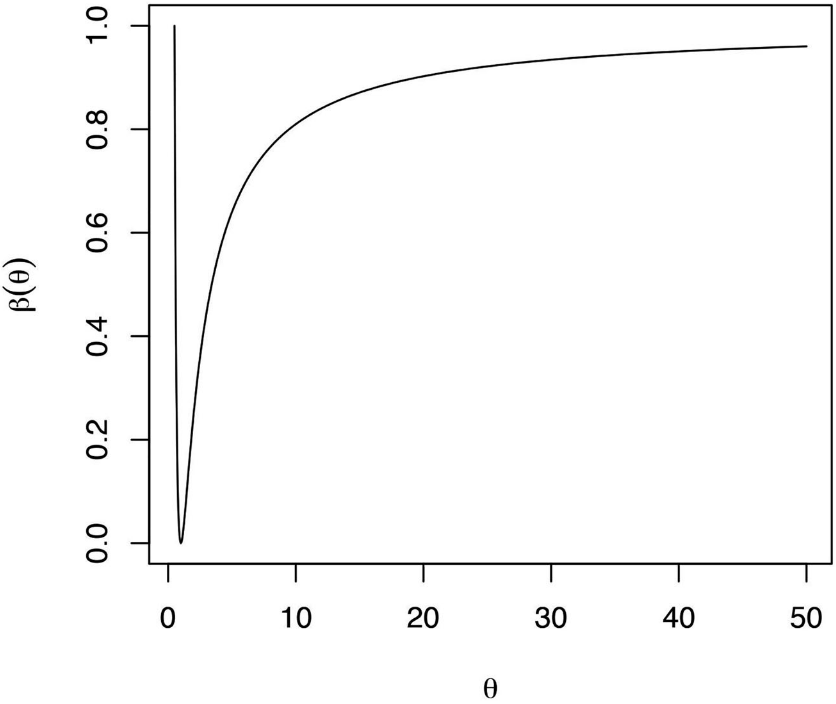

Fig. 5.4 Representation of $\beta(\theta) = \left(\frac{\theta-1}{\theta}\right)^2$. It can be noted that $\beta(\theta) \to +\infty$, for $\theta \to 0$; while $\beta(\theta) \to 1$, for $\theta \to +\infty$

5.9

(a) Let's calculate $L(\theta; \mathbf{x})$ (see Fig. 5.6).

$$L(\theta; \boldsymbol{x}) = \frac{\theta^n}{\prod x_i^2} \prod \mathbb{I}_{[0,+\infty]}(x_i) = \frac{\theta^n}{\prod x_i^2} \mathbb{I}_{[0,X_{(1)}]}(\theta).$$

We apply the definition of LRT.

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{0<\theta\leq 1} L(\theta; \boldsymbol{x})}{\sup\limits_{\theta>0} L(\theta; \boldsymbol{x})}.$$

The sup of the denominator corresponds to $L(\theta; \boldsymbol{x})$ evaluated at $\hat{\theta}_{MLE} = X_{(1)}$.

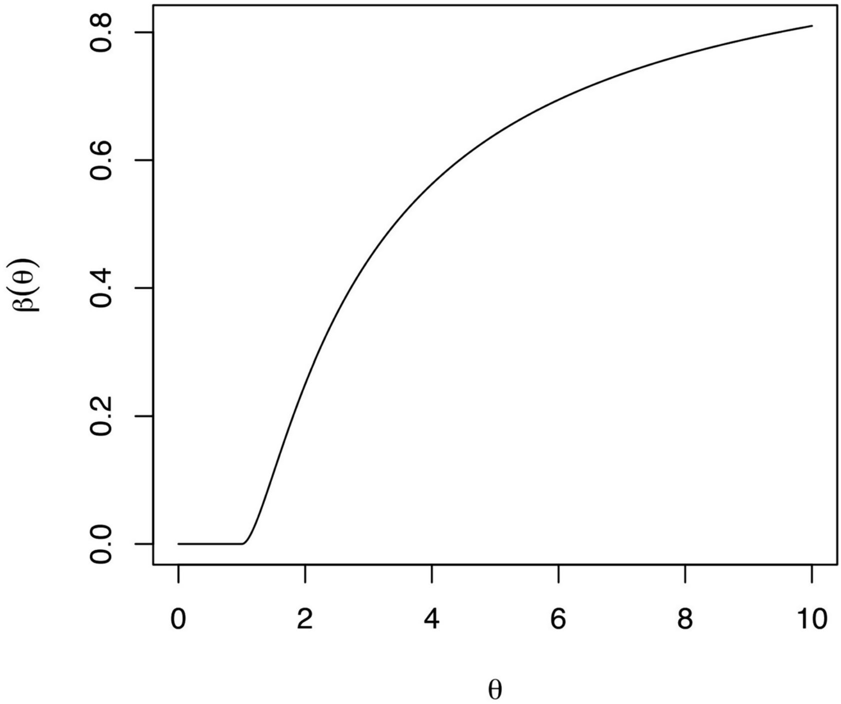

Fig. 5.5 Representation of $\beta(\theta) = \left(1 - \frac{2}{\theta} + \frac{1}{\theta^2}\right) \mathbf{1}_{\{\theta>=1\}}$

$$\lambda(\boldsymbol{x}) = \begin{cases} \frac{X_{(1)}^n}{\prod x_i^2} / \frac{X_{(1)}^n}{\prod x_i^2} = 1, & \text{if} \quad 1 \geq X_{(1)}; \\ \frac{1}{\prod x_i^2} / \frac{X_{(1)}^n}{\prod x_i^2} = \left(\frac{1}{X_{(1)}}\right)^n, & \text{if} \quad 1 < X_{(1)}. \end{cases}$$

We impose that the R is of level α and focus on $c \in (0, 1)$:

$$R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\} = \{\boldsymbol{x} : 1 \leq c, X_{(1)} \leq 1\} \cup \{\boldsymbol{x} : \left(\frac{1}{X_{(1)}}\right)^n \leq c, X_{(1)} > 1\} =$$

$$= \emptyset \cup \{\boldsymbol{x} : X_{(1)} \geq 1/c^{1/n} = 1/k, X_{(1)} > 1\} = \{\boldsymbol{x} : X_{(1)} \geq 1/k\}.$$

$$\alpha = \sup_{0<\theta\leq 1} \mathbb{P}\{\boldsymbol{x} \in R\} = \sup_{0<\theta\leq 1} \mathbb{P}\{X_{(1)} \geq 1/k\} = \sup_{0<\theta\leq 1} \left(\int_{1/k}^{+\infty} \theta/x^2 \, dx\right)^n =$$

$$= \sup_{0<\theta\leq 1} \left(\theta \cdot \left.\frac{-1}{x}\right|_{1/k}^{+\infty}\right)^n = \sup_{0<\theta\leq 1} \left(k\theta\right)^n = k^n \implies k = \sqrt[n]{\alpha}.$$

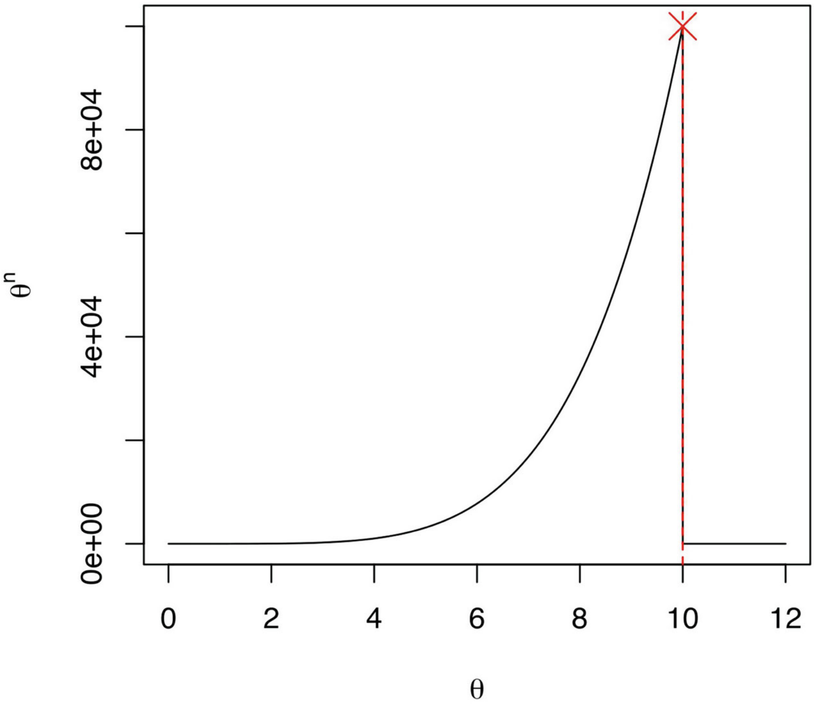

Fig. 5.6 Representation of $L(\theta; \boldsymbol{x})$. The maximum of $L(\theta; \boldsymbol{x})$ is represented with a cross and is reached at $\theta = X_{(1)}$ (10 in this case)

$$R_\alpha = \{\boldsymbol{x} : X_{(1)} \geq 1/\sqrt[n]{\alpha}\}.$$

(b) Let's calculate the power function of the test found in (a):

$$\beta(\theta) = \mathbb{P}\{\boldsymbol{x} \in R\} = \left(\int_{1/\sqrt[n]{\alpha}}^{+\infty} \theta/x^2\,dx\right)^n = \begin{cases} 1, & \text{if } \theta \geq 1/\sqrt[n]{\alpha}; \\ \alpha\theta^n, & \text{if } \theta < 1/\sqrt[n]{\alpha}. \end{cases}$$

The function $\beta(\theta)$ is represented in Fig. 5.7.

(c) Let's calculate the minimum value of n such that the test of size $\alpha = 0.04$ found in (a) has power 1 against $\theta = 3$.

$$\beta(\theta) = 1 \implies \theta \geq 1/\sqrt[n]{\alpha} \implies \alpha \geq 1/\theta^n \implies \log\alpha \geq -n\log\theta$$
$$\implies n \geq -\frac{\log\alpha}{\log\theta} = -\frac{\log 0.04}{\log 3} = 2.93.$$

We conclude that $n \geq 3$.

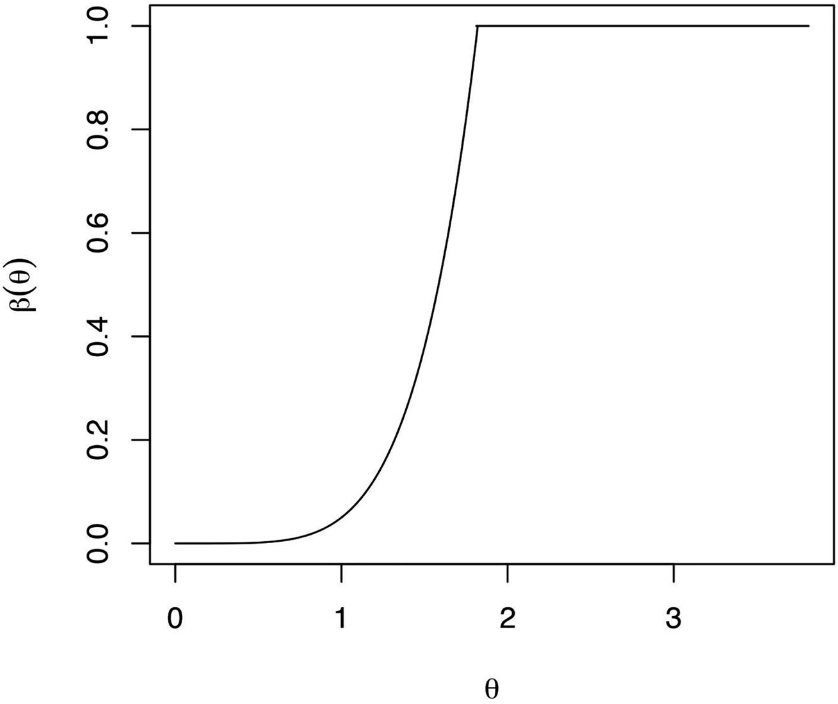

Fig. 5.7 Representation of $\beta(\theta)$. The angular point is recorded at $\theta = 1/\sqrt[n]{\alpha}$ (in this case about 1.82, since we chose $n = 5$ and $\alpha = 0.05$)

5.10

(a) We show that the size of R is indeed α, remembering that $\overline{X} \sim N(\theta, \sigma^2/n)$:

$$\begin{aligned}
\alpha &= \sup_{\theta \leq \theta_0} \mathbb{P}\{\boldsymbol{x} \in R\} = \sup_{\theta \leq \theta_0} \mathbb{P}\{\overline{X} > \theta_0 + z_{1-\alpha}\sqrt{\sigma^2/n}\} = \\
&= \sup_{\theta \leq \theta_0} \mathbb{P}\left\{\frac{\overline{X} - \theta}{\sqrt{\sigma^2/n}} > \frac{\theta_0 + z_{1-\alpha}\sqrt{\sigma^2/n} - \theta}{\sqrt{\sigma^2/n}}\right\} = \\
&= \sup_{\theta \leq \theta_0} 1 - \Phi\left(\frac{\theta - \theta_0}{\sqrt{\sigma^2/n}} + z_{1-\alpha}\right) = \\
&= 1 - \Phi(z_{1-\alpha}) = 1 - (1 - \alpha) = \alpha.
\end{aligned}$$

We now show that the same R can be obtained through LRT:

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{\theta \le \theta_0} L(\theta; \boldsymbol{x})}{\sup\limits_{\theta \in \mathbb{R}} L(\theta; \boldsymbol{x})} = \frac{\sup\limits_{\theta \le \theta_0} (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum (x_i-\theta)^2}{2\sigma^2}\}}{\sup\limits_{\theta \in \mathbb{R}} (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum (x_i-\theta)^2}{2\sigma^2}\}}.$$

The sup of the denominator corresponds to $L(\theta; \boldsymbol{x})$ evaluated at $\hat{\theta}_{MLE} = \overline{X}$.

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{\theta \le \theta_0} \exp\{-\frac{\sum (x_i-\theta)^2}{2\sigma^2}\}}{\exp\{-\frac{\sum (x_i-\overline{X})^2}{2\sigma^2}\}} = \frac{\sup\limits_{\theta \le \theta_0} \exp\{-\frac{\sum (x_i-\overline{X}+\overline{X}-\theta)^2}{2\sigma^2}\}}{\exp\{-\frac{\sum (x_i-\overline{X})^2}{2\sigma^2}\}} =$$

$$= \frac{\sup\limits_{\theta \le \theta_0} \exp\{-\frac{\sum (x_i-\overline{X})^2+(\overline{X}-\theta)^2+2(x_i-\overline{X})(\overline{X}-\theta)}{2\sigma^2}\}}{\exp\{-\frac{\sum (x_i-\overline{X})^2}{2\sigma^2}\}} =$$

$$= \frac{\sup\limits_{\theta \le \theta_0} \exp\{-\frac{\sum (x_i-\overline{X})^2+(\overline{X}-\theta)^2}{2\sigma^2}\}}{\exp\{-\frac{\sum (x_i-\overline{X})^2}{2\sigma^2}\}} = \sup_{\theta \le \theta_0} \exp\left\{ -\frac{\sum (\overline{X}-\theta)^2}{2\sigma^2} \right\}.$$

We can conclude that:

$$\lambda(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \overline{X} < \theta_0; \\ \exp\left\{ -\frac{n \cdot (\overline{X}_n-\theta_0)^2}{2\sigma^2} \right\}, & \text{if } \overline{X} \ge \theta_0. \end{cases}$$

We set the R, focusing on $c \in (0, 1)$:

$$R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \le c\} =$$

$$= \{\boldsymbol{x} : 1 \le c, \overline{X} < \theta_0\} \cup \{\boldsymbol{x} : \exp\left\{ -\frac{n \cdot (\overline{X}-\theta_0)^2}{2\sigma^2} \right\} \le c, \overline{X} \ge \theta_0\} =$$

$$= \emptyset \cup \{\boldsymbol{x} : \exp\left\{ -\frac{n \cdot (\overline{X}-\theta_0)^2}{2\sigma^2} \right\} \le c, \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \overline{X} \ge \sqrt{-\frac{2\sigma^2}{n} \log c} + \theta_0,\ \overline{X} \ge \theta_0\}.$$

Now we impose the size of R equal to α:

$$\alpha = \sup_{\theta \le \theta_0} \mathbb{P}\{\boldsymbol{x} \in R\} = \sup_{\theta \le \theta_0} \mathbb{P}\Big\{\overline{X} \ge \sqrt{-\frac{2\sigma^2}{n}\log c} + \theta_0\Big\} =$$

$$= \sup_{\theta \le \theta_0} \mathbb{P}\left\{\frac{\overline{X} - \theta}{\sqrt{\sigma^2/n}} \ge \frac{\sqrt{-\frac{2\sigma^2}{n}\log c} + \theta_0 - \theta}{\sqrt{\sigma^2/n}}\right\} =$$

$$= \sup_{\theta \le \theta_0} 1 - \Phi\left(\frac{\theta_0 - \theta + \sqrt{-\frac{2\sigma^2}{n}\log c}}{\sqrt{\sigma^2/n}}\right) =$$

$$= 1 - \Phi\left(\sqrt{-2\log c}\right) \implies \sqrt{-2\log c} = z_{1-\alpha} \implies c = e^{-\frac{z_{1-\alpha}^2}{2}}.$$

Then $R_\alpha = \{\boldsymbol{x} : \overline{X} > \theta_0 + z_{1-\alpha}\sqrt{\sigma^2/n}\}$.

(b) We show that the size of this R is indeed α, remembering that $\frac{\overline{X}-\theta}{\sqrt{S^2/n}} \sim t_{n-1}$.

$$\alpha = \sup_{\theta \le \theta_0} \mathbb{P}\{\boldsymbol{x} \in R\} = \sup_{\theta \le \theta_0} \mathbb{P}\{\overline{X} > \theta_0 + t_{n-1,1-\alpha}\sqrt{S^2/n}\} =$$

$$= \sup_{\theta \le \theta_0} \mathbb{P}\left\{\frac{\overline{X} - \theta}{\sqrt{S^2/n}} > \frac{\theta_0 + t_{n-1,1-\alpha}\sqrt{S^2/n} - \theta}{\sqrt{S^2/n}}\right\} =$$

$$= \sup_{\theta \le \theta_0} 1 - t_{n-1}\left(\frac{\theta_0 - \theta}{\sqrt{S^2/n}} + t_{n-1,1-\alpha}\right) = 1 - (1-\alpha) = \alpha.$$

We now show that the same R can be obtained through LRT:

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \le \theta_0,\, \sigma^2 > 0} L(\theta; \boldsymbol{x})}{\sup_{\theta \in \mathbb{R},\, \sigma^2 > 0} L(\theta; \boldsymbol{x})} = \frac{\sup_{\theta \le \theta_0,\, \sigma^2 > 0} (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum(x_i-\theta)^2}{2\sigma^2}\}}{\sup_{\theta \in \mathbb{R},\, \sigma^2 > 0} (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum(x_i-\theta)^2}{2\sigma^2}\}}.$$

The sup of the denominator corresponds to $L(\theta; \boldsymbol{x})$ evaluated at $\hat{\theta}_{MLE} = \overline{X}$ and $\hat{\sigma}^2_{MLE} = \hat{\sigma}^2 = \frac{\sum(x_i - \overline{X})^2}{n} = \frac{n-1}{n}S^2$.

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \le \theta_0,\, \sigma^2 > 0} \sigma^{-n} \exp\{-\frac{\sum(x_i-\theta)^2}{2\sigma^2}\}}{\hat{\sigma}^{-n} \exp\{-\frac{\sum(x_i-\overline{X})^2}{2\hat{\sigma}^2}\}}$$

$$= \frac{\sup\limits_{\theta \le \theta_0,\, \sigma^2 > 0} \sigma^{-n} \exp\{-\frac{\sum(x_i - \overline{X} + \overline{X} - \theta)^2}{2\sigma^2}\}}{\hat{\sigma}^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2}{2\hat{\sigma}^2}\}} =$$

$$= \frac{\sup\limits_{\theta \le \theta_0,\, \sigma^2 > 0} \sigma^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2 + (\overline{X} - \theta)^2 + 2(x_i - \overline{X})(\overline{X} - \theta)}{2\sigma^2}\}}{\hat{\sigma}^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2}{2\hat{\sigma}^2}\}} =$$

$$= \frac{\sup\limits_{\theta \le \theta_0,\, \sigma^2 > 0} \sigma^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2 + (\overline{X} - \theta)^2}{2\sigma^2}\}}{\hat{\sigma}^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2}{2\hat{\sigma}^2}\}}.$$

Writing $\hat{\sigma}_0^2 = \frac{\sum(x_i - \theta_0)^2}{n} = S_0^2 = (\overline{X} - \theta_0)^2 + \frac{n-1}{n} S^2$, we can conclude that:

$$\lambda(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \overline{X} < \theta_0; \\ \frac{\hat{\sigma}_0^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2 + (\overline{X} - \theta_0)^2}{2\hat{\sigma}_0^2}\}}{\hat{\sigma}^{-n} \exp\{-\frac{\sum(x_i - \overline{X})^2}{2\hat{\sigma}^2}\}} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}, & \text{if } \overline{X} \ge \theta_0. \end{cases}$$

We set the R, focusing on $c \in (0, 1)$:

$$R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \le c\} = \{\boldsymbol{x} : 1 \le c, \overline{X} < \theta_0\} \cup \{\boldsymbol{x} : \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} \le c, \overline{X} \ge \theta_0\} =$$

$$= \emptyset \cup \{\boldsymbol{x} : \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \le c^{2/n} = k, \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \frac{\frac{n-1}{n} S^2}{(\overline{X} - \theta_0)^2 + \frac{n-1}{n} S^2} \le k,\ \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \frac{(\overline{X} - \theta_0)^2}{S^2} + \frac{n-1}{n} \ge \frac{n-1}{nk},\ \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \left(\frac{\overline{X} - \theta_0}{S}\right)^2 \ge \frac{n-1}{k} - \frac{n-1}{n},\ \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \left(\frac{\overline{X} - \theta_0}{S/\sqrt{n}}\right)^2 \ge \frac{n-1}{k} - (n-1),\ \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \frac{\overline{X} - \theta_0}{S/\sqrt{n}} \ge \sqrt{\frac{n-1}{k} - (n-1)} = \tilde{k},\ \overline{X} \ge \theta_0\} =$$

$$= \{\boldsymbol{x} : \overline{X} \ge \theta_0 + \frac{S}{\sqrt{n}} \cdot \tilde{k}\}.$$

Now we impose the size of R equal to α:

$$
\begin{aligned}
\alpha &= \sup_{\theta \le \theta_0} \mathbb{P}\{\boldsymbol{x} \in R\} = \sup_{\theta \le \theta_0} \mathbb{P}\left\{\overline{X} \ge \theta_0 + \frac{S}{\sqrt{n}} \cdot \tilde{k}\right\} = \\
&= \sup_{\theta \le \theta_0} 1 - \mathbb{P}\left\{\frac{\overline{X} - \theta}{\sqrt{S^2/n}} \le \frac{\theta_0 + \frac{S}{\sqrt{n}} \cdot \tilde{k} - \theta}{\sqrt{S^2/n}}\right\} = \\
&= \sup_{\theta \le \theta_0} 1 - t_{n-1}\left(\frac{\theta_0 + \frac{S}{\sqrt{n}} \cdot \tilde{k} - \theta}{\sqrt{S^2/n}}\right) = \\
&= 1 - t_{n-1}(\tilde{k}) \implies t_{n-1}(\tilde{k}) = t_{n-1,1-\alpha}.
\end{aligned}
$$

Then $R_\alpha = \{\boldsymbol{x} : \overline{X} > \theta_0 + t_{n-1,1-\alpha}\sqrt{S^2/n}\}$.

Chapter 6
Uniformly Most Powerful Test

6.1 Theory Recap

Definition 6.1 (Uniformly Most Powerful (UMP) Test) Let C be a class of tests $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$. A test of the class C with power function $\beta(\theta)$ is the uniformly most powerful test, UMP, of the class C, if:

$$\beta(\theta) \geq \beta'(\theta) \qquad \forall \theta \in \Theta_0^c, \quad \forall \beta' \text{ power function associated with a test in } C.$$

Theorem 6.1 (Neyman-Pearson) *Consider the following class of tests:*

$$\begin{cases} H_0 : \theta = \theta_0; \\ H_1 : \theta = \theta_1; \end{cases}$$

where the probability density associated with $\boldsymbol{X}$ is $f(\boldsymbol{X}; \theta_i)$ with $i \in \{0, 1\}$. If we use a test whose rejection region satisfies:

$$\begin{aligned} &\boldsymbol{x} \in R \quad \textit{if} \quad f(\boldsymbol{x}; \theta_1) > k f(\boldsymbol{x}; \theta_0) \\ &\textit{and} \\ &\boldsymbol{x} \in R^c \quad \textit{if} \quad f(\boldsymbol{x}; \theta_1) < k f(\boldsymbol{x}; \theta_0) \end{aligned} \tag{6.1}$$

for some $k \geq 0$ and

$$\alpha = \mathbb{P}_{\theta_0}\{\boldsymbol{X} \in R\}. \tag{6.2}$$

Then:

- *Every test that satisfies Eq. (6.1), (6.2) is a UMP test of level α.*

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_6

Table 6.1 Karlin-Rubin Theorem (Theorem 6.2) varying the hypotheses and the MLR

TEST	MLR	R
$H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$	Non-decreasing	$R = \{T > t_0\}$
$H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$	Non-increasing	$R = \{-T > t_0\}$
$H_0 : \theta \ge \theta_0$ against $H_1 : \theta < \theta_0$	Non-decreasing	$R = \{T < t_0\}$
$H_0 : \theta \ge \theta_0$ against $H_1 : \theta < \theta_0$	Non-increasing	$R = \{-T < t_0\}$

- *If there exists a test that satisfies Eq.* (6.1), (6.2) *with* $k > 0$*, then every UMP test of level* α *is also a test of size* α *(satisfies Eq.* (6.2)*), and every UMP test of level* α *satisfies Eq.* (6.1) *except for a set* A*, which satisfies* $\mathbb{P}_{\theta_0}\{\boldsymbol{X} \in A\} = \mathbb{P}_{\theta_1}\{\boldsymbol{X} \in A\} = 0$.

Definition 6.2 (Monotone Likelihood Ratio) A family of probability densities $\{g(t;\theta) : \theta \in \Theta \subset \mathbb{R}\}$ for a r.v. T has a monotone likelihood ratio, MLR, if, $\forall\,\theta_2$ and $\forall\,\theta_1$ such that $\theta_2 > \theta_1$, $g(t;\theta_2)/g(t;\theta_1)$ is a monotone function (non-increasing or non-decreasing) of t.

Theorem 6.2 (Karlin-Rubin) *Consider the following class of tests:*

$$\begin{cases} H_0 : \theta \le \theta_0; \\ H_1 : \theta > \theta_0. \end{cases}$$

Suppose that T *is a sufficient statistic for* θ *and that the family of probability densities* $\{g(t;\theta) : \theta \in \Theta\}$ *of* T *has a non-decreasing MLR. Then for every* t_0*, the test that rejects* H_0 *if and only if* $T > t_0$ *is a UMP test of level* α*, where* $\alpha = \mathbb{P}_{\theta_0}\{T > t_0\}$*. The other cases are reported in Table 6.1.*

6.2 Exercises

Exercise 6.1 Given the family of laws:

$$f_X(x;\theta) = \frac{2}{\theta^2}\big(\theta - x\big)\, I_{(0,\theta)}(x)\,;$$

we want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, with $0 < \theta_1 < \theta_0$.

(a) Find a most powerful test of level α based on a sample of size 1.
(b) Calculate the power of the test in the previous point against θ_1.

Exercise 6.2 Find a most powerful test of level α based on a sample of size 1 to verify $H_0 : X \sim N(0,1)$ against $H_1 : X \sim C(0,1)$, i.e., X is a Cauchy variable with median 0.

Exercise 6.3 Let $X_1, \ldots, X_n$ be a random sample from a population $N(\mu, \sigma_0^2)$, with $\mu \in \mathbb{R}$ unknown and $\sigma_0^2 > 0$ known.

(a) Find a most powerful test of level α for $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$, with $\mu_1 > \mu_0$.
(b) Deduce from (a) a uniformly most powerful test of level α for $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$.

Exercise 6.4 For a variable X, consider the statistical model defined by:

$$f(x; \theta) = \theta\, x^{\theta-1}, \qquad\qquad 0 < x < 1.$$

(a) Find a Neyman-Pearson test of size α (based on a sample of size 1) for $H_0 : \theta = 1$ against $H_1 : \theta = \theta_1$, with $\theta_1 > 1$.
(b) Is the test found in (a) biased?
(c) For the statistical hypotheses of point (a), given $\theta_1 > 1$, calculate the maximum power that an arbitrary test of size α can have.
(d) Deduce from (a) a uniformly most powerful test of level α for $H_0 : \theta = 1$ against $H_1 : \theta > 1$.

Exercise 6.5 Find a most powerful test of level α based on a sample of size 1 to verify $H_0 : X \sim f_0$ against $H_1 : X \sim f_1$, where:

$$f_0(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \qquad\qquad f_1(x) = \frac{e^{-|x|}}{2}.$$

Exercise 6.6 Show that the statistical model defined by

$$f(x; \theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \qquad\qquad \theta \in \mathbb{R},$$

does not have a likelihood ratio that is monotonic in X.

Exercise 6.7 Given two natural numbers $n < N$, show that the hypergeometric statistical model $G(N, M, n)$, $0 \leq M \leq N$, has a monotonic likelihood ratio.

Exercise 6.8 Given the family of exponential laws $\mathcal{E}(\lambda)$, $\lambda > 0$, find a uniformly most powerful test of level α for $H_0 : \lambda \leq \lambda_0$ against $H_1 : \lambda > \lambda_0$ based on a sample of size n.

Exercise 6.9 Consider a random sample $X_1, \ldots, X_n$ from a population $U\big([0, \theta]\big)$, $\theta > 0$. To test $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, consider the rejection region:

$$R_\alpha = \Big\{x_{(n)} > (1 - \alpha)^{1/n}\theta_0\Big\}.$$

(a) Verify that R_α has size α and calculate its power function.
(b) Is the test given by R_α biased?
(c) Show that the model has a likelihood ratio that is monotonic with respect to $T = X_{(n)}$.
(d) Deduce from (c) that R_α defines a uniformly most powerful test of any test of level α for $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

Exercise 6.10 Let X be a sample of size one from a distribution with density:

$$f_X(x;\theta) = \frac{2}{\theta^2}(\theta - x)I_{(0,\theta)}(x)$$

with $\theta \in (0, \infty)$. Consider the problem of testing the hypotheses:

$$H_0 : \theta = 1 \quad \text{vs.} \quad H_1 : \theta > 1.$$

Given $\alpha \in (0, 1)$, construct the rejection region of the uniformly most powerful test δ_1 of level α and calculate its power function.

Exercise 6.11 Consider a single variable X described by the statistical model:

$$f_X(x;\theta) = \frac{e^{x-\theta}}{\left(1 + e^{x-\theta}\right)^2}, \qquad -\infty < x < +\infty, \qquad -\infty < \theta < +\infty.$$

Let $\alpha \in (0, 1)$.

(a) Find a most powerful test of level α for $H_0 : \theta = 0$ against $H_1 : \theta = 1$.
(b) Find a uniformly most powerful test of level α for $H_0 : \theta = 0$ against $H_1 : \theta > 0$.
(c) Show that the model has a likelihood ratio that is monotonic in X.
(d) Find a UMP test of level α for $H_0 : \theta \leq 0$ against $H_1 : \theta > 0$.
(e) Calculate the power of the test in (a) in the case $\alpha = 0.3$.

Exercise 6.12 For $n \geq 1$, let $X_1, ..., X_n$ be a random sample from a distribution having density:

$$f_X(x;\theta) = \begin{cases} \dfrac{1}{\theta} m x^{m-1} e^{-\frac{x^m}{\theta}} & \text{if } x > 0; \\ 0 & \text{otherwise;} \end{cases}$$

where m is a known natural number and θ is an unknown positive parameter. Given $\theta_0 > 0$, determine the rejection region of the level test $\alpha \in (0, 1)$ uniformly most powerful for testing the hypotheses:

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta > \theta_0.$$

6.3 Solutions

6.1

(a) To find a more powerful level α test based on a sample of size 1, we apply N-P (Theorem 6.1).
We calculate the rejection region:

$$
\begin{aligned}
R &= \{x : f(x;\theta_1) > k \cdot f(x;\theta_0)\} = \\
&= \left\{x : \frac{2}{\theta_1^2}\big(\theta_1 - x\big)\, I_{(0,\theta_1)}(x) > k \cdot \frac{2}{\theta_0^2}\big(\theta_0 - x\big)\, I_{(0,\theta_0)}(x)\right\} \\
&= \begin{cases}
\left\{x : \frac{2}{\theta_1^2}(\theta_1 - x) > k \cdot \frac{2}{\theta_0^2}\big(\theta_0 - x\big)\, I_{(0,\theta_0)}(x)\right\} & \text{if} \quad 0 < x < \theta_1; \\
\left\{x : 0 > k \cdot \frac{2}{\theta_0^2}\big(\theta_0 - x\big)\, I_{(0,\theta_0)}(x)\right\} = \emptyset & \text{if} \quad \theta_1 \le x \le \theta_0; \\
\left\{x : 0 > 0\right\} = \emptyset & \text{if} \quad x > \theta_0.
\end{cases}
\end{aligned}
$$

We consider the only non-trivial case, namely $x \in (0, \theta_1)$. Then:

$$
R = \left\{x : \frac{\big(\theta_1 - x\big)}{\big(\theta_0 - x\big)} > k \cdot \frac{\theta_1^2}{\theta_0^2} = \tilde{k}\right\}.
$$

We define $g(x) = \frac{(\theta_1 - x)}{(\theta_0 - x)}$. This is a homographic function that has $x = \theta_1$ as a vertical asymptote and $y = 1$ as a horizontal asymptote (see Fig. 6.1, first graph from the left). In the right graph of Fig. 6.1, $g(x)$ is represented in its real domain, i.e., $x \in (0, \theta_1)$, while the dashed line corresponds to a possible value of $\tilde{k}$. We can therefore conclude that the rejection region is of the form $\{X < c\}$, where $c \in (0, \theta_1)$.
We then impose that the test is of level α:

$$
\begin{aligned}
\alpha &= \mathbb{P}_{\theta_0}\{X \in R\} = \mathbb{P}_{\theta_0}\{X < c\} = \\
&= \int_0^c \frac{2}{\theta_0^2}\big(\theta_0 - x\big)\,\mathrm{d}x = \frac{2}{\theta_0}x - \frac{x^2}{\theta_0^2}\bigg|_0^c = \\
&= \frac{c}{\theta_0}\left(2 - \frac{c}{\theta_0}\right) \qquad \Rightarrow \qquad c = \theta_0(1 - \sqrt{1-\alpha}).
\end{aligned}
$$

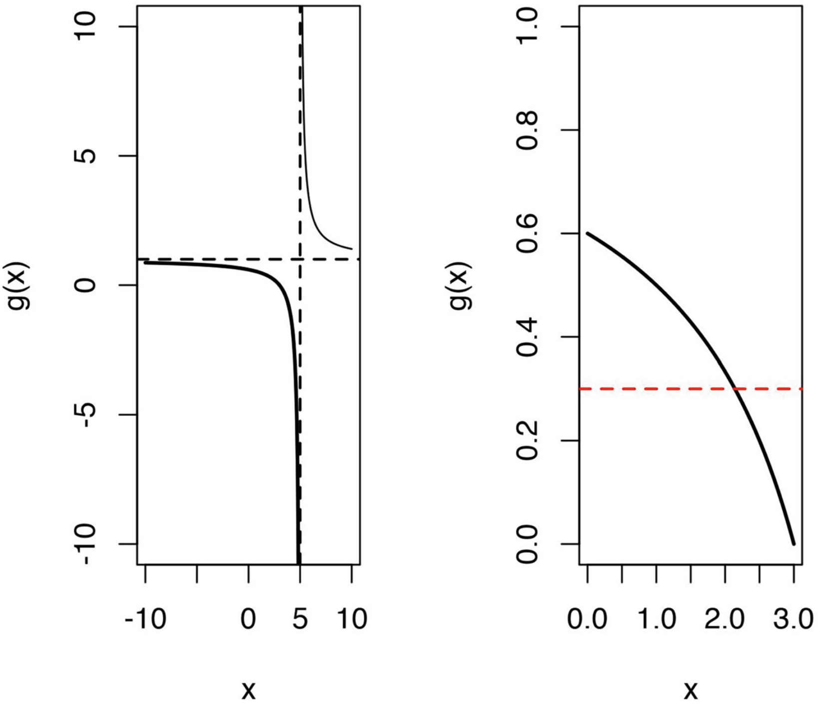

Fig. 6.1 Representation of $g(x) = \frac{(\theta_1 - x)}{(\theta_0 - x)}$. In the left graph, the function is represented on $\mathbb{R}$ and the asymptotes are highlighted. In the right graph, the function is evaluated only on $x \in (0, \theta_1)$, which is the real domain of our function. The dashed line in the right graph represents a possible value of $y = \tilde{k}$

(b)

$$
\begin{aligned}
\beta(\theta_1) &= \mathbb{P}_{\theta_1}\{X < \theta_0(1 - \sqrt{1-\alpha})\} = \\
&= \int_0^{\theta_0(1-\sqrt{1-\alpha})} \frac{2}{\theta_1^2}\big(\theta_1 - x\big)\,\mathrm{d}x = \frac{2}{\theta_1}x - \frac{x^2}{\theta_1^2}\Big|_0^{\theta_0(1-\sqrt{1-\alpha})} = \\
&= \frac{\theta_0(1-\sqrt{1-\alpha})}{\theta_1}\left(2 - \frac{\theta_0(1-\sqrt{1-\alpha})}{\theta_1}\right).
\end{aligned}
$$

6.2 In this case we can apply N-P (Theorem 6.1).

$$
R = \Big\{x : f_1(x) > k \cdot f_0(x)\Big\} \qquad k \geq 0.
$$

We calculate R:

$$R = \left\{x : f_1(x) > k \cdot f_0(x)\right\} = \left\{x : \frac{1}{\pi(1+x^2)} > \frac{1}{\sqrt{2\pi}} k e^{-x^2/2}\right\} =$$
$$= \left\{x : \frac{e^{x^2/2}}{(1+x^2)} > \frac{\pi}{\sqrt{2\pi}} k = \tilde{k}\right\}.$$

We observe that $g(x) = \frac{e^{x^2/2}}{(1+x^2)}$ is a non-negative even function, defined on all $\mathbb{R}$. Given that $g(x)$ is even, it is sufficient to study $g(t) = \frac{e^{t/2}}{(1+t)}$ with $t = x^2 \geq 0$.

$$g'(t) = \frac{e^{t/2}}{(1+t)} \frac{t-1}{2}.$$

Then $g(t)$ is increasing for $t \geq 1$, while it is decreasing for $t < 1$.

$max(g(x^2)) = 1$, reached at $x = 0$ (note: local maximum); $min(g(x^2)) = \exp\{1/2\}/2$, reached at $x = \pm 1$ (note: global minima).

We represent the function $g(x^2)$ in Fig. 6.2, highlighting three different possible $\tilde{k}$ with different lines. To define the rejection region we must distinguish based on $\tilde{k}$:

$$R = \left\{x : \frac{e^{x^2/2}}{(1+x^2)} > \frac{\pi}{\sqrt{2\pi}} k = \tilde{k}\right\} =$$
$$= \begin{cases} \mathbb{R} & \text{if} \quad \tilde{k} < \exp\{1/2\}/2; \\ \{-c_1 < x < c_1\} \cup \{x < -c_2\} \cup \{x > c_2\} & \text{if} \quad \exp\{1/2\}/2 < \tilde{k} < 1; \\ \{x < -c_2\} \cup \{x > c_2\} & \text{if} \quad \tilde{k} > 1. \end{cases}$$

We then set the significance level of the test.

The first case is trivial. Let's focus on the second and third case.

Case 2: $\exp\{1/2\}/2 < \tilde{k} < 1$.

$$\alpha = \mathbb{P}_{H_0}\{-c_1 < X < c_1\} + \mathbb{P}_{H_0}\{X < -c_2\} + \mathbb{P}_{H_0}\{X > c_2\} =$$
$$= \phi(c_1) - \phi(-c_1) + \phi(-c_2) + 1 - \phi(c_2) =$$
$$= 2\phi(c_1) - 1 + 2(1 - \phi(c_2)) =$$
$$= 2\phi(c_1) - 1 + 2 - 2\phi(c_2) =$$
$$= 1 + 2\phi(c_1) - 2\phi(c_2).$$

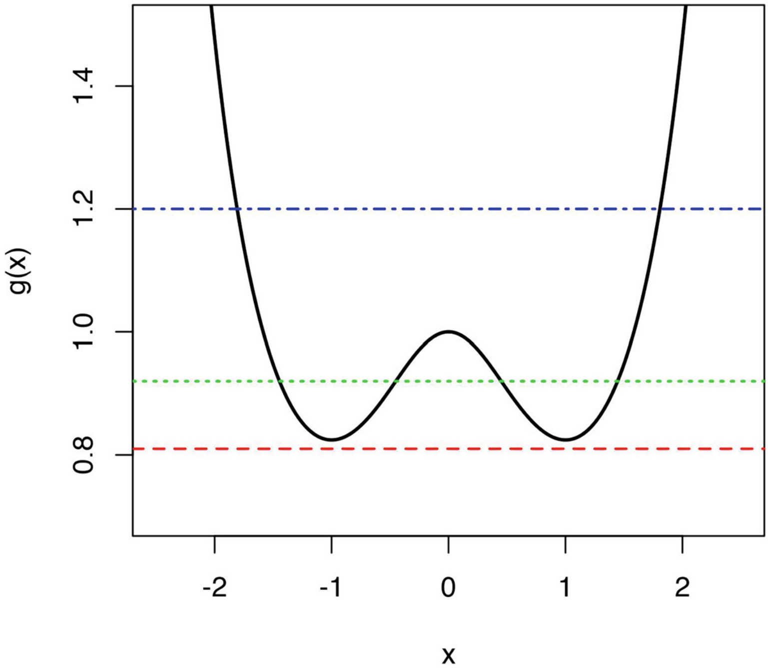

Fig. 6.2 Representation of $g(x) = \frac{e^{x^2/2}}{(1+x^2)}$. In particular: – – – – – – $\tilde{k} < \exp\{1/2\}/2$ - - - - - - - - - - - $\exp\{1/2\}/2 < \tilde{k} < 1$ — · — · — · $\tilde{k} > 1$

We can find c_1 and c_2, by numerically solving the following system:

$$\begin{cases} \alpha = 1 + 2\phi(c_1) - 2\phi(c_2); \\ g(c_1) = g(c_2). \end{cases} \tag{6.3}$$

Case 3: $\tilde{k} > 1$.

$$\alpha = \mathbb{P}_{H_0}\{X < -c_2\} + \mathbb{P}_{H_0}\{X > c_2\} = 2(1 - \phi(c_2)). \implies c_2 = z_{1-\alpha/2}.$$

6.3

(a) We apply N-P (Theorem 6.1):

$$f(\boldsymbol{x};\mu)=\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu)^2\right\}=$$
$$=\underbrace{\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2\right\}}_{h(\boldsymbol{x})}\cdot\underbrace{\exp\left\{\frac{n\overline{x}_n\mu}{\sigma_0^2}-\frac{n\mu^2}{\sigma_0^2}\right\}}_{g(t;\mu)}.$$

We apply N-P (Theorem 6.1) for sufficient statistics.
Observation

$$g(t;\mu_1)>kg(t;\mu_0) \iff \frac{nt\mu_1}{\sigma_0^2}-\frac{n\mu_1^2}{2\sigma_0^2}>\log k+\left(\frac{nt\mu_0}{\sigma_0^2}-\frac{n\mu_0^2}{2\sigma_0^2}\right)$$
$$\iff nt(\mu_1-\mu_0)>\log k'+\frac{n}{2}(\mu_1^2-\mu_0^2)$$
$$\overset{\mu_1>\mu_0}{\iff} t>\frac{2\log k'+n(\mu_1^2-\mu_0^2)}{n(\mu_1-\mu_0)}.$$

Then $R=\{\boldsymbol{x}:\overline{X}>c\}$. We then impose that the test is of level α:

$$\alpha=\mathbb{P}_{H_0}\{\boldsymbol{X}\in R\}=\mathbb{P}\left\{\frac{\overline{X}_n-\mu_0}{\sigma_0/\sqrt{n}}>\frac{c-\mu_0}{\sigma_0/\sqrt{n}}\right\} \iff \frac{c-\mu_0}{\sigma_0/\sqrt{n}}=z_{1-\alpha}.$$

We can therefore write the rejection region of the UMP test of level α as:

$$R_\alpha=\left\{\overline{\boldsymbol{X}}_n>\mu_0+\frac{\sigma_0}{\sqrt{n}}z_{1-\alpha}\right\}.$$

(b) Given that the likelihood ratio is *monotonic* and $\overline{X}$ is a sufficient statistic, then the test characterised by $R_\alpha=\left\{\overline{\boldsymbol{X}}_n>\mu_0+\frac{\sigma_0}{\sqrt{n}}z_{1-\alpha}\right\}$ is still UMP at level α.

6.4

(a) We apply N-P (Theorem 6.1). We derive the rejection region:

$$R=\{x: f(x;\theta_1)>k\cdot f(x;1)\}=\{x:\theta_1 x^{\theta_1-1}\,\mathbb{I}_{(0,1)}(x)>k\cdot\mathbb{I}_{(0,1)}(x)\}=$$
$$=\{x:x>\left(\frac{k}{\theta_1}\right)^{1/(\theta_1-1)}=\tilde{k}\}=\{x:x>\tilde{k}\}.$$

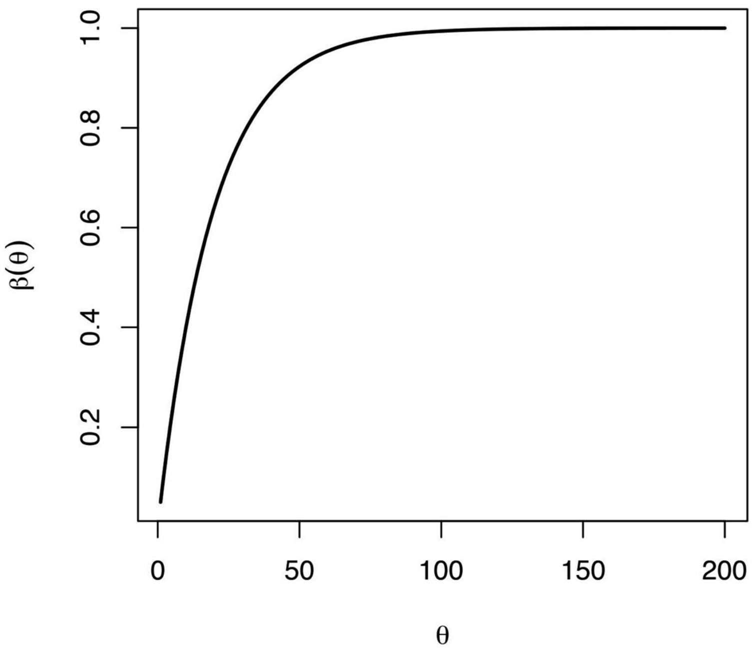

Fig. 6.3 Representation of the power function $\beta(\theta)$ associated with the test ($\theta \geq 1$)

We set the test level equal to α.

$$\alpha = \mathbb{P}_1\{X \in R\} = \mathbb{P}_1\{X > \tilde{k}\} = \int_{\tilde{k}}^{1} \mathrm{dx} = 1 - \tilde{k} \implies \tilde{k} = 1 - \alpha.$$

So: $R_\alpha = \{X > 1 - \alpha\}$.

(b) To answer the question we calculate the power function:

$$\beta(\theta) = \mathbb{P}_\theta\{X > 1 - \alpha\} = \int_{1-\alpha}^{1} \theta\, x^{\theta-1}\, \mathrm{dx} = x^\theta \Big|_{1-\theta}^{1} = 1 - (1-\alpha)^\theta.$$

We immediately notice that the power function is monotonically increasing therefore it satisfies the requirement to be an *unbiased* test:

$$\beta(\theta') \geq \beta(\theta'') \qquad \forall \theta' > 1,\ \theta'' = 1.$$

See Fig. 6.3.

(c) Since N-P (Theorem 6.1) guarantees us to have found a UMP test, by definition of UMP, we can affirm that:

$$\beta'(\theta) \leq 1 - (1-\alpha)^{\theta_1}.$$

β' is the power function associated with a generic test of the same class as the considered test (level α test).

(d) Since R_α does not depend on θ_1, we can affirm that: the level α test, associated with the test $H_0 : \theta = 1$ against $H_1 : \theta > 1$, is UMP at level α.

6.5

$$\begin{cases} H_0 : X \sim f_0 & f_0(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}; \\ H_1 : X \sim f_1 & f_1(x) = \frac{e^{-|x|}}{\sqrt{2}}. \end{cases}$$

We apply N-P (Theorem 6.1):

$$\begin{aligned} R = \{f_1 > kf_0\} &= \left\{ \frac{e^{-|x|}}{\sqrt{2}} > k \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right\} = \\ &= \left\{ \frac{e^{-|x|}}{\sqrt{2}} \cdot \frac{\sqrt{2\pi}}{e^{-x^2/2}} > k \right\} = \\ &= \left\{ \frac{e^{-|x|}}{e^{-x^2/2}} > c \right\} = \\ &= \left\{ e^{x^2/2-|x|} > c \right\} = \\ &= \left\{ x^2/2 - |x| > c' \right\}. \end{aligned}$$

The function $g(x) = x^2/2 - |x|$ is even, and exists on all $\mathbb{R}$. In Fig. 6.4 $g(x)$ is represented and the lines relative to different possible values of c'.

We therefore have to distinguish 4 cases:

CASE 1 $c' < -0.5$ $\Rightarrow$ $R = \mathbb{R}$;

CASE 2 $c' = -0.5$ $\Rightarrow$ $R = \mathbb{R} \setminus \{\pm 1\}$;

CASE 3 $-0.5 < c' < 0$ $\Rightarrow$ $R = \{|X| \leq x_1\} \cup \{|X| > x_2\}$

$d0 \leq x_1 \leq 1 < x_2 < 2$;

CASE 4 $c' \geq 0$ $\Rightarrow$ $R = \{|X| > x_3\}$ $x_3 > 2$.

CASE 1 $\alpha = 1$;

CASE 2 $\alpha = 1$;

CASE 3 $\alpha = 1 - 2[\phi(x_2) - \phi(x_1)]$;

CASE 4 $\alpha = 2(1 - \phi(x_3))$.

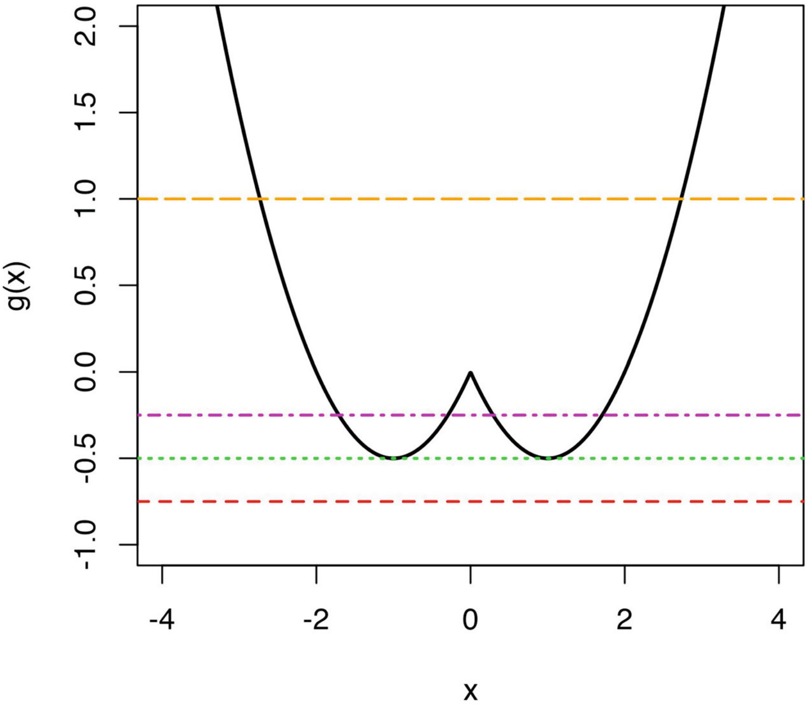

Fig. 6.4 Representation of $g(x) = x^2/2 - |x|$ and different values of c'. In particular: red dashed line $c' < -0.5$, green dotted line $c' = -0.5$, magenta dash-dotted line $-0.5 < c' < 0$, orange long-dashed line $c' \geq 0$

6.6 Let $\theta_1 < \theta_2$.

$$\frac{f(x;\theta_1)}{f(x;\theta_2)} = \frac{1 + (x-\theta_2)^2}{1 + (x-\theta_1)^2}.$$

$$\lim_{x\to\pm\infty} \frac{1 + (x-\theta_2)^2}{1 + (x-\theta_1)^2} = 1.$$

Therefore it cannot be monotonic in x.

6.7

$$n < N, \qquad G(N, M, n), \qquad 0 \leq M \leq N.$$

$$X \sim G(N, M, n) \qquad f(x; M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}.$$

Consider M and $M+1$ and calculate the likelihood ratio:

$$\frac{f(x; M+1)}{f(x; M)} = \frac{\binom{M+1}{x}\binom{N-M-1}{n-x}}{\binom{M}{x}\binom{N-M}{n-x}} = \frac{M+1}{N-M} \cdot \frac{N-M-n+x}{M+1-x};$$

which is increasing in x.

6.8 We apply the K-R Theorem (Theorem 6.2).

(a) We derive a sufficient statistic for θ. We write the joint density:

$$f_{\boldsymbol{x}}(\boldsymbol{x}; \theta) = \lambda^n e^{-\lambda \sum x_i}.$$

We immediately notice that the density $\mathcal{E}$ belongs to the exponential family, therefore:

$$T(X) = \sum X_i$$

is a sufficient statistic (it can be seen immediately that it is also complete because $\lambda \in (0, +\infty)$, which contains an open set of $\mathbb{R}$).

(b) We derive the law of T: $T(X) \sim g(t; \lambda)$.

$$X_i \sim \mathcal{E}(\lambda) \stackrel{d}{=} \Gamma(1, \lambda) \qquad \Longrightarrow \qquad T = \sum X_i \sim \Gamma(n, \lambda).$$

(c) We verify that T has MLR (Monotone Likelihood Ratio).
Let $\lambda_2 > \lambda_1$.

$$\frac{g(t; \lambda_2)}{g(t; \lambda_1)} = \frac{1/\Gamma(n)e^{-\lambda_2 t} \cdot \lambda_2^2 \cdot t^{n-1}}{1/\Gamma(n)e^{-\lambda_1 t} \cdot \lambda_1^2 \cdot t^{n-1}} = e^{(\lambda_1-\lambda_2)t} \cdot \left(\frac{\lambda_2}{\lambda_1}\right)^n.$$

The likelihood ratio is *monotonically decreasing*.

Given the assumptions for the application of K-R (Theorem 6.2), we can say that:

$$R = \{-T > t_0\} = \{T < -t_0 = \tilde{t}_0\}.$$

We set the test level:

$$\alpha = \sup_{0<\lambda\leq\lambda_0} \mathbb{P}\{T < \tilde{t}_0\} = \mathbb{P}_{\lambda_0}\{T < \tilde{t}_0\} \qquad \Longrightarrow \qquad \tilde{t}_0 = \gamma^{\alpha}_{n,\lambda_0}.$$

6.9

(a) We calculate the law of $X_{(n)}$:

$$\mathbb{P}\{X_{(n)} \le t\} = (\mathbb{P}\{X_i \le t\})^n = \left(\frac{t}{\theta}\right)^n \mathbb{I}_{(0,\theta)}(t) = \left(\frac{t}{\theta}\right)^n \mathbb{I}_{(t,+\infty)}(\theta).$$

Then we calculate α:

$$\begin{aligned}\alpha &= \sup_{\theta\le\theta_0} \mathbb{P}\{\mathbf{X} \in R_\alpha\} = \sup_{\theta\le\theta_0} \mathbb{P}\left\{X_{(n)} > (1-\alpha)^{1/n}\theta_0\right\} = \\ &= \sup_{\theta\le\theta_0} 1 - \mathbb{P}\left\{X_{(n)} \le (1-\alpha)^{1/n}\theta_0\right\} = \\ &= \sup_{\theta\le\theta_0} \left[1 - \left(\frac{1}{\theta}(1-\alpha)^{1/n}\theta_0\right)^n\right] = \\ &= [1-(1-\alpha)] = \alpha.\end{aligned}$$

The *sup* is reached for $\theta = \theta_0$. We then calculate the power function $\beta(\theta)$:

$$\begin{aligned}\beta(\theta) &= \mathbb{P}_\theta\{\mathbf{X} \in R_\alpha\} = \mathbb{P}\left\{X_{(n)} > (1-\alpha)^{1/n}\theta_0\right\} = \\ &= 1 - \mathbb{P}\left\{X_{(n)} \le (1-\alpha)^{1/n}\theta_0\right\} = \\ &= 1 - \frac{[(1-\alpha)^{1/n}\theta_0]^n}{\theta^n}\mathbb{I}_{(0,\theta)}((1-\alpha)^{1/n}\theta_0).\end{aligned}$$

(b) We immediately notice that the power function is monotonically increasing therefore it satisfies the requirement to be an *undistorted* test:

$$\beta(\theta') \ge \beta(\theta'') \qquad \forall\theta' > \theta_0,\ \theta'' \le \theta_0.$$

(c) From point (a) we immediately derive that:

$$f_{X_{(n)}}(t) = n\frac{t^{(n-1)}}{\theta^n}\mathbb{I}_{(0,\theta)}(t).$$

Let's calculate the MLR, considering $\theta_2 > \theta_1$:

$$\frac{g(t;\theta_2)}{g(t;\theta_1)} = \frac{n\frac{t^{(n-1)}}{\theta_2^n}\mathbb{I}_{(0,\theta_2)}(t)}{n\frac{t^{(n-1)}}{\theta_1^n}\mathbb{I}_{(0,\theta_1)}(t)} = \frac{\theta_1^n}{\theta_2^n}\frac{\mathbb{I}_{(0,\theta_2)}(t)}{\mathbb{I}_{(0,\theta_1)}(t)} = \begin{cases}\left(\frac{\theta_1}{\theta_2}\right)^n, & \text{if } t \le \theta_1;\\ +\infty, & \text{if } t > \theta_1.\end{cases}$$

We immediately notice that the MLR is *monotonically increasing*.

(d) Let's consider the following test: $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. Since the test is not simple, to find the UMP of level α, we try to apply K-R (Theorem 6.2). We know that $T = X_{(n)}$ is a sufficient statistic for θ and that it has a *monotonically increasing* MLR. Given the assumptions for the application of K-R (Theorem 6.2), we can say that the UMP test has the following rejection region:

$$R = \{X_{(n)} > t_0\}.$$

Furthermore, from the previous points we can conclude that:

$$R_\alpha = \left\{X_{(n)} > (1-\alpha)^{1/n}\theta_0\right\}.$$

6.10 To answer the question we set up the following test, in order to apply N-P (Theorem 6.1).

$$H_0 : \theta = 1 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \quad \theta_1 > 1.$$

We define the rejection region:

$$\begin{aligned}
R &= \left\{x : f(x;\theta_1) > k \cdot f(x;1)\right\} \\
&= \left\{x : \frac{2}{\theta_1^2}(\theta_1 - x)I_{(0,\theta_1)}(x) > k \cdot 2(1-x)I_{(0,1)}(x)\right\} = \\
&= \begin{cases} \left\{x : \frac{2}{\theta_1^2}(\theta_1 - x) > k \cdot 2(1-x)\right\} = \{x : \frac{\theta - x}{1-x} > k\theta^2\} & \text{if} \quad 0 < x \leq 1; \\ \left\{x : \frac{2}{\theta_1^2}(\theta_1 - x) > 0\right\} = \{1 < x \leq \theta_1\} & \text{if} \quad 1 < x \leq \theta_1; \\ \left\{x : 0 > 0\right\} = \emptyset & \text{if} \quad x > \theta_1. \end{cases}
\end{aligned}$$

We focus on the only non-trivial case and notice that $g(x) = \frac{\theta - x}{1-x}$ is monotonically increasing and has as codomain $[\theta_1, +\infty]$. Therefore:

$$R = \{X > c\}.$$

We impose that the level of the test is α.

$$\begin{aligned}
\alpha &= \sup_{\theta \in \Theta_0} \mathbb{P}\{X > c\} = \mathbb{P}_1\{X > c\} = \\
&= \int_c^1 2(1-x)\,dx = 2x - x^2\Big|_c^1 = \\
&= 2 - 1 - (2c - c^2) = (1-c)^2.
\end{aligned}$$

We can conclude that: $c = 1 - \sqrt{\alpha}$.

$$R_\alpha = \{X > 1 - \sqrt{\alpha}\}.$$

N-P (Theorem 6.1) guarantees that this test is UMP of level α.

We evaluate the power function:

$$\begin{aligned}\beta(\theta) &= \mathbb{P}_\theta\left\{X > 1-\sqrt{\alpha}\right\} = \int_{1-\sqrt{\alpha}}^{\theta} \frac{2}{\theta^2}(\theta - x)dx = \frac{2}{\theta}x - \frac{x^2}{\theta^2}\bigg|_{1-\sqrt{\alpha}}^{\theta} = \\ &= 2 - 1 - \frac{2}{\theta}(1-\sqrt{\alpha}) + \frac{1}{\theta^2}(1-\sqrt{\alpha})^2 = \\ &= 1 - \frac{2}{\theta}(1-\sqrt{\alpha}) + \frac{1}{\theta^2}(1-\sqrt{\alpha})^2 = \\ &= \left(1 - \frac{(1-\sqrt{\alpha})}{\theta}\right)^2.\end{aligned}$$

This is valid if $1 - \sqrt{\alpha} < \theta$, otherwise $\beta(\theta) = 0$.

6.11

(a)

$$\begin{cases} H_0 : \theta = 0; \\ H_1 : \theta = 1. \end{cases}$$

We apply N-P (Theorem 6.1):

$$\begin{aligned}R = \{x : f(x;1) > kf(x;0)\} &= \left\{\frac{e^{x-1}}{(1+e^{x-1})^2} > k\frac{e^x}{(1+e^x)^2}\right\} = \\ &= \left\{x : \frac{(1+e^x)^2}{(1+e^{x-1})^2} > k\frac{e^x}{e^{x-1}} = k'\right\} = \\ &= \left\{x : \frac{1+e^x}{1+e^{x-1}} > \tilde{k}\right\} = \\ &= \left\{x : 1 + e^x > \tilde{k}(1+e^{x-1})\right\} = \\ &= \left\{x : e^x(1-\tilde{k}) > \tilde{k} - 1 + \frac{\tilde{k}}{c}\right\}.\end{aligned}$$

So the rejection region is of the form: $\{X > \gamma\}$.

We therefore impose that the test level is equal to α:

$$\alpha = \mathbb{P}_0\{X > \gamma\} = \int_{\gamma}^{+\infty} \frac{e^x}{(1+e^x)^2}\,\mathrm{d}x = -\frac{1}{1+e^x}\bigg|_{\gamma}^{+\infty} = \frac{1}{1+e^{\gamma}} = \alpha.$$

Hence:

$$R_\alpha = \left\{X > \log\left(\frac{1-\alpha}{\alpha}\right)\right\}$$

is a UMP test of level α.

(b) Given that R, in the case of simple hypotheses, does not depend on H_1:

$$R = \left\{x : X > \log\left(\frac{1-\alpha}{\alpha}\right)\right\}.$$

(c) Let $\theta_2 > \theta_1$:

$$\frac{f(x;\theta_2)}{f(x;\theta_1)} = e^{x-\theta_2}e^{-x+\theta_1}\left(\frac{1+e^{x-\theta_1}}{1+e^{x-\theta_2}}\right)^2.$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1+e^{x-\theta_1}}{1+e^{x-\theta_2}}\right) = \frac{e^{x-\theta_1}(1+e^{x-\theta_2})-(1-e^{x-\theta_1})e^{x-\theta_2}}{(1+e^{x-\theta_2})^2} = \frac{e^{x-\theta_1}-e^{x-\theta_2}}{(1+e^{x-\theta_2})^2} > 0.$$

The likelihood ratio is *monotonically increasing* in x.

(d) UMP test of level α for the following hypotheses:

$$\begin{cases} H_0 : \theta \le 0; \\ H_1 : \theta > 0. \end{cases}$$

is of the form $R = \{X > k\}$, according to K-R (Theorem 6.2) We therefore impose that $\mathbb{P}_0\{X \in R\} = \alpha$. Then:

$$R_\alpha = \left\{X > \log\left(\frac{1-\alpha}{\alpha}\right)\right\}.$$

(e)

$$\alpha = 0.3 \qquad \Rightarrow \qquad \left\{X > \log\left(\frac{0.7}{0.3}\right) = 0.8473\right\}.$$

$$\beta_1(\theta) = \mathbb{P}_1\{X > 0.8473\} = \int_{0.8473}^{+\infty} \frac{e^{x-1}}{(1+e^{x-1})^2}\,\mathrm{d}x =$$

$$= -\frac{1}{1+e^{x-1}}\Bigg|_{0.8473}^{+\infty} = \frac{1}{1+e^{0.8473-1}} \simeq 0.5381.$$

6.12 I proceed using N-P (Theorem 6.1) on the following test:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1 \qquad \theta_1 > \theta_0.$$

If the rejection region does not depend on θ_1, then we can say we have found the UMP test of level α.

$$\begin{aligned}
R &= \Big\{x : f(x;\theta_1) > k \cdot f(x;\theta_0)\Big\} = \\
&= \Big\{x : \frac{1}{\theta_1} m x^{m-1} e^{-\frac{x^m}{\theta_1}} \mathbb{I}_{(0,+\infty)}(x) > k \cdot \frac{1}{\theta_0} m x^{m-1} e^{-\frac{x^m}{\theta_0}} \mathbb{I}_{(0,+\infty)}(x)\Big\} = \\
&= \Big\{x : e^{\frac{x^m}{\theta_0} - \frac{x^m}{\theta_1}} > k \cdot \frac{\theta_1}{\theta_0}\Big\} = \\
&= \Big\{x : x^m \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) > \log\left(k \cdot \frac{\theta_1}{\theta_0}\right)\Big\} = \\
&= \Big\{x : x > \left[\log\left(k \cdot \frac{\theta_1}{\theta_0}\right) \frac{\theta_0\theta_1}{\theta_1 - \theta_0}\right]^{1/m} = c\Big\}.
\end{aligned}$$

So the rejection region is of the form: $R = \{x : x > c\}$.
We therefore impose that it is of level α:

$$\begin{aligned}
\alpha &= \sup_{\theta=\theta_0} = \mathbb{P}\{X > c\} = 1 - \mathbb{P}\{X \le c\} = \\
&= 1 - \int_0^c \frac{1}{\theta_0} m x^{m-1} e^{-\frac{x^m}{\theta_0}}\, \mathrm{dx} = 1 + e^{-\frac{x^m}{\theta_0}}\Bigg|_0^c = 1 + e^{-\frac{c^m}{\theta_0}} - 1 = e^{-\frac{c^m}{\theta_0}}.
\end{aligned}$$

Then:

$$c = (-\theta_0 \log \alpha)^{1/m}.$$

So $R_\alpha = \{X > (-\theta_0 \log \alpha)^{1/m}\}$. Given that it does not depend on θ_1, we conclude that this rejection region is also the rejection region of level α of the UMP test:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$

N.B. The exercise can also be carried out using $T = \sum_i X_i$ as a sufficient statistic.

Chapter 7
Confidence Intervals

7.1 Theory Recap

Definition 7.1 (Interval Estimation) The interval estimation of a real parameter θ is constituted by any pair of statistics $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$ of the sample $\boldsymbol{X}$ that satisfy $L(\boldsymbol{X}) \leq U(\boldsymbol{X})$. The random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is said to be the interval estimate, or confidence interval, for θ.

Definition 7.2 (Coverage Probability) The coverage probability of an interval estimate $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ for θ is defined as:

$$\mathbb{P}_\theta(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]).$$

Definition 7.3 (Confidence Level) The confidence level of an interval estimate $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ for θ is defined as:

$$\inf_\theta \mathbb{P}_\theta(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]).$$

Theorem 7.1 (Confidence Interval and Acceptance Region) *For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of level α of the test $H_0 : \theta = \theta_0$. For each $\boldsymbol{x} \in \mathcal{X}$, define an interval $IC(\boldsymbol{x})$ as:*

$$IC(\boldsymbol{x}) = \{\theta_0 : \boldsymbol{x} \in A(\theta_0)\}.$$

Then the random interval $IC(\boldsymbol{X})$ is a confidence interval of level $1-\alpha$. Alternatively, let $IC(\boldsymbol{X})$ be a confidence interval of level $1-\alpha$. For every $\theta_0 \in \Theta$, define:

$$A(\theta_0) = \{\boldsymbol{x} : \theta_0 \in C(\boldsymbol{x})\}.$$

Then $A(\theta_0)$ is the acceptance region of level α associated with the test $H_0 : \theta = \theta_0$.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_7

Definition 7.4 (Pivotal Quantity) A random variable $Q(\boldsymbol{X};\theta)$ is a pivotal quantity (or pivot) if the distribution of $Q(\boldsymbol{X};\theta)$ does not depend on θ.

Theorem 7.2 (Pivoting of the Distribution Function) *Let T be a statistic with a continuous distribution function $F_T(t;\theta)$. Let α_1 and α_2 be two fixed values, such that $\alpha_1+\alpha_2=\alpha$ and $\alpha\in(0,1)$. Suppose that $\forall t\in\mathcal{T}$, the functions $\theta_L(t)$ and $\theta_U(t)$ can be defined as follows:*

- *If $F_T(t;\theta)$ is a decreasing function of θ $\forall t$, we define $\theta_L(t)$ and $\theta_U(t)$ as:*

$$F_T(t;\theta_U(t))=\alpha_1,\qquad F_T(t;\theta_L(t))=1-\alpha_2.$$

- *If $F_T(t;\theta)$ is an increasing function of θ $\forall t$, we define $\theta_L(t)$ and $\theta_U(t)$ as:*

$$F_T(t;\theta_U(t))=1-\alpha_2,\qquad F_T(t;\theta_L(t))=\alpha_1.$$

Then the random interval $[\theta_L(t),\theta_U(t)]$ is a confidence interval of level $1-\alpha$ for θ.

Theorem 7.3 (Minimum Length and Unimodality of the Density) *Let $f_X(x)$ be a unimodal probability density. If the interval $[a,b]$ satisfies the following characteristics:*

- *$\int_a^b f_X(x)\,dx=1-\alpha$;*
- *$f(a)=f(b)>0$;*
- *$a\le x*\le b$, where $x*$ is the mode of $f_X(x)$.*

Then $[a,b]$ is the interval of minimum length among those that satisfy the first condition.

7.2 Exercises

Exercise 7.1 Consider the statistical model given by the exponential laws $\mathcal{E}(\nu)$, $\nu>0$, and let $X_1,\ldots,X_n$ be a random sample drawn from a population described by this model. Find the confidence intervals for ν of level $\gamma=1-\alpha$ constructed based on:

(a) LRT for $\nu=\nu_0$ against $\nu\neq\nu_0$.
(b) Pivotal quantity $Q=2\nu\sum_{i=1}^n X_i$.

Exercise 7.2 For a sample of size 1 from the law:

$$f(x;\theta)=\frac{2}{\theta^2}(\theta-x),\qquad 0<x<\theta;$$

find the confidence intervals for θ of level $\gamma = 1 - \alpha$ constructed through:

(a) LRT for $\theta = \theta_0$ against $\theta \neq \theta_0$.
(b) Pivotal quantity $F_\theta(X)$.
(c) Pivotal quantity X/θ, choosing that of the type $\big(x, f(x)\big)$.
(d) Which interval would you choose for an interval estimation of θ at a level $\gamma = 0.95$?

Exercise 7.3 Let $X_1, \ldots, X_n$ be a random sample from a population $N(\mu, \sigma^2)$. Find a confidence interval for σ^2 of level $\gamma = 1 - \alpha$ in the cases:

(a) μ known.
(b) μ unknown.

Exercise 7.4 Consider the random samples $X_1, \ldots, X_n$ from a population $N(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_m$ from a population $N(\mu_2, \sigma^2)$. Find a confidence interval for $\mu_1 - \mu_2$ at level $\gamma = 1 - \alpha$ in the cases:

(a) σ^2 known.
(b) σ^2 unknown.

Exercise 7.5 Let X be a single observation from a Beta(θ,1):

$$f(x; \theta) = \theta\, x^{\theta - 1}\, I_{(0,1)}(x), \qquad \theta > 0.$$

(a) Find the law of $Y = -\dfrac{1}{\log X}$ and calculate the confidence level of the interval $\left(\dfrac{Y}{2}, Y\right)$ for θ.
(b) Show that X^θ is a pivotal quantity and use it to construct a confidence interval for θ at an arbitrary level $1 - \alpha$, $\alpha \in (0, 1)$, choosing the one with the smallest width.
(c) Compare the interval $\left(\dfrac{Y}{2}, Y\right)$ with the interval found in (b) at the same level.

Exercise 7.6 Consider the statistical model:

$$f(x; \mu) = \frac{1}{2}\, \mathrm{e}^{-|x-\mu|}, \qquad -\infty < x < +\infty, \quad -\infty < \mu < +\infty.$$

Given X a sample of size 1:

(a) Verify that the quantity $Q = X - \mu$ is pivotal.
(b) Determine for μ the confidence interval based on Q at level $1 - \alpha$ and of minimum length.

7.3 Solutions

7.1

(a) Comparing with the result of Exercise 5.3, the rejection region of level α of the test:

$$H_0 : \nu = \nu_0 \qquad \text{vs} \qquad H_1 : \nu \neq \nu_0;$$

turns out to be:

$$R = \left\{\overline{X}_n < \bar{t}_1/\nu_0\right\} \cup \left\{\overline{X}_n > \bar{t}_2/\nu_0\right\}$$
$$\Downarrow$$
$$\left\{\frac{\bar{t}_1}{\nu} \leq \overline{X}_n \leq \frac{\bar{t}_2}{\nu}\right\} = \left\{\frac{\bar{t}_1}{\overline{X}_n} \leq \nu \leq \frac{\bar{t}_2}{\overline{X}_n}\right\}.$$

Choosing $\bar{t}_1 = \gamma_{\alpha/2}(n, n)$ and $\bar{t}_2 = \gamma_{1-\alpha/2}(n, n)$ we get:

$$IC(1-\alpha) = \left[\frac{\gamma_{\alpha/2}(n, n)}{\overline{X}_n}; \frac{\gamma_{1-\alpha/2}(n, n)}{\overline{X}_n}\right].$$

(b) Consider the r.v. $Q = 2\nu \sum X_i$ and observe that:

$$\sum X_i \sim \Gamma(n, \nu) \quad \Rightarrow \quad 2\nu \sum X_i \sim \Gamma(n, 1/2) \overset{d}{=} \chi^2(2n).$$

We then define the IC based on Q, with confidence level $1 - \alpha$:

$$\begin{cases} IC & = \left[a \leq 2\nu \sum X_i \leq b\right]; \\ \mathbb{P}\{a \leq Q \leq b\} & = 1 - \alpha. \end{cases}$$

Choosing $a = \chi^2_{\alpha/2}(2n)$ and $b = \chi^2_{1-\alpha/2}(2n)$, we obtain:

$$IC_{(1-\alpha)} = \left[\frac{\chi^2_{\alpha/2}(2n)}{2\sum X_i} \leq \nu \leq \frac{\chi^2_{1-\alpha/2}(2n)}{2\sum X_i}\right].$$

Observation Remember that $\gamma_\alpha(n, n) = \frac{\chi^2(2n)}{2n}$.
In fact, let $X \sim \Gamma(n, n)$, then:

$$\alpha = \mathbb{P}\left\{X \leq \gamma_\alpha(n, n)\right\} = \mathbb{P}\left\{2nX \leq 2n\gamma_\alpha(n, n)\right\}$$

and $n2X \sim \Gamma(n, 1/2) \overset{d}{=} \chi^2(2n)$.

Therefore $\gamma_\alpha(n,n) = \frac{\chi^2(2n)}{2n}$ and

$$IC(1-\alpha) = \left[\frac{n\gamma_{\alpha/2}(n,n)}{\sum X_i}; \frac{n\gamma_{1-\alpha/2}(n,n)}{\sum X_i}\right] = \left[\frac{\chi^2_{\alpha/2}(2n)}{2\sum X_i}; \frac{\chi^2_{1-\alpha/2}(2n)}{2\sum X_i}\right].$$

We conclude that the two ICs obtained in point (a) and point (b) coincide.

7.2

(a) Observe that on a sample of size 1 we have:

$$L(\theta; x) = \frac{2}{\theta^2}(\theta - x)\mathbb{I}_{[x;+\infty)}(\theta).$$

$$\frac{\partial L}{\partial \theta} = -\frac{2}{\theta^2} + \frac{4x}{\theta^3} = 0 \quad \Rightarrow \quad \hat{\theta}_{MLE} = 2X.$$

From here, the LRT for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$:

$$\lambda(x) = \frac{\frac{2}{\theta_0^2}(\theta_0 - x)\mathbb{I}_{[0;\theta_0]}(x)}{\frac{2}{4x^2}(2x - x)} = \frac{4x}{\theta_0^2}(\theta_0 - x)\mathbb{I}_{[0,\theta_0]}(x).$$

$$R = \{\lambda(x) \le c\} = \{x > \theta_0\} \cup \left\{\frac{4x}{\theta_0^2}(\theta_0 - x) < c\right\} =$$

$$= \{4x(\theta_0 - x) < k'\} = \{4x\theta_0 - 4x^2 < k'\} =$$

$$= \{x^2 - \theta_0 x + h > 0\} \qquad \Rightarrow \qquad x = \frac{\theta_0 \pm \sqrt{\theta_0^2 - 4h}}{2} = \frac{\theta_0}{2} \pm \tilde{h}.$$

$$\alpha = 1 - \int_{\frac{\theta_0}{2}-\tilde{h}}^{\frac{\theta_0}{2}+\tilde{h}} \frac{2}{\theta_0^2}(\theta_0 - x)\,\mathrm{d}x =$$

$$= 1 - \frac{1}{\theta_0^2}\left[-(\theta_0 - x)^2\right]_{\frac{\theta_0}{2}-\tilde{h}}^{\frac{\theta_0}{2}+\tilde{h}} =$$

$$= 1 - \frac{1}{\theta_0^2}[2\theta_0\tilde{h}] = 1 - \frac{2\tilde{h}}{\theta_0} \qquad \Rightarrow \qquad \tilde{h} = \frac{\theta_0}{2}(1-\alpha).$$

$$R_\alpha = \left\{X < \frac{1}{2}\theta_0\alpha\right\} \cup \left\{X > \theta_0\left(1 - \frac{\alpha}{2}\right)\right\}.$$

From which:

$$IC_{(1-\alpha)} = \left[\frac{2X}{2-\alpha} \le \theta \le \frac{2X}{\alpha}\right].$$

(b) $Q = F_X(x;\theta) \sim U_{[0,1]}$:

$$F_X(x) = \begin{cases} 0 & x \le 0; \\ \frac{2}{\theta^2}\left(\theta x - \frac{x^2}{2}\right) & 0 < x \le \theta; \\ 1 & x > \theta. \end{cases}$$

Then $F_X(X) = \frac{2}{\theta^2}\left(\theta X - \frac{X^2}{2}\right)$.
We observe that:

$$g(\theta) = \frac{2}{\theta^2}\left(\theta x - \frac{x^2}{2}\right).$$

$$g'(\theta) = \frac{2\theta^2 x - 2\theta(\theta x - x^2/2)}{\theta^4} = \frac{2x(x-\theta)}{\theta^3} < 0.$$

$$g(\theta) = k \iff \frac{2}{\theta^2}\left(\theta x - \frac{x^2}{2}\right) = k$$

$$\iff \frac{x}{\theta^2}(2\theta - x) = k \iff \theta = \frac{x + \sqrt{x^2(1-k)}}{k}.$$

Therefore:

$$\begin{aligned} IC_{(1-\alpha)} &= \left[\frac{\alpha}{2} \le F_\theta(x) \le 1 - \frac{\alpha}{2}\right] \\ &= \left[X\left(\frac{1+\sqrt{\alpha/2}}{1-\alpha/2}\right); X\left(\frac{1+\sqrt{1-\alpha/2}}{\alpha/2}\right)\right]. \end{aligned}$$

(c) We consider the pivotal quantity $Q = \frac{X}{\theta}$.

$$F_Q(q) = \mathbb{P}\left\{\frac{X}{\theta} \le q\right\} = \mathbb{P}\{X \le \theta q\} = \frac{2\theta^2 q}{\theta^2} - \frac{\theta^2 q^2}{\theta^2} = q(2-q).$$

Since we are looking for an IC of the type $[X; f(X)]$ of level $1-\alpha$, we impose that:

$$\mathbb{P}\{c < Q < 1\} = 1-\alpha.$$

$$2q - q^2\Big|_c^1 = 1-\alpha \iff 1-2c+c^2 = (c-1)^2 = 1-\alpha.$$

$$c = 1-\sqrt{1-\alpha}.$$

Therefore, we conclude that:

$$IC_{(1-\alpha)} = \left[X \leq \theta \leq \frac{X}{1-\sqrt{1-\alpha}}\right].$$

(d) $1-\alpha = 0.95$. Substituting this value into the ICs calculated in the previous points, we obtain:

$$\begin{cases} \text{(a)} \quad L = 38.97X; \\ \text{(b)} \quad L = 78.31X; \\ \text{(c)} \quad L = 38.49X. \end{cases}$$

We therefore choose the IC found in point (c).

7.3

(a) If μ is known:

$$T := \frac{\sum(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n).$$

T is a pivotal quantity. We are therefore looking for an IC with a confidence level of $1-\alpha$:

$$\begin{cases} IC & = [a < T < b]; \\ \mathbb{P}\{a < T < b\} & = 1-\alpha. \end{cases}$$

We can choose $a = \chi^2_{\alpha/2}(n)$ and $b = \chi^2_{1-\alpha/2}(n)$ so that:

$$IC_{(1-\alpha)} = \left[\frac{\sum(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}; \frac{\sum(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}\right].$$

(b) If μ is unknown:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2(n-1).$$

So, similarly to the previous point, we obtain:

$$IC_{(1-\alpha)} = \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}; \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right].$$

7.4

(a) If σ^2 is known, we observe that:

$$\overline{X}_n - \overline{Y}_n \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

By standardising, we obtain:

$$Q = \frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1).$$

Q is a pivotal quantity. Therefore:

$$IC(1-\alpha) = \left[\overline{X}_n - \overline{Y}_n \pm z_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right].$$

(b) If σ^2 is unknown, we recall the definition of S_p^2 (pooled estimator of the variance):

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{(n+m-2)}.$$

$$Q = \frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2).$$

Q is a pivotal quantity. Therefore:

$$IC(1-\alpha) = \left[\overline{X}_n - \overline{Y}_n \pm t_{1-\frac{\alpha}{2}}(n+m-2)S_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right].$$

7.5

(a)

$$Y = -\frac{1}{\log X} = g(X).$$

$$-\frac{1}{Y} = \log X.$$

$$X = e^{-\frac{1}{Y}}.$$

$$g_Y(y) = f_X(g^{-1}(y))|g^{-1}(y)|.$$

$$g_Y(y) = \theta \left(e^{-\frac{1}{y}}\right)^{\theta-1} \frac{e^{-\frac{1}{y}}}{y^2} I_{(0,+\infty)}(y) = \theta \frac{e^{-\frac{\theta}{y}}}{y^2} I_{(0,+\infty)}(y).$$

We consider:

$$IC = \left(\frac{Y}{2}, Y\right);$$

and calculate its confidence level:

$$\inf_\theta \mathbb{P}_\theta \left(\frac{Y}{2} < \theta < Y\right) = \inf_\theta \mathbb{P}_\theta \left(\theta < Y < 2\theta\right) =$$

$$= \inf_\theta \int_\theta^{2\theta} \theta \frac{e^{-\frac{\theta}{y}}}{y^2} \, \mathrm{d}y =$$

$$= \inf_\theta e^{-\frac{\theta}{y}} \Big|_\theta^{2\theta} = e^{-\frac{1}{2}} - e^{-1} = 0.2333.$$

Note that it does not depend on θ.

(b) $Q = X^\theta$ is a pivotal quantity, indeed:

$$Q = h(X) = X^\theta.$$

$$X = h^{-1}(Q) = Q^{\frac{1}{\theta}}.$$

$$f_Q(q) = f_X(q^{\frac{1}{\theta}}) \left|\frac{1}{\theta} q^{\frac{1-\theta}{\theta}}\right| = \theta q^{\frac{\theta-1}{\theta}} \frac{1}{\theta} q^{\frac{1-\theta}{\theta}} = 1 \qquad 0 < q < 1.$$

$$Q \sim U_{[0,1]}.$$

We write the confidence interval with a confidence level of $1-\alpha$:

$$\begin{cases} IC & = [a < X^{\theta} < b] \iff \left[\log a < \theta \log X < \log b\right] \\ & \qquad\qquad\quad\;\; \iff \left[\frac{\log b}{\log X} < \theta < \frac{\log a}{\log X}\right]; \\ b-a & = 1-\alpha. \end{cases}$$

We minimise the length of the IC, $L \propto (\log a - \log b)$:

$$\begin{cases} \min_{(b,a)}(\log b - \log a); \\ a = b-(1-\alpha). \end{cases}$$

$$\Downarrow$$

$$\min_{b\in[0,1]} (\log b - \log(b-(1-\alpha))) =$$

$$= \min_{b\in[0,1]} \log\left(\frac{b}{b-(1-\alpha)}\right) =$$

$$= \max_{b\in[0,1]} \log\left(1 - \frac{1-\alpha}{b}\right).$$

The minimum length is reached for $b=1$.
Therefore, the IC of minimum length and confidence level equal to $1-\alpha$ is:

$$IC(1-\alpha) = \left[0, \frac{\log \alpha}{\log X}\right].$$

(c)

$$IC_{(1-\alpha)} = \left(\frac{Y}{2}, Y\right) = \left(-\frac{1}{2\log X}, -\frac{1}{\log X}\right)$$

is of the same type as the IC calculated in point (b), with:

$$\begin{cases} \log b = -\frac{1}{2} & \iff b = e^{-\frac{1}{2}}; \\ \log a = -1 & \iff a = e^{-1}. \end{cases}$$

Since the ICs calculated in point (a) and point (b) are of the same type and the IC in point (b) is of minimum length, the IC in point (a) certainly has a length greater than the IC in point (b).

7.6

(a) X is a sample of size 1. Let's calculate the law of Q:

$$f_Q(q) = \mathbb{P}\{X - \mu \le q\} = \mathbb{P}\{X \le q + \mu\} =$$
$$= \int_{-\infty}^{q+\mu} \frac{1}{2} e^{-|x-\mu|}\, \mathrm{d}x \overset{y=x-\mu}{=} \int_{-\infty}^{q} \frac{1}{2} e^{-|y|}\, \mathrm{d}y.$$

Therefore, Q is a pivotal quantity ($Q \sim f_X(x; 0)$).

(b)

$$\begin{cases} CI & = [a < X - \mu < b] \quad \Rightarrow \quad [X - b < \mu < X - a]; \\ 1 - \alpha & = \frac{1}{2} \int_a^b e^{-|x|}\, \mathrm{d}x. \end{cases}$$

The length of the interval is $b - a$. Moreover, $f_X(x; 0)$ is a unimodal density. Therefore, the minimum length is obtained for $|a| = |b|$, $a = -b$. We then impose that the confidence level of the interval is equal to $1 - \alpha$:

$$1-\alpha = \frac{1}{2} \int_{-b}^{b} e^{-|x|}\, \mathrm{d}x = \int_0^b e^{-x}\, \mathrm{d}x = 1-e^{-b} \quad \Rightarrow \quad b = -\log(\alpha) = \log \frac{1}{\alpha}.$$

We therefore obtain:

$$CI_{(1-\alpha)} = \left[X \pm \log \frac{1}{\alpha} \right].$$

Chapter 8
Asymptotic Statistics

8.1 Theory Recap

Definition 8.1 (Consistency) A sequence of estimators $W_n = W_n(\boldsymbol{X})$ is consistent for the parameter θ if, $\forall \varepsilon > 0$ and $\forall \theta \in \Theta$, it holds:

$$\lim_{n\to+\infty} \mathbb{P}_\theta \left(|W_n - \theta| < \varepsilon\right) = 1;$$

that is $W_n \xrightarrow{p} \theta$.

Theorem 8.1 *Let W_n be a sequence of estimators consistent for θ, such that:*

- $\lim_{n\to+\infty} \mathbb{E}_\theta[W_n] = \theta$*;*
- $\lim_{n\to+\infty} \text{Var}_\theta(W_n) = 0$*;*

then W_n is a consistent estimator for θ.

Theorem 8.2 (Consistency for MLE) *Let $X_1, \ldots, X_n$ be i.i.d. random variables such that $X_i \sim f_X(x; \theta)$ and $L(\theta; \boldsymbol{x})$ the corresponding likelihood. Let $\hat{\theta}$ be the MLE of θ and $\tau(\theta)$ a continuous function of θ. Under suitable regularity conditions for $f_X(x; \theta)$ and $L(\theta; \boldsymbol{x})$ (see Miscellaneous 10.6.2 [3]), then $\forall \varepsilon > 0$ and $\forall \theta \in \Theta$:*

$$\lim_{n\to+\infty} \mathbb{P}_\theta\{|\tau(\hat{\theta}) - \tau(\theta)| \geq \varepsilon\} = 0.$$

Then $\tau(\hat{\theta})$ is a consistent estimator for $\tau(\theta)$.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_8

Definition 8.2 (Limit Variance) Consider an estimator T_n. If, given a sequence k_n, it holds:

$$\lim_{n\to+\infty} k_n Var(T_n) = \tau^2 < +\infty;$$

then τ^2 is called limit variance.

Definition 8.3 (Asymptotic Variance) Consider an estimator T_n, such that:

$$k_n(T_n - \tau(\theta)) \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

Then σ^2 is called the asymptotic variance of T_n.

Definition 8.4 (Asymptotic Efficiency) A sequence of estimators W_n is asymptotically efficient for the parameter $\tau(\theta)$ if:

$$\sqrt{n}\,(W_n - \tau(\theta)) \xrightarrow{\mathcal{L}} N(0, v(\theta));$$

where:

$$v(\theta) = \frac{(\tau'(\theta))^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(x;\theta)\right)^2\right]}.$$

Therefore, the asymptotic variance coincides with the Cramér-Rao limit.

Theorem 8.3 (Asymptotic Efficiency for MLE) *Let $X_1, \ldots, X_n$ be i.i.d. random variables such that $X_i \sim f_X(x;\theta)$ and let $\hat{\theta}$ be the MLE of θ and $\tau(\theta)$ a continuous function of θ. Under suitable regularity conditions for $f_X(x;\theta)$ and $L(\theta;\boldsymbol{x})$ (see Miscellaneous 10.6.2 [3]), it holds:*

$$\sqrt{n}\left(\tau(\hat{\theta}) - \tau(\theta)\right) \xrightarrow{\mathcal{L}} N(0, v(\theta));$$

where $v(\theta)$ is the Cramér-Rao lower limit. Therefore, $\tau(\hat{\theta})$ is a consistent and asymptotically efficient estimator for $\tau(\theta)$.

Definition 8.5 (Asymptotic Relative Efficiency) Consider two estimators W_n and V_n for $\tau(\theta)$ such that:

$$\sqrt{n}\,(W_n - \tau(\theta)) \xrightarrow{\mathcal{L}} N(0, \sigma_W^2);$$

$$\sqrt{n}\,(V_n - \tau(\theta)) \xrightarrow{\mathcal{L}} N(0, \sigma_V^2).$$

The Asymptotic Relative Efficiency (ARE) is defined as the following ratio:

$$ARE(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}.$$

8.2 Exercises

Exercise 8.1 The response time of a computer to the input of a terminal is modelled by a random variable of exponential law $\mathcal{E}(\theta)$, with θ unknown. We measure n response times $T_1, \ldots, T_n$ to estimate the expected response time $1/\theta$ and to estimate the parameter θ.

Let $\overline{T}_n = \frac{1}{n}\sum_{i=1}^n T_i$ be the estimator of $1/\theta$.

(a) Show that it is an unbiased estimator.
(b) Determine its law.
(c) Study its asymptotic normality, consistency and asymptotic efficiency.
Now consider the estimation of θ.
(d) Derive from $1/\overline{T}_n$ an unbiased estimator $\hat{\theta}_n$ of θ.
(e) Study its asymptotic normality, consistency and asymptotic efficiency.
(f) Construct a critical region of level (approximately) α to test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.
(g) Deduce from $\hat{\theta}_n$ an asymptotically pivotal quantity with which to construct a confidence interval for θ level (approximately) $1 - \alpha$.
(h) Find a transformation $g : [0, +\infty) \to \mathbb{R}$ that stabilises the asymptotic variance of $g(\hat{\theta}_n)$, i.e. such that the asymptotic variance of $g(\hat{\theta}_n)$ is independent of θ.
(i) Propose a confidence interval for θ of level (approximately) $1 - \alpha$ constructed based on $g(\hat{\theta}_n)$. Compare this confidence interval with the one obtained in point (g).

Exercise 8.2 Let $X_1, \ldots, X_n$ be a random sample from a uniform law on the interval $[0, \theta]$, $\theta > 0$.

(a) Study the consistency of $X_{(n)}$, the maximum likelihood estimator of θ, and of $X_{(n)}(n + 1)/n$, the unbiased estimator of θ.
(b) For the confidence interval for θ of level $1 - \alpha$ of minimum length that can be constructed with the pivotal quantity $X_{(n)}/\theta$, study the limit of this length for $n \to \infty$.

Exercise 8.3 Let $X_1, \ldots, X_n$ be a family of independent random variables all distributed according to an exponential law with mean τ. Each X_i represents the disintegration time of a nucleus of a certain radioactive element. For each fixed $t \geq 0$, let Y_i be the random variable that is 1 if the i-th nucleus is still alive at time t and 0 otherwise.

Consider the following estimators of τ:

- V_n, MLE based on the sample $Y_1, \ldots, Y_n$.
- $W_n = n \min\{X_1, \ldots, X_n\}$.
- T_n, UMVUE based on the sample $X_1, \ldots, X_n$.

Answer the following questions:

(a) Determine the laws of the estimators W_n and T_n.
(b) Study the asymptotic normality and consistency of V_n, W_n and T_n.
(c) Which is the best estimator?

Exercise 8.4 Let $(X_1, ..., X_n)$ be a random sample drawn from a Poisson distribution with parameter $\lambda > 0$. Let $\tau(\lambda) = e^{-\lambda}(1+\lambda)$. To estimate τ, consider the MLE and the UMVUE. Determine if they are consistent estimators.

Exercise 8.5 Let $X_1, \ldots, X_n$ be a sample of independent random variables with beta(θ, 1) density,

$$f_X(x; \theta) = \theta x^{\theta-1} \mathbb{I}_{(0,1)}(x), \qquad \theta > 0;$$

and let $\hat{\theta}_n$ and $\hat{\theta}_{ML}$ be the UMVUE and MLE estimators of θ, respectively.

(a) Discuss the consistency, asymptotic normality and asymptotic efficiency of the two estimators.
(b) Construct the critical regions of level (approximately) α to test $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$.

Exercise 8.6 Let $X_1, ..., X_n$ be a random sample from a $\Gamma(2, 1/\theta)$ with $\theta > 0$. We therefore have

$$f_X(x; \theta) = \theta^{-2}\, x\, \mathrm{e}^{-x/\theta} I_{(0,+\infty)}(x).$$

Let $\hat{\theta}_n$ be the maximum likelihood estimator for θ.

(a) Verify that $\hat{\theta}_n$ is consistent.
(b) Determine the asymptotic distribution of $\hat{\theta}_n$.
(c) Determine the maximum likelihood estimator $\hat{\sigma}_n^2$ for the variance of X_1.
(d) Determine the asymptotic distribution of $\hat{\sigma}_n^2$.

Exercise 8.7 Let $X_1, \ldots, X_n$ be a random sample from

$$f_X(x; \theta) = \frac{1}{2}\Big(1 + \theta\, x\Big)\, \mathbb{I}_{(-1,1)}(x), \qquad \theta \in [-1, 1].$$

(a) Using the method of moments, determine an estimator $\hat{\theta}_n$ of θ.
(b) Determine the asymptotic distribution of $\hat{\theta}_n$.
(c) Propose an asymptotic confidence interval of level $1 - \alpha$ for θ.

Exercise 8.8 Given a random sample $X_1, \ldots, X_n$ from a Bernoulli distribution $Be(p)$, consider $V_n(\boldsymbol{X}) = n\overline{X}_n(1 - \overline{X}_n)/(n-1)$ the UMVUE estimator of the variance σ^2 of the distribution.

(a) Show that $V_n(\boldsymbol{X})$ is consistent for σ^2.
(b) Determine the asymptotic law of $V_n(\boldsymbol{X})$.

8.3 Solutions

8.1

(a)

$$\mathbb{E}[\overline{T}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[T_i] = \frac{1}{\theta}.$$

Hence, $\overline{T}_n$ is an unbiased estimator for $\frac{1}{\theta}$.

(b)

$$\sum_{i=1}^{n} T_i \sim \Gamma(n, \theta) \implies \overline{T}_n \sim \Gamma(n, n\theta).$$

(c) By the CLT (Theorem 1.16):

$$\sqrt{n}\left(\overline{T}_n - \frac{1}{\theta}\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{\theta^2}\right).$$

Asymptotic Normality implies consistency.
We calculate the Cramér-Rao Limit, considering that $\tau(\theta) = \frac{1}{\theta}$:

$$\begin{aligned}\frac{(\tau'(\theta))^2}{I_1(\theta)} &= \frac{\frac{1}{\theta^4}}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]} = \\ &= \frac{1}{\theta^4\mathbb{E}\left[\frac{\partial}{\partial\theta}(\log\theta - \theta x)^2\right]} = \\ &= \frac{1}{\theta^4\mathbb{E}[\left(X - \frac{1}{\theta}\right)^2]} = \\ &= \frac{1}{\theta^4 Var(T_i)} = \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2}.\end{aligned}$$

Hence, $\overline{T}_n$ is asymptotically efficient.

(d)

$$\mathbb{E}\left[\frac{1}{\overline{T}_n}\right] = n\mathbb{E}\left[\frac{1}{\sum T_i}\right] = \frac{n}{n-1}\theta.$$

Hence, an unbiased estimator for θ is:

$$\hat{\theta}_n = \frac{n-1}{n}\frac{1}{\overline{T}_n} = \frac{n-1}{\sum T_i}.$$

(e) Starting from the results of point (c), we apply the delta method (Theorem 1.17) with $g(t) = 1/t$ and $g'(t) = -1/t^2$ and we obtain:

$$\sqrt{n}\left(\frac{1}{\overline{T}_n} - \theta\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{\theta^2}\left(-\frac{1}{\left(\frac{1}{\theta^2}\right)^2}\right)^2\right) = N\left(0, \theta^2\right).$$

Applying the Slutsky Theorem (1.15), we conclude that:

$$\hat{\theta}_n = \frac{n-1}{n}\frac{1}{\overline{T}_n}$$

is asymptotically normal, that is:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \theta^2).$$

The Cramér Rao Limit is:

$$\frac{1}{Var(T_i)} = \theta^2;$$

hence $\hat{\theta}_n$ is also asymptotically efficient.

(f) We consider the test:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0.$$

Given that $\hat{\theta}_n \approx N(\theta, \theta^2)$, the rejection region of approximately level α is:

$$R_\alpha = \left\{\frac{|\hat{\theta}_n - \theta_0|}{\theta_0/\sqrt{n}} > z_{1-\frac{\alpha}{2}}\right\} = \left\{|\hat{\theta}_n - \theta_0| > z_{1-\frac{\alpha}{2}}\frac{\theta_0}{\sqrt{n}}\right\}.$$

(g) Similarly to the previous point, given that:

$$\frac{\hat{\theta}_n - \theta}{\theta/\sqrt{n}} \approx N(0, 1);$$

$$1 - \alpha = \mathbb{P}\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\hat{\theta}_n - \theta}{\theta/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right\} =$$
$$= \mathbb{P}\left\{\theta\left(1 - z_{1-\frac{\alpha}{2}}/\sqrt{n}\right) \leq \hat{\theta}_n \leq \theta\left(1 + z_{1-\frac{\alpha}{2}}/\sqrt{n}\right)\right\};$$

the asymptotic CI is:

$$IC_{1-\alpha} = \left[\frac{\hat{\theta}_n}{1 + z_{1-\frac{\alpha}{2}}/\sqrt{n}}; \frac{\hat{\theta}_n}{1 - z_{1-\frac{\alpha}{2}}/\sqrt{n}}\right].$$

(h) $g(\hat{\theta}_n)$ has asymptotic variance $\theta^2 g'(\theta)^2$ (according to the Delta Method 1.17). We impose that the asymptotic variance is equal to 1, therefore $g'(\theta) = \frac{1}{\theta}$, or $g(\theta) = \log\theta$. We define:

$$W_n = \log\hat{\theta}_n;$$

which is such that:

$$\sqrt{n}\left(W_n - \log\theta\right) \xrightarrow{\mathcal{L}} N(0, 1).$$

(i) The asymptotic CI based on W_n can be constructed by observing that asymptotically:

$$1 - \alpha = \left[W_n - z_{1-\frac{\alpha}{2}}/\sqrt{n} \leq \log\theta \leq W_n + z_{1-\frac{\alpha}{2}}/\sqrt{n}\right];$$

therefore:

$$IC_{1-\alpha} = \left[e^{W_n}e^{-z_{1-\frac{\alpha}{2}}/\sqrt{n}}; e^{W_n}e^{z_{1-\frac{\alpha}{2}}/\sqrt{n}};\right].$$

Developing the extremes of this CI to the I order, we find the one obtained in point (g).

8.2

(a) For the calculation of the MLE and its mean and variance, see Exercise 4.2.

$$\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta;$$

$$Var(X_{(n)}) = \frac{n}{(n+1)^2(n+2)}\theta^2.$$

The MLE law is:

$$F_{X_{(n)}}(t) = \begin{cases} 0 & t < 0; \\ \left(\frac{t}{\theta}\right)^n & 0 \le t \le \theta; \\ 1 & t > \theta. \end{cases}$$

Therefore $X_{(n)} \xrightarrow{\mathcal{L}} \theta$ and since θ is constant, $X_{(n)} \xrightarrow{p} \theta$.
We now consider $T_n = X_{(n)}(n+1)/n$.
T_n is an unbiased estimator for θ, in fact: $\mathbb{E}[T_n] = \theta$.
Then:

$$MSE(T_n) = Var(T_n) = \frac{(n+1)^2}{n^2}\frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{\theta^2}{n(n+2)} \to 0.$$

Therefore, by Theorem 8.1, T_n is consistent.

(b) $Q = X_{(n)}/\theta$ is a pivotal quantity since:

$$F_Q(t) = \mathbb{P}\{X_{(n)}/\theta \le t\} = \mathbb{P}\{X_{(n)} \le t\theta\} = \begin{cases} 0 & t < 0; \\ t^n & 0 \le t \le 1; \\ 1 & t > 1. \end{cases}$$

We therefore calculate a CI that has a confidence level equal to $1-\alpha$.

$$\begin{cases} CI & = \left[a \le X_{(n)}/\theta \le b\right] = \left[\frac{X_{(n)}}{b}; \frac{X_{(n)}}{a}\right]; \\ 1-\alpha & = b^n - a^n. \end{cases}$$

The length of the CI is proportional to $1/a - 1/b$. We therefore identify the pair (a, b) that allows us to have the CI of minimum length. We solve the following constrained optimal problem:

$$\begin{cases} \min\limits_{a,b} \frac{1}{a} - \frac{1}{b}; \\ 1-\alpha = b^n - a^n. \end{cases}$$

We derive both expressions with respect to a:

$$\begin{cases} \frac{\partial L}{\partial a} & = -\frac{1}{a^2} + \frac{1}{b^2}\frac{db}{da} \\ 0 & = nb^{n-1}\frac{db}{da} - na^{n-1} \end{cases} \Rightarrow \begin{cases} -\frac{1}{a^2} + \frac{1}{b^2}\left(\frac{a}{b}\right)^{n-1} & = \frac{a^{n+1}-b^{n+1}}{a^2 b^{n+1}} < 0; \\ \frac{db}{da} & = \left(\frac{a}{b}\right)^{n-1}. \end{cases}$$

Therefore the minimum length is obtained at the maximum value of a, so $b = 1$ and $a = \sqrt[n]{\alpha}$. The CI is:

$$CI_{1-\alpha} = \left[X_{(n)}; \frac{X_{(n)}}{\sqrt[n]{\alpha}} \right];$$

whose length is $\frac{1}{\sqrt[n]{\alpha}} - 1 \overset{n \to +\infty}{\to} 0$.

8.3

(a) From the text we deduce that:

$$X_1, \dots, X_n \sim \mathcal{E}\left(\frac{1}{\tau}\right), \qquad Y_i = \mathbb{I}_{\{X_i > t\}} \implies Y_i \sim Be(e^{-t/\tau}).$$

Since $\overline{Y}_n$ is MLE for $e^{-t/\tau}$, by the principle of invariance the following estimator V_n:

$$V_n := -\frac{t}{\log \overline{Y}_n}$$

is MLE for τ.
We consider the estimator W_n, defined as follows:

$$W_n := n \min\{X_1, \dots, X_n\} = nX_{(1)}.$$

Let's calculate its law.
We know that: $X_{(1)} \sim \mathcal{E}\left(\frac{n}{\tau}\right)$, therefore:

$$\mathbb{P}\{W_n > t\} = \mathbb{P}\{nX_{(1)} > t\} = \mathbb{P}\{X_{(1)} > t/n\} = e^{-t/\tau}.$$

Therefore:

$$W_n \sim \mathcal{E}\left(\frac{1}{\tau}\right).$$

We consider the estimator T_n, UMVUE for τ:

$$T_n := \overline{X}_n.$$

Let's calculate its law.
We know that: $X_1 \sim \mathcal{E}\left(\frac{1}{\tau}\right)$, therefore:

$$\sum X_i \sim \Gamma\left(n, \frac{1}{\tau}\right) \implies T_n \sim \Gamma\left(n, \frac{n}{\tau}\right).$$

(b) We evaluate the asymptotic normality of V_n, W_n and T_n. For the CLT (1.16) we know that:

$$\sqrt{n}\left(\overline{Y}_n - e^{-t/\tau}\right) \xrightarrow{\mathcal{L}} N\left(0, e^{-t/\tau}(1 - e^{-t/\tau})\right).$$

To evaluate the asymptotic normality of $V_n = -\frac{t}{\log \overline{Y}_n}$, we exploit the delta method (Theorem 1.17), with $g(\overline{Y}_n) = -\frac{t}{\log \overline{Y}_n}$.

$$g(x) = -\frac{t}{\log x} \qquad g'(x) = \frac{t}{(\log x)^2}\frac{1}{x},$$

therefore:

$$\sqrt{n}\,(V_n - \tau) \xrightarrow{\mathcal{L}} N\left(0, e^{-t/\tau}(1 - e^{-t/\tau}) \cdot \frac{t^2}{\left(-\frac{t}{\tau}\right)^4} e^{2t/\tau}\right)$$
$$= N\left(0, \frac{\tau^4}{t^2} e^{t/\tau}(1 - e^{-t/\tau})\right).$$

W_n is not asymptotically normal and is not consistent.

$$\sqrt{n}(T_n - \tau) \xrightarrow{\mathcal{L}} N(0, \tau^2).$$

(c) To evaluate the best estimator, we calculate the ARE between V_n and T_n:

$$ARE(T_n, V_n) = \frac{\tau^2 t^2}{\tau^4 e^{t/\tau}(1 - e^{-t/\tau})}.$$

The ARE depends on τ, so we cannot identify a better estimator.

8.4

$$X_i \sim P(\lambda), \qquad \tau(\lambda) = e^{-\lambda}(1 + \lambda).$$

By the principle of invariance, the MLE of $\tau(\lambda)$ is $(1 + \overline{X}_n)e^{-\overline{X}_n}$.

By the SLLN 1.11, it holds:

$$(1 + \overline{X}_n)e^{-\overline{X}_n} \xrightarrow{q.c.} \tau(\lambda);$$

therefore the MLE is consistent.

We consider the UMVUE:

$$\underbrace{\left(1-\frac{1}{n}\right)^{n\overline{X}_n}}_{\to (e^{-1})^{\lambda}}\underbrace{\left(1+\overline{X}_n\frac{n}{n-1}\right)}_{\to(1+\lambda)} \to e^{-\lambda}(1+\lambda).$$

Therefore the UMVUE is consistent. For the calculation of the UMVUE, refer to Exercise 4.10.

8.5

(a)

$$X_i \sim \theta x^{\theta-1}\mathbb{I}_{(0,1)}(x)\,, \qquad \theta > 0.$$

In Exercise 4.8 the UMVUE $\hat{\theta}_n$ and its expected value and variance were calculated. We report the results obtained below:

$$\hat{\theta}_n = -\frac{n-1}{\sum_i \log X_i}.$$

$$\mathbb{E}[\hat{\theta}_n] = \theta.$$

$$Var(\hat{\theta}_n) = \frac{\theta^2}{n-2}.$$

Using Theorem 8.1, we conclude that $\hat{\theta}_n$ is consistent.
We now evaluate the asymptotic normality.

$$Y_i = -\log X_i \sim \mathcal{E}(\theta) \overset{CLT}{\Longrightarrow} \sqrt{n}\left(\overline{Y}_n - \frac{1}{\theta}\right) \overset{\mathcal{L}}{\to} N\left(0, \frac{1}{\theta^2}\right).$$

Through the delta method (Theorem 1.17):

$$-\frac{n}{\sum \log X_i} \Longrightarrow \sqrt{n}\left(\frac{1}{\overline{Y}_n} - \theta\right). \overset{\mathcal{L}}{\to} N\left(0, \theta^2\right)$$

Therefore, by Slutsky's Theorem (1.15):

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{\mathcal{L}}{\to} N(0, \theta^2).$$

$\hat{\theta}_n$ is asymptotically efficient $\frac{\theta^2}{n} = \frac{1}{nI(\theta)}$.

(b) We consider the test:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$

The critical region is:

$$R_\alpha = \left\{ \hat{\theta}_n > \theta_0 + z_{1-\alpha} \frac{\theta_0}{\sqrt{n}} \right\}.$$

8.6

(a)

$$f(x;\theta) = \theta^{-2}\, x\, \mathrm{e}^{-x/\theta} \mathbb{I}_{(0,+\infty)}(x).$$

$$L(\theta; \boldsymbol{x}) = \theta^{-2n} \prod x_i \mathrm{e}^{-\frac{\sum_i x_i}{\theta}} \mathbb{I}_{(0,+\infty)}(x_i).$$

$$l(\theta; \boldsymbol{x}) \propto -2n \log\theta - \frac{\sum_i x_i}{\theta}.$$

$$\frac{\partial l(\theta; \boldsymbol{x})}{\partial\theta} = -\frac{2n}{\theta} + \frac{\sum_i x_i}{\theta^2} > 0 \iff \theta < \frac{\overline{X}_n}{2}.$$

Therefore $\hat{\theta}_n = \frac{\overline{X}_n}{2}$ is the MLE for θ.
By the SLLN (1.11) it holds:

$$\overline{X}_n \stackrel{q.c.}{\rightarrow} \mathbb{E}[X_i] = 2\theta.$$

Therefore $\hat{\theta}_n$ is consistent for θ.

(b) The CLT (1.16) guarantees that:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \stackrel{\mathcal{L}}{\rightarrow} N\left(0, \frac{2\theta^2}{4}\right) = N\left(0, \frac{\theta^2}{2}\right).$$

(c)

$$Var(X_i) = 2\theta^2;$$

therefore by the principle of invariance it holds:

$$\hat{\sigma}_n^2 = 2\hat{\theta}_n^2 = \frac{\overline{X}_n^2}{2}.$$

(d) We use the delta method (Theorem 1.17), considering $g(x) = \frac{1}{2}x^2$ and $g'(x) = x$:

$$\sqrt{n}\left(\hat{\sigma}_n^2 - 2\theta^2\right) \stackrel{\mathcal{L}}{\rightarrow} N\left(0, 2\theta^2 \cdot (2\theta)^2\right) = N(0, 8\theta^4).$$

8.7

(a)

$$\mathbb{E}[X_i] = \frac{1}{2}\int_{-1}^{1}(1+\theta x)x\,\mathrm{d}x = \frac{\theta}{2}\int_{-1}^{1}x^2\,\mathrm{d}x = \frac{\theta}{2}\frac{2}{3} = \frac{\theta}{3}.$$

Therefore, the estimator obtained by the method of moments is:

$$\hat{\theta}_n = 3\overline{X}_n.$$

Note that $\hat{\theta}_n \stackrel{q.c.}{\to} \theta$.

(b)

$$Var(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \frac{1}{2}\int_{-1}^{1}x^2\,\mathrm{d}x - \frac{\theta^2}{9} = \frac{1}{3} - \frac{\theta^2}{9} = \frac{3-\theta^2}{9}.$$

Then, the CLT (1.16) ensures that:

$$\sqrt{n}(\hat{\theta}_n - \theta) \stackrel{\mathcal{L}}{\to} N(0, 3-\theta^2).$$

(c) Using the Slutsky's Theorem (1.15), we obtain:

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\frac{3-\hat{\theta}_n^2}{n}}} \approx N(0, 1).$$

Therefore

$$IC_{(1-\alpha)} = \left[\hat{\theta}_n \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{3-\hat{\theta}_n^2}{n}}\right] = \left[3\overline{X}_n \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{3-9\overline{X}_n^2}{n}}\right].$$

8.8

(a) For the SLLN (1.11):

$$V_n(\boldsymbol{X}) = n\overline{X}_n(1-\overline{X}_n)/(n-1) \stackrel{q.c.}{\to} p(1-p) = \sigma^2.$$

Therefore V_n is a consistent estimator for σ^2.

(b) For the CLT (1.16):

$$\sqrt{n}\left(\overline{X}_n - p\right) \stackrel{\mathcal{L}}{\to} N(0, p(1-p)).$$

To prove the asymptotic normality of V_n we use Slutsky's Theorem (1.15) with $g(x) = x(1-x)$, $g'(x) = 1-2x$ and $g''(x) = -2$.
If $p \neq 1/2$:

$$\sqrt{n}\left(\overline{X}_n(1-\overline{X}_n) - p(1-p)\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{(1-2p)^2 p(1-p)}{n}\right).$$

If $p = 1/2$:

$$\sqrt{n}\left(\overline{X}_n(1-\overline{X}_n) - \frac{1}{4}\right) \xrightarrow{\mathcal{L}} -\frac{1}{4}\chi^2(1).$$

Therefore $V_n \xrightarrow{\mathcal{L}} \frac{1}{4} - \frac{1}{4n}\chi^2(1)$.

Part II
Regression Models and Analysis of Variance

Chapter 9
Linear Regression

9.1 Theory Recap

We tackle the statistical study of the behaviour of a random variable Y (called response or dependent variable) with respect to other quantities $X_1, X_2, .., X_r$ (called predictors or independent variables) which in this study will be assumed deterministic.

We assume that the following relationship may exist:

$$Y = \beta_0 + \beta_1 X_1 + .. + \beta_r X_r + \varepsilon; \tag{9.1}$$

where ε is a random variable with zero mean and variance σ^2. β_0, β_1,..,β_r and σ^2 are real and unknown parameters.

Assuming we have a sample of n joint observations Y_i and their relative x_{ij}, $j = 1, .., r$ and $i = 1, .., n$, then for each observation it holds:

$$Y_i = \beta_0 + \beta_1 x_{i1} + .. + \beta_r x_{ir} + \varepsilon_i \quad i = 1, .., n; \tag{9.2}$$

Equation (9.2) can be written compactly as:

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \tag{9.3}$$

where:

$\mathbf{Y} \in \mathbb{R}^n$ is the vector of random responses, $\mathbf{y}$ are the relative realisations $\mathbf{y} = (y_1, .., y_n)^T$.
$\mathbb{X} \in \mathbb{R}^{n\times(r+1)}$ is the design matrix, in which the first column is the unit vector $(1, .., 1)^T$ and the subsequent columns are the vectors $\mathbf{x}_j = (x_{1j}, .., x_{nj})^T$.
$\boldsymbol{\beta} \in \mathbb{R}^{r+1}$ is the vector of unknown regression parameters.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_9

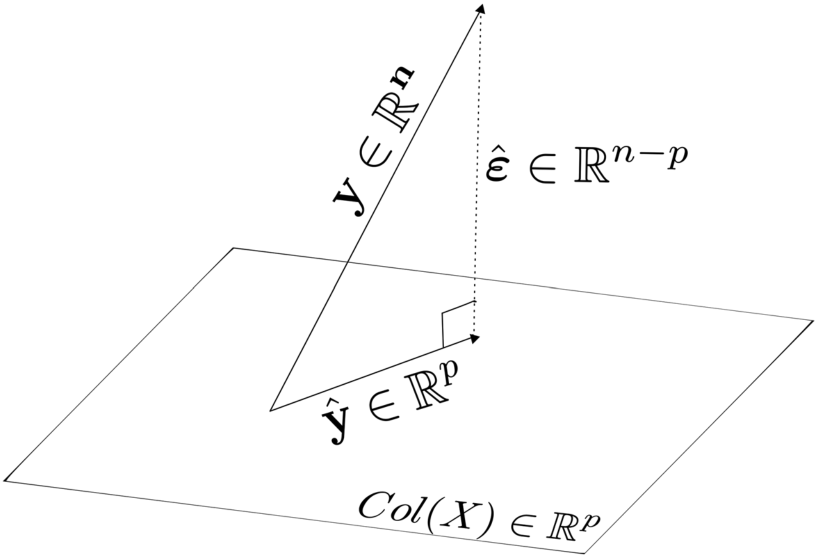

Fig. 9.1 Graphical representation of the least squares method

$\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the vector of errors, random and unknown, such that $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbb{I}_n$.

The model described in Eq. (9.2) is called linear, as it is linear with respect to $\boldsymbol{\beta}$.

9.1.1 Estimators of the Unknown Parameters of the Regression: Least Squares Method

The least squares (LS) method is an approach used to obtain the estimator of $\boldsymbol{\beta}$. This method is based on the minimisation of the square of the error, namely:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \varepsilon_i^2 = \arg\min_{\boldsymbol{\beta}} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbb{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbb{X}\boldsymbol{\beta}). \tag{9.4}$$

Differentiating with respect to $\boldsymbol{\beta}$ and setting the differential to $\mathbf{0}$, we obtain the estimator:

$$\hat{\boldsymbol{\beta}} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbf{y}. \tag{9.5}$$

The estimates obtained with the least squares method are the projection of $\mathbf{y}$ onto the column space of the design matrix $\mathbb{X}$, $Col(\mathbb{X})$ (see Fig. 9.1).

Furthermore, we can define the following quantities:

$\hat{\mathbf{y}}$, the estimated responses: $\hat{\mathbf{y}} = \mathbb{X}\hat{\boldsymbol{\beta}} = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y} = H\mathbf{y}$, where H is the projection matrix onto the space $Col(\mathbb{X})$.
$\hat{\boldsymbol{\varepsilon}}$, the estimated error vector: $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbb{X}\hat{\boldsymbol{\beta}}$.

Theorem 9.1 (Gauss-Markov) *Given the linear model in Eq. (9.3), the estimators obtained with the least squares method are unbiased and have minimum variance (BLUE, Best Linear Unbiased Estimator).*

Theorem 9.2 *Given the linear model in Eq. (9.3), assume that the rank of* $\mathbb{X}$ *is* $p = r + 1$*, i.e., the design matrix has full rank, then:*

- $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.
- $Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$.
- $\mathbb{E}[\hat{\boldsymbol{\varepsilon}}] = \mathbf{0}$.
- $Cov(\hat{\boldsymbol{\varepsilon}}) = \sigma^2(\mathbb{I} - H)$.
- $\mathbb{E}[\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}] = \sigma^2(n - p)$.

Theorem 9.3 *Given the linear model in Eq. (9.3), assume that the rank of* $\mathbb{X}$ *is* $p = r + 1$*, i.e., the design matrix has full rank, and* $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbb{I}_n)$*, then:*

- $\hat{\boldsymbol{\beta}} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y}$ *is the maximum likelihood estimator for* $\boldsymbol{\beta}$.
- $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}}{n}$ *is the maximum likelihood estimator for* σ^2 *(often calculated using* $S^2 = \frac{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}}{n-p}$*).*
- $\hat{\boldsymbol{\beta}} \sim N(\hat{\boldsymbol{\beta}}, \sigma^2(\mathbb{X}^T\mathbb{X})^{-1})$.
- $\hat{\boldsymbol{\varepsilon}} \sim N(\mathbf{0}, \sigma^2(\mathbb{I} - H))$.
- $\hat{\boldsymbol{\varepsilon}} \perp\!\!\!\perp \hat{\boldsymbol{\beta}}$.
- $n\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \sim \sigma^2\chi^2(n - p)$.

Corollary

$$\frac{1}{\sigma^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbb{X}^T\mathbb{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi^2_p. \tag{9.6}$$

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}. \tag{9.7}$$

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbb{X}^T\mathbb{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{pS^2} \sim F_{p,n-p}. \tag{9.8}$$

9.1.2 Inference

Assuming the hypotheses of Theorem 9.3 are valid, we can perform the following three types of tests:

- Significance of all predictors.

- Significance of a single predictor.
- Difference in significance between nested models.

9.1.3 Confidence Regions and Intervals for Predictors

Starting from Eq. (9.8) we define the confidence region of level $(1-\alpha)$ for $\boldsymbol{\beta}$ as:

$$R_{(1-\alpha)}(\boldsymbol{\beta}) = \left\{ \boldsymbol{\beta} \in \mathbb{R}^p : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbb{X}^T \mathbb{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq p S^2 F_{p,n-p}(1-\alpha) \right\};$$

where $F_{p,n-p}(1-\alpha)$ indicates the quantile of order $(1-\alpha)$ of a Fisher distribution of parameters $p, n-p$. The confidence region has an ellipsoidal shape and does not correspond to the Cartesian product of the marginal confidence intervals, i.e., those related to the individual β_i, $i \in \{0, .., r\}$.

Using the results of Theorem 9.3, we can write the confidence interval of level $(1-\alpha)$ for the individual β_i, $i \in \{0, .., r\}$, as:

$$IC_{(1-\alpha)}(\beta_i) = [\hat{\beta}_i \pm t_{n-p}(1-\alpha/2) S \sqrt{(\mathbb{X}^T \mathbb{X})^{-1}_{ii}}];$$

where $t_{n-p}(1-\alpha/2)$ is the quantile of level $1-\alpha/2$ of a Student's t with $n-p$ degrees of freedom.

9.1.4 Confidence Intervals for Prediction

In the case of a new data point $\mathbf{x}_0$, the first quantity of interest to calculate is the point prediction $\hat{y}_0$, as:

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} + \varepsilon_0, \qquad \varepsilon_0 \sim N(0, \sigma^2) \quad \wedge \quad \varepsilon_0 \perp\!\!\!\perp \hat{\boldsymbol{\beta}}. \tag{9.9}$$

We can also calculate the variability associated with $\hat{y}_0$:

$$Var(\hat{y}_0) = Var(\mathbf{x}_0^T \hat{\boldsymbol{\beta}} + \varepsilon_0) = \mathbf{x}_0^T (\mathbb{X}^T \mathbb{X})^{-1} \mathbf{x}_0 \sigma^2 + \sigma^2. \tag{9.10}$$

Given the variance, we can define the prediction interval of level $1-\alpha$ for the point value given $\mathbf{x}_0$:

$$\text{IP}(\hat{y}_0; \mathbf{x}_0) = \hat{y}_0 \pm t_{n-p}(1-\alpha/2) S \sqrt{\mathbf{x}_0^T (\mathbb{X}^T \mathbb{X})^{-1} \mathbf{x}_0 + 1}. \tag{9.11}$$

We can define the confidence interval of level $1-\alpha$ for the mean of the predicted values given $\mathbf{x}_0$:

$$\mathrm{IC}(\hat{y}_0; \mathbf{x}_0) = \hat{y}_0 \pm t_{n-p}(1-\alpha/2) S \sqrt{\mathbf{x}_0^T (\mathbb{X}^T \mathbb{X})^{-1} \mathbf{x}_0}. \tag{9.12}$$

It is immediately observed that the prediction interval in Eq. (9.11) is wider than the confidence interval Eq. (9.12).

9.1.5 Model Goodness of Fit (GOF)

A measure of the goodness of the model is the $\mathtt{R}^2$ coefficient, also known as the coefficient of determination, and in the case of multiple linear regression, the adjusted $\mathtt{R}^2$.

Definition 9.1 $\mathtt{R}^2$ and adjusted $\mathtt{R}^2$

$$\mathtt{R}^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} = 1 - \frac{SS_{res}}{SS_{tot}};$$

where:

$$SS_{TOT} = SS_{reg} + SS_{res};$$
$$\sum_{i=1}^n (y_i - \bar{\mathbf{y}})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{\mathbf{y}})^2 + \sum_{i=1}^n (y_i - \bar{\hat{\mathbf{y}}})^2.$$

It can be shown that if the design matrix has a constant column then $\mathtt{R}^2 \in [0, 1]$. $\mathtt{R}^2$ can also be expressed as SS_{reg}/SS_{tot} and represents the percentage of variability explained by the regressors, so the closer it is to 1, the more the model explains the response variable.

We can define $\mathtt{R}^2_{adj}$, a measure of the goodness of the model that also takes into account its complexity:

$$\mathtt{R}^2_{adj} = 1 - \frac{SS_{res}}{SS_{tot}} \frac{n-1}{n-r-1}.$$

$\mathtt{R}^2_{adj}$ by definition is always less than or equal to $\mathtt{R}^2$. It is used to evaluate the goodness of the model in the case of multiple linear regression, because it allows to take into account the complexity of the model.

Online Supplementary Material

```
A supplement to this chapter is available online,
containing data, further insights and exercises.
```

9.1.6 Libraries

```
library( car )
## Loading required package: carData
library( ellipse )
##
## Attaching package: 'ellipse'
## The following object is masked from 'package:car':
##
##     ellipse
## The following object is masked from 'package:graphics':
##
##     pairs
library( faraway )
##
## Attaching package: 'faraway'
## The following objects are masked from 'package:car':
##
##     logit, vif
library( leaps )
library( qpcR )
## Loading required package: MASS
## Loading required package: minpack.lm
## Loading required package: rgl
## Loading required package: robustbase
##
## Attaching package: 'robustbase'
## The following object is masked from 'package:faraway':
##
##     epilepsy
## Loading required package: Matrix
```

9.2 Exercises

Exercise 9.1 Manually calculate the estimate of the $\boldsymbol{\beta}$ in the following cases:

(a) No predictors are considered.
(b) Only one predictor is considered.

Exercise 9.2 Derive the test that evaluates the significance of all predictors, under the assumptions of Theorem 9.3.

Exercise 9.3 Derive the test that evaluates the significance of the single predictor, under the assumptions of Theorem 9.3.

Exercise 9.4 Describe the main steps of the study of a regression model, highlighting the R commands to use.

Exercise 9.5 Consider the `savings` dataset in the `faraway` package. This dataset contains information about 50 US states. The covariates are:

- `sr` is personal savings divided by disposable income.
- `pop15` is the percentage of the population under 15 years old.
- `pop75` is the percentage of the population over 75 years old.
- `dpi` is per-capita income in dollars, net of taxes.
- `ddpi` is purchasing power—an aggregate economic index, expressed as a percentage.

These data are averaged over the period 1960–1970, to remove any short-term cycles or fluctuations.

Answer the following questions:

(a) Load the dataset and perform a graphical exploration.
(b) Propose a complete linear model to explain personal savings and comment on all the items in the model.
(c) Explicitly perform the F test on the significance of the model.
(d) Explicitly perform the test on the significance of a regression coefficient related to `pop15`.
(e) Calculate the 95% confidence interval for the regression coefficient related to `pop75`.
(f) Calculate the 95% confidence interval for the regression coefficient related to `ddpi`.
(g) Represent the 95% confidence region for the regression coefficients associated with `pop15` and `pop75`, adding the point (0, 0).
(h) Identify any influential points in the dataset using: H projection matrix, standardised residuals, studentised residuals and Cook's distance.
(i) Compare the influential points identified with the techniques proposed above, using the commands `influencePlot` and `influence.measures`.
(j) Evaluate the impact of the different influential points on the model.
(k) Assess the homoscedasticity of the residuals.
(l) Assess the normality of the residuals.

Exercise 9.6 Load the dataset `data_es2.RData`, available in the online supplementary material. This dataset is related to rocks found in the Casentino forests. The dataset contains the following information:

- `height`: height of the rock [m].
- `iron`: percentage of iron in a cubic millimetre of rock.
- `calcium`: percentage of calcium in a cubic millimetre of rock.

Researchers are interested in assessing whether the percentage of these two elements can be predictive of the height of the rock.

Answer the following questions:

(a) Load the dataset and perform a graphical exploration of the variables.
(b) Propose a model to answer the researchers' question.
(c) Verify the assumptions of the model.
(d) Evaluate a possible transformation of the response variable and redo all the analyses.

Exercise 9.7 Consider the dataset `state`, available in `R`, in which data related to 50 US states are collected. The variables are estimated in July 1975:

- `Income`: per capita income (1974).
- `Illiteracy`: illiteracy (1970, % of population).
- `Life Exp`: life expectancy in years (1969–71).
- `Murder`: murder rate per 100,000 inhabitants (1976).
- `HS Grad`: percentage of high school graduates (1970).
- `Frost`: average number of days with minimum temperature equal to 32° (1931–1960).
- `in capital or large city`.
- `Area`: area (in square miles).

Consider `life expectancy` as the response variable and answer the following questions:

(a) Analyse the data with graphical methods.
(b) Evaluate and comment on a complete linear model.
(c) Assess the validity of the model assumptions.
(d) Evaluate an appropriate model reduction.

Exercise 9.8 We want to study a possible relationship between the height of tomato plants and the average weight in grams of the tomatoes harvested.

The available data are as follows:

```
weight   = c( 60, 65, 72, 74, 77, 81, 85, 90 )
height = c( 160, 162, 180, 175, 186, 172, 177, 184 )
```

Answer the following questions:

(a) Represent the data.
(b) Evaluate a simple linear model that predicts the average weight of the tomatoes as the response variable.
(c) Calculate the confidence interval for the prediction of the mean responses, considering 15 elements that have height within the range of the dataset values.
(d) Calculate the prediction interval of the responses, considering the elements from the previous point.
(e) Compare the intervals obtained in points (c) and (d).

9.3 Solutions

9.1

(a) The model we want to consider in this case is the following:

$$\mathbf{y} = \beta_0 + \varepsilon; \tag{9.13}$$

in which the design matrix is constituted by the unit vector alone ($\mathbb{X} = \mathbb{I}$).

$$\hat{\beta}_0 = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y} = \frac{1}{n}\mathbb{I}^T\mathbf{y} = \frac{\sum_i^n y_i}{n} = \bar{\mathbf{y}};$$

therefore, in the absence of information, the best estimate we can provide is the sample mean.

(b) The model we want to consider in this case is the following:

$$\mathbf{y} = \beta_0 + \beta_1 x + \varepsilon; \tag{9.14}$$

The design matrix is:

$$\mathbb{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \tag{9.15}$$

Let's then calculate:

$$\mathbb{X}^T\mathbb{X} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}. \tag{9.16}$$

$$(\mathbb{X}^T\mathbb{X})^{-1} = \frac{1}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i^2)}{n}} \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}. \tag{9.17}$$

We solve Eq. (9.5).

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i - \sum_i x_i y_i / n}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i^2)}{n}} = \frac{S_{xy}}{S_{xx}}; \tag{9.18}$$

$$\hat{\beta}_0 = \frac{1}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i^2)}{n}} \left[\bar{y} \left(\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n} \right) + \frac{\sum x_i}{n} \sum_i x_i \bar{y} - \bar{x} \sum_i X_i y_i \right] =$$
$$= \frac{1}{S_{xx}} \left[\bar{y} S_{xx} - \bar{x} \left(\sum_i x_i y_i - \frac{x_i y_i}{n} \right) \right] =$$
$$= \frac{1}{S_{xx}} \left[\bar{y} S_{xx} - \bar{x} S_{xy} \right] =$$
$$= \bar{y} - \bar{x} \hat{\beta}_1.$$

9.2 The test we want to perform is the following:

$$H_0 : \beta_0 = \beta_1 = .. = \beta_p = 0 \quad \text{vs} \quad H_1 : \exists i \in \{1, .., p\} | \beta_i \neq 0.$$

The test statistic to answer this test is based on the total variability SS_{TOT} and the residual variability SS_{res}:

$$SS_{TOT} = \sum_{i=1}^n (y_i - \bar{y})^2;$$
$$SS_{res} = \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

The test statistic is the following:

$$F = \frac{SS_{TOT} - SS_{res}/(p-1)}{SS_{res}/(n-p)};$$

and it is distributed as a Fisher with parameters $p - 1$ and $n - p$.

If the p-value associated with F is less than 5%, we reject the null hypothesis, i.e., there is at least one regression coefficient different from zero.

9.3 The test we want to perform is the following:

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_1 : \beta_i \neq 0.$$

To perform the test, we construct the following test statistic T:

$$T = \frac{|\hat{\beta}_i - 0|}{se(\hat{\beta}_i)};$$

where $se(\hat{\beta}_i)$ is the standard error of the coefficient estimate:

$$se(\hat{\beta}_i) = \sqrt{\hat{\sigma}^2 \cdot (X^T X)^{-1}_{ii}}.$$

Considering the assumptions of Theorem 9.3, it can be shown that $T \sim t(n-p)$. We then calculate the p-value of the two-sided test and if it is less than 5% we can conclude that the regression coefficient is different from zero.

9.4 The steps to be performed are:

(a) Visualisation of the dataset using the `pairs` command. To correctly analyse this graph, one must focus on three elements: (1) observe the trend of the response variable with respect to the other variables in the dataset and determine whether these trends suggest a linear regression model; (2) observe the relationship between the variables in the dataset that we would like to use as regressors, if the correlation is high probably, one of the two will be redundant and superfluous within the model. The correlation can be measured with the `cor` command. (3) Notice the possible presence of influential points in the dataset.
If the response variable is continuous and a linear trend can be assumed between this and the predictors, then we can proceed with a linear regression model.

(b) Evaluation of a linear regression model using the `mod = lm(y` $\sim x_1 + x_2 + .. + x_r$`)` command. The parameters to analyse are: (1) the goodness of the model through R^2 and R^2_{adj} and (2) the significance of the regressors through the F test and T test on the individual regressors. These elements can be obtained automatically through the `summary(mod)` command.

(c) Verification of the model assumptions. The assumptions to verify are: (1) homoscedasticity of the residuals and (2) normality of the residuals. Homoscedasticity can be evaluated graphically through a scatterplot of the residuals, which sees the residuals on the y-axis and the $\hat{y}$, the responses estimated by the model on the x-axis, command `plot(mod$fit, mod$res)`. If the points are scattered around zero, we conclude that the assumption of homoscedasticity is valid, if instead we observe a particular pattern the assumption is violated.
The normality assumption can be verified both graphically (through qqplot using the commands `qqnorm(mod$res)` and `qqline(g$res)`) and mathematically through the Shapiro-Wilks test, `shapiro.test(mod$res)`.

These are the main steps for constructing and analysing a regression model. However, we may encounter some issues:

- Presence of influential points.
- Violation of the normality assumption.
- Predictors to which a parameter β is associated for which there is no statistical evidence that it is different from 0.

These issues are identified through:

- Analysis of the projection matrix H, standardised residuals or Cook's distance.
- Shapiro test and qqplot.
- Analysis of the p-values of the t-tests associated with the regressor β_i.

Finally, they can be resolved:

- By removing from the dataset those points defined as influential.
- By transforming the response variable (for example through Box-Cox transformation).
- By reducing the model.

Every time one of these three operations is performed, it is very important to reconsider the validity of the model assumptions.

9.5

(a) Load the dataset.

```
data( savings )

# Dimensions
dim( savings )
## [1] 50  5

# Overview of the first rows
head( savings )
##                sr pop15 pop75     dpi ddpi
## Australia 11.43 29.35  2.87 2329.68 2.87
## Austria   12.07 23.32  4.41 1507.99 3.93
## Belgium   13.17 23.80  4.43 2108.47 3.82
## Bolivia    5.75 41.89  1.67  189.13 0.22
## Brazil    12.88 42.19  0.83  728.47 4.56
## Canada     8.79 31.72  2.85 2982.88 2.43
```

In Fig. 9.2, we visualise the dataset using the `pairs` command, which presents a matrix of `r+1` x `r+1` plots, where `r` represents the number of regressors (4 in this case).

```
pairs(savings[ , c( 'sr', 'pop15', 'pop75', 'dpi', 'ddpi' )])
```

Let's focus on the first row of the `pairs` output. On the y axis of all 4 graphs, the values of `sr`, which is the response variable, are plotted against `pop15`, `pop75`, `dpi` and `ddpi`, which are the predictors.

It is possible to notice a linear trend of `sr` with respect to `pop75` and `ddpi`, while there is no evident trend with respect to `pop15` and `dpi`.

Observing also the other plots, we can say that `pop15` and `pop75` have a strong negative correlation; `pop75` and `dpi` present a positive linear relationship, while `pop15` and `dpi` seem to present a quadratic relationship. Finally, there do not appear

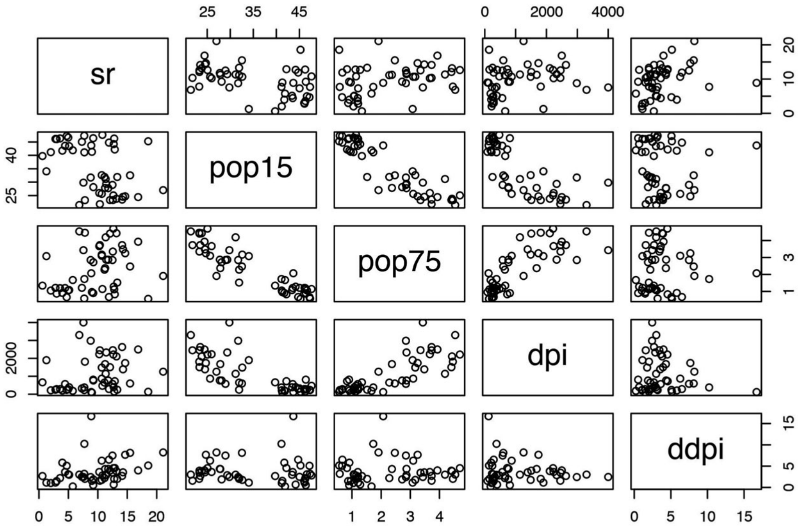

Fig. 9.2 Data visualisation

to be evident relationships between the variable `ddpi` and the other considered variables.

It is important to note that there are influential points, possible outliers (see the last column of plots related to `ddpi`).

(b) Evaluate a complete linear model. To do this, we use the `lm` command and set `sr` as the response variable.

```
g = lm( sr ~ pop15 + pop75 + dpi + ddpi, data = savings )
#g = lm( sr ~ ., savings )
summary( g )
##
## Call:
##lm(formula = sr ~ pop15 + pop75 + dpi + ddpi, data = savings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2422 -2.6857 -0.2488  2.4280  9.7509
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 28.5660865  7.3545161   3.884 0.000334 ***
## pop15       -0.4611931  0.1446422  -3.189 0.002603 **
## pop75       -1.6914977  1.0835989  -1.561 0.125530
## dpi         -0.0003369  0.0009311  -0.362 0.719173
## ddpi         0.4096949  0.1961971   2.088 0.042471 *
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
```

```
## Residual standard error: 3.803 on 45 degrees of freedom
## Multiple R-squared:  0.3385, Adjusted R-squared:  0.2797
## F-statistic: 5.756 on 4 and 45 DF,  p-value: 0.0007904

gs = summary( g )

#names( g )
```

From the complete model, we deduce that $\beta_1 \neq 0$ and $\beta_4 \neq 0$, therefore `pop15` and `ddpi` are predictive with respect to `sr`.

An indication of the goodness of the model (GOF) is given by the R^2 index (33.85%) and R^2_{adj} (27.97%). The values obtained in this model are low, we probably should consider a model reduction.

We now evaluate the estimated regression coefficients, both from the model output and by calculating them explicitly.

```
X = model.matrix(g)
round( g$coefficients, 3 ) #beta_hat
## (Intercept)       pop15       pop75         dpi        ddpi
##      28.566      -0.461      -1.691       0.000       0.410
stopifnot(all.equal(solve(t(X)%*%X)%*%t(X)%*% savings$sr,
                      as.matrix( g$coefficients ) ) )
```

We have used the `stopifnot` command to verify that the outcome of the explicit calculation of the $\hat{\beta}$ is identical (`all.equal`) to the output of the `lm` model.

We evaluate the $\hat{y}$, both manually and from the model output.

```
y_hat_man = X %*% g$coefficients

stopifnot(all.equal(y_hat_man, as.matrix(g$fitted.values)))
```

The residuals of the model can be obtained through `g$residuals`. We find `p=r+1`, with the command `g$rank`.

(c) The F statistic and the related test are represented in the model output. The test performed is the following:

$$H_0 : \beta_i = 0 \quad \forall i \quad vs \quad H_1 : \exists i | \quad \beta_i \neq 0.$$

We calculate the F test manually:

```
# SStot = Sum ( yi-ybar )^2
SS_tot = sum( ( savings$sr-mean( savings$sr ) )^2 )

# SSres = Sum ( residuals^2 )
SS_res = sum( g$res^2 )

p = g$rank # p = 5
n = dim(savings)[1] # n = 50
```

```
f_test = ( ( SS_tot - SS_res )/(p-1) )/( SS_res/(n-p) )

1 - pf( f_test, p - 1, n - p )
## [1] 0.0007903779
```

We observe that the p-value is equal to 0.0007904 (the same value we read in the last line of `summary(g)`).

We therefore conclude that for standard confidence values we reject the null hypothesis, so there is at least one regression coefficient that is not null.

(d) We manually evaluate the significance of β_1 (the parameter associated with `pop_15`), that is, we perform:

$$H_0 : \beta_1 = 0 \qquad vs \qquad H_1 : \beta_1 \neq 0.$$

There are various ways to perform this test:

- *t-test.*

```
X = model.matrix( g )

sigma2 = (summary( g )$sigma)^2
#manually
sigma2 = sum( ( savings$sr - g$fitted.values )^2 ) / ( n -p )

se_beta_1 = summary( g )$coef[ 2, 2 ]
#manually
se_beta_1 = sqrt( sigma2 * diag( solve( t( X ) %*% X ) )[2] )

T.0 = abs( ( g$coefficients[ 2 ] - 0 )/ se_beta_1 )

2*( 1-pt( T.0, n-p ) )
##       pop15
## 0.002603019
```

- *F-test on nested models.*

To perform this test, we evaluate the nested model that includes all the variables considered in the model `g` except for the variable whose effect we are evaluating. Then we perform an F test on the residuals of the two models.

The test statistic we want to evaluate is the following:

$$F_0 = \frac{\frac{SS_{res}(\text{complete\_model}) - SS_{res}(\text{nested\_model})}{df(\text{complete\_model}) - df(\text{nested\_model})}}{\frac{SS_{res}(\text{complete\_model})}{df(\text{complete\_model})}}.$$

$$F_0 \sim F(df(\text{complete\_model}) - df(\text{nested\_model}), df(\text{complete\_model}));$$

where df are the degrees of freedom.

```
g2 = lm( sr ~ pop75 + dpi + ddpi, data = savings )
summary( g2 )
##
## Call:
## lm(formula = sr ~ pop75 + dpi + ddpi, data = savings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.0577 -3.2144  0.1687  2.4260 10.0763
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.4874944  1.4276619   3.844  0.00037 ***
## pop75       0.9528574  0.7637455   1.248  0.21849
## dpi         0.0001972  0.0010030   0.197  0.84499
## ddpi        0.4737951  0.2137272   2.217  0.03162 *
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 4.164 on 46 degrees of freedom
## Multiple R-squared:  0.189,  Adjusted R-squared:  0.1361
## F-statistic: 3.573 on 3 and 46 DF,  p-value: 0.02093
SS_res_2 = sum( g2$residuals^2 )

f_test_2 = ( ( SS_res_2 - SS_res ) / 1 )/( SS_res / (n-p) )

1 - pf( f_test_2, 1, n-p )
## [1] 0.002603019
```

NB It is not the F test that is reported in the last line of `summary(g)`.

- *ANOVA between the two nested models.*

```
anova( g2, g )
## Analysis of Variance Table
##
## Model 1: sr ~ pop75 + dpi + ddpi
## Model 2: sr ~ pop15 + pop75 + dpi + ddpi
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     46 797.72
## 2     45 650.71  1    147.01 10.167 0.002603 **
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
```

We observe that the results obtained in all three ways lead us to assert that β_1 is significantly different from 0.

(e) We calculate the 95% confidence interval for the regression coefficient related to `pop75`.

The interval we want to calculate is:

$$IC_{(1-\alpha)}(\beta_2) = [\hat{\beta}_2 \pm t_{1-\alpha/2}(n-p) \cdot se(\hat{\beta}_2)]\,,$$

where $\alpha = 5\%$ and $df = n - p = 45$.

```
alpha = 0.05
t_alpha2 = qt( 1-alpha/2, n-p )
beta_hat_pop75 = g$coefficients[3]
se_beta_hat_pop75 = summary( g )[[4]][3,2]

IC_pop75 = c( beta_hat_pop75 - t_alpha2 * se_beta_hat_pop75,
              beta_hat_pop75 + t_alpha2 * se_beta_hat_pop75 )
IC_pop75
##      pop75      pop75
## -3.8739780  0.4909826
```

We observe that $IC_{(1-\alpha)}(\beta_2)$ includes 0, so we have no evidence to reject H_0 : $\beta_2 = 0$, with a confidence level of 5%. This result is in line with what was obtained in the model output (p-value equal to 12.5%).

(f) We calculate the 95% confidence interval for the regression parameter associated with `ddpi`.

```
alpha = 0.05
t_alpha2 = qt( 1-alpha/2, n-p )
beta_hat_ddpi = g$coefficients[5]
se_beta_hat_ddpi = summary( g )[[4]][5,2]

IC_ddpi = c( beta_hat_ddpi - t_alpha2 * se_beta_hat_ddpi,
             beta_hat_ddpi + t_alpha2 * se_beta_hat_ddpi )
IC_ddpi
##       ddpi       ddpi
## 0.01453363 0.80485623
```

In this case, we observe that $IC_{(1-\alpha)}(\beta_4)$ does not include 0, we have evidence to reject $H_0 : \beta_4 = 0$, at 5% confidence. However, the lower limit of the interval $IC_{(1-\alpha)}(\beta_4)$ is very close to 0. We can see in fact from the output that the p-value is equal to 4.2%, slightly less than 5%, which confirms the above.

Furthermore, the confidence interval is quite wide, given that the upper limit is 80 times the lower limit. This testifies a high level of variability relative to the effect of `ddpi` on the response variable.

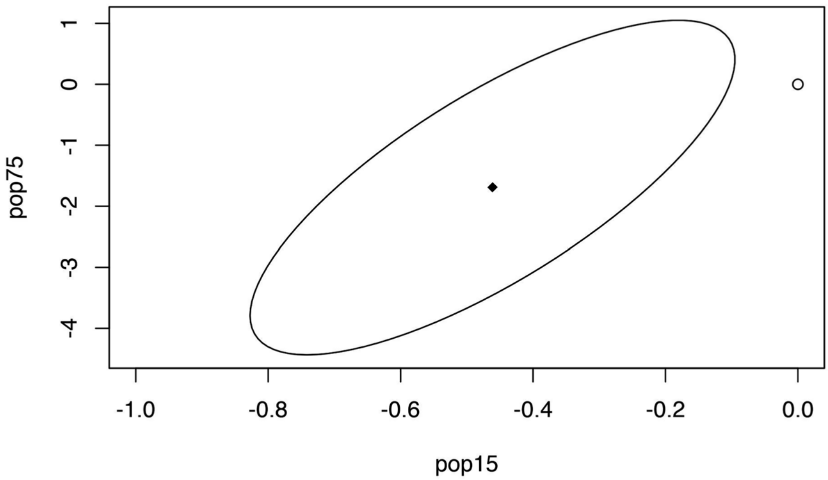

Fig. 9.3 95% confidence region for the regression coefficients associated with pop15 and pop75. The black diamond represents the centre of the ellipse, while the circle with the black border the null hypothesis that is to be tested

(g) We construct in Fig. 9.3 the confidence region at 95% for the regression coefficients associated with pop15 and pop75.

```
#help( ellipse )
plot( ellipse( g, c( 2, 3 ) ), type = "l", xlim = c( -1, 0 ) )

#vector we are testing in the null hypothesis
points( 0, 0 )
points( g$coef[ 2 ] , g$coef[ 3 ] , pch = 18, col = 1 )
```

The coordinates of the centre of the ellipse, represented by a black square, are $(\hat{\beta}_1, \hat{\beta}_2)$. The circle with the black border represents the null hypothesis tested, namely (0,0), and is outside the confidence region.

We are interested in evaluating this test:

$$H_0 : (\beta_1, \beta_2) = (0, 0) \qquad vs \qquad H_1 : (\beta_1, \beta_2) \neq (0, 0).$$

Since the point $(0, 0)$ is outside the confidence region, we reject H_0 with a level equal to 5%. This means that at least one of the two regression coefficients is different from 0.

Observation It is important to underline that the confidence region is different from the Cartesian product of the two individual confidence intervals: $IC_{(1-\alpha)}(\beta_1)$ X $IC_{(1-\alpha)}(\beta_2)$. We represent in Fig. 9.4 the Cartesian product of the marginal confidence intervals.

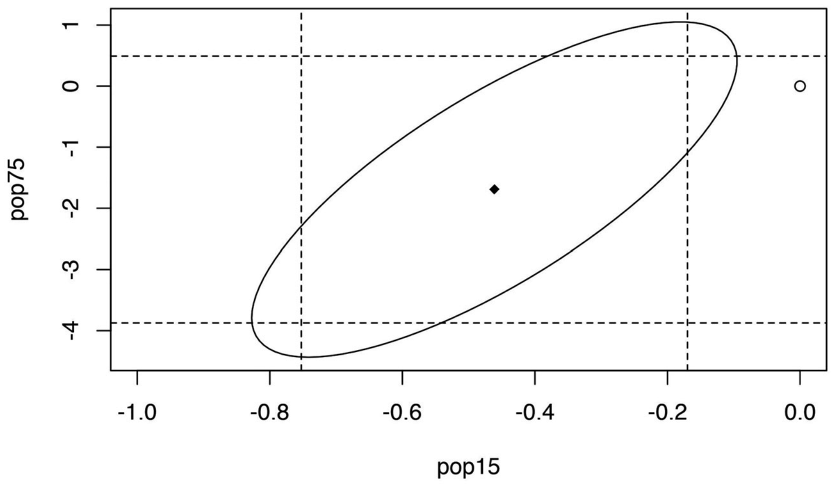

Fig. 9.4 95% confidence region for the regression coefficients associated with pop15 and pop75. The 95% confidence intervals for the individual predictors are highlighted with dashed lines

```
beta_hat_pop15 = g$coefficients[2]
se_beta_hat_pop15 = summary( g )[[4]][2,2]

IC_pop15 = c( beta_hat_pop15 - t_alpha2 * se_beta_hat_pop15,
              beta_hat_pop15 + t_alpha2 * se_beta_hat_pop15 )
IC_pop15
##      pop15      pop15
## -0.7525175 -0.1698688

plot( ellipse( g, c( 2, 3 ) ), type = "l", xlim = c( -1, 0 ) )

points( 0, 0 )
points( g$coef[ 2 ] , g$coef[ 3 ] , pch = 18 )

#new part
abline( v = c( IC_pop15[1], IC_pop15[2] ), lty = 2 )
abline( h = c( IC_pop75[1], IC_pop75[2] ), lty = 2 )
```

Observation The **0** is included in the interval $IC_{(1-\alpha)}(\beta_2)$ and is not included in the interval $IC_{(1-\alpha)}(\beta_1)$, as one might expect from the previous point.

Observation It may happen to accept the null hypothesis that you want to test, analysing the Cartesian product of the marginal ICs and to reject, considering the joint confidence region (case represented by the grey triangle in Fig. 9.5). It may happen to reject the null hypothesis that you want to test, analysing the Cartesian product of the marginal ICs and to accept, considering the joint confidence region (case represented by the grey circle in Fig. 9.5). In these ambiguous situations, we

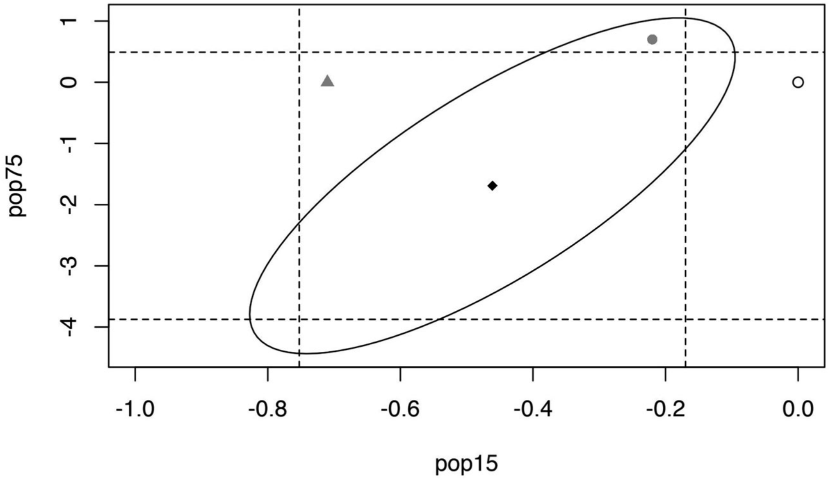

Fig. 9.5 95% confidence region for the regression coefficients associated with `pop15` and `pop75`. The 95% confidence intervals for the individual predictors are highlighted with dashed lines. The grey circle and triangle represent two possible null hypotheses that you want to test

must always refer to the joint confidence region, because it takes into account the possible dependence present between the estimators of the two tested coefficients.

```
plot( ellipse( g, c( 2, 3 ) ), type = "l", xlim = c( -1, 0 ) )

points( 0, 0 )
points( g$coef[ 2 ] , g$coef[ 3 ] , pch = 18 )

abline( v = c( IC_pop15[1], IC_pop15[2] ), lty = 2 )
abline( h = c( IC_pop75[1], IC_pop75[2] ), lty = 2 )

#new part
points( -0.22, 0.7, col = "gray60", pch = 16, lwd = 2 )
points( -0.71, 0, col = "gray60", pch = 17, lwd = 2 )
```

```
cor( savings$pop15, savings$pop75 )
## [1] -0.9084787
```

In this case, the ellipse has high eccentricity, which makes us think of a strong correlation between the two variables `pop15` and `pop75`. This intuition is confirmed by the correlation coefficient very close to -1.

Observation This intuition was also reported in the comment to the `pairs` plot.

(h) We evaluate the presence of any influential points in the dataset through the following techniques:

- *H projection matrix (leverage points).*
- *Standardised Residuals.*
- *Studentised Residuals.*
- *Cook's Distance.*
- The *leverage points* are defined as the elements of the diagonal of the projection matrix $H = X(X^T X)^{-1} X^T$.

```
X = model.matrix( g )

lev = hat( X )
round( lev, 3 )

# similarly
lev = hatvalues( g )


#manually
H = X %*% solve( t( X ) %*% X ) %*% t( X )
lev = diag( H )

#trace
sum(lev)
## [1] 5
```

Observation The trace of the matrix H ($tr(H) = \sum_i h_{ii}$) is equal to the rank of the matrix X, which is $p = r + 1$, assuming that the covariates are uncorrelated with each other and $p < n$. p is the dimension of the column space of X ($col(X)$). According to the geometric interpretation of the least squares estimate of the coefficients, H is the projection matrix onto $col(X)$. In fact, the estimates $\hat{\mathbf{y}}$ are obtained as $H\mathbf{y}$.

Rule of Thumb A data point is defined as a leverage point if:

$$h_{ii} > 2 \cdot \frac{p}{n}.$$

```
plot( g$fitted.values, lev,  xlab = 'Fitted values',
      ylab = "Leverages", pch = 16, col = 'black' )

abline( h = 2 * p/n, lty = 2, col = 1 )

watchout_points_lev = lev[ which( lev > 2 * p/n  ) ]
watchout_ids_lev = seq_along( lev )[ which( lev > 2 * p/n ) ]
points( g$fitted.values[ watchout_ids_lev ],
        watchout_points_lev,
        col = 'gray60', pch = 16 )
```

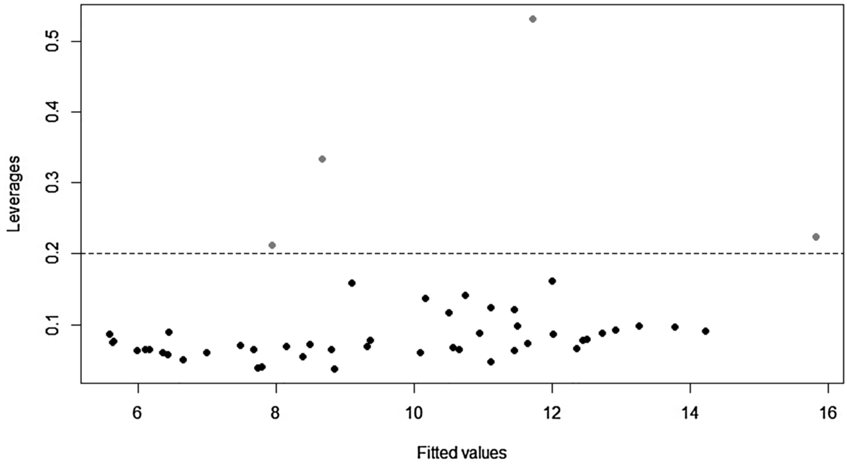

Fig. 9.6 Identification of leverage points in grey. The dashed line is y = 2p/n

```
sum( lev )      # check: sum_i hat( x )_i = r + 1
## [1] 5

lev [ lev >  2 * 5 / 50 ]
##       Ireland         Japan United States         Libya
##     0.2122363     0.2233099     0.3336880     0.5314568
sum( lev [ lev >  2 * 5 / 50 ] )
## [1] 1.300691
```

In Fig. 9.6 we therefore identify Ireland, Japan, the USA and Libya as leverage points.

We visualise the leverage points using `pairs` in Fig. 9.7 and notice that these points are indeed at the extremes of the plots.

```
colors = rep( 'black', nrow( savings ) )
colors[ watchout_ids_lev ] = rep('gray60',
                                 length( watchout_ids_lev ) )

pairs( savings[ , c( 'sr', 'pop15', 'pop75', 'dpi', 'ddpi' ) ],
       pch = 16, col = colors,
       cex = 1 + 0.5 * as.numeric( colors != 'black' ) )
```

- We now evaluate the influential points through the *standardised residuals*.

  We define the standardised residuals as:

$$r_i^{std} = \frac{y_i - \hat{y}_i}{S}.$$

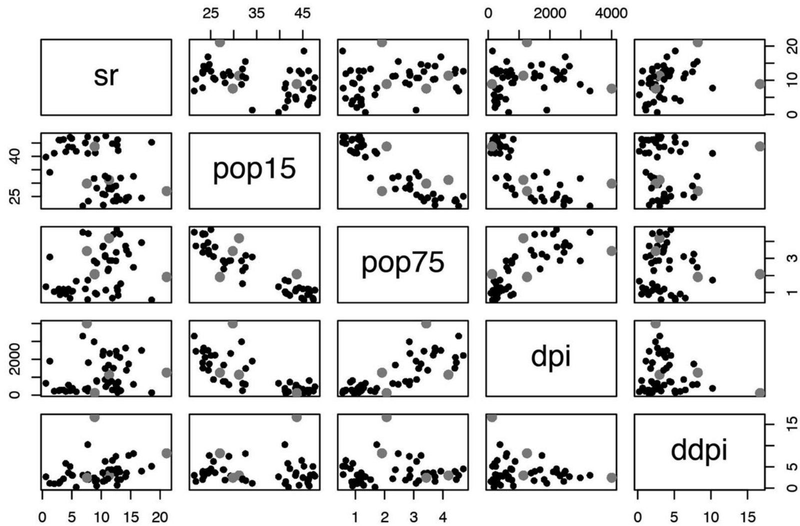

Fig. 9.7 Visualisation of the data under analysis. Influential points are represented in light grey

Rule of Thumb We define influential points as the data for which the following inequality holds:

$$|r_i^{std}| > 2.$$

We represent the standardised residuals (on the y-axis) and the $\hat{\mathbf{y}}$ (on the x-axis) and highlight the influential points in Fig. 9.8 based on the standardised residuals and the leverages.

```
gs = summary(g)
res_std = g$res/gs$sigma
watchout_ids_rstd = which( abs( res_std ) > 2 )
watchout_rstd = res_std[watchout_ids_rstd ]
watchout_rstd
##     Chile    Zambia
## -2.167486  2.564229

# Standardised residuals (not studentised)
par( xpd = T, mar = par()$mar + c(0,0,1,0))
plot( g$fitted.values, res_std,
      xlab = 'Fitted values',
      ylab = "Standardised residuals")
segments( 5, -2, 16, -2, lty = 2, col = 1 )
segments( 5, 2, 16, 2, lty = 2, col = 1 )
points( g$fitted.values[watchout_ids_rstd],
        res_std[watchout_ids_rstd],
        col = 'grey60', pch = 16 )
points( g$fitted.values[watchout_ids_lev],
```

```
       res_std[watchout_ids_lev],
       col = 'gray60', pch = 17 )
legend("top", inset=c(0,-0.2), horiz = T,
    col = rep('gray70',3),
       c('Std. Residuals', 'Leverages'),
       pch = c( 16, 17 ), bty = 'n' )
```

```
#sort( g$res/gs$sigma )
sort( g$res/gs$sigma ) [ c( 1, 50 ) ]
##     Chile    Zambia
## -2.167486  2.564229

#countries = row.names( savings )
#identify( 1:50, g$res/gs$sigma, countries )
```

The `identify` command allows you to identify the points of the graph, in fact by double-clicking on a certain point the label attached to it appears.

Representing the residuals on the y-axis simply in the order in which they appear in the dataset, allows us to say if there is a particular trend with respect to the sampling order. This graph is represented in Fig. 9.9.

```
plot( g$res/gs$sigma, xlab = "Order of appearance",
      ylab = "Standardised residuals" )
```

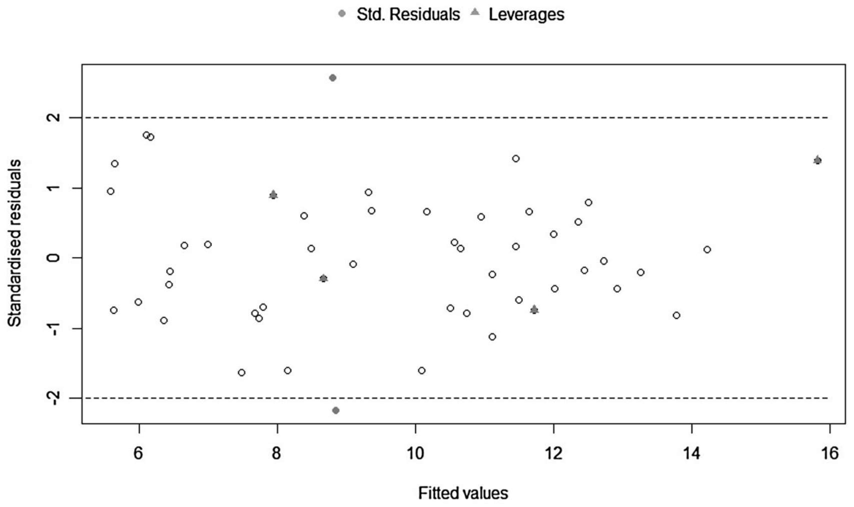

Fig. 9.8 Representation of the standardised residuals. The grey circles represent the influential points identified according to the criterion of standardised residuals. The grey triangles represent the influential points identified according to the leverages

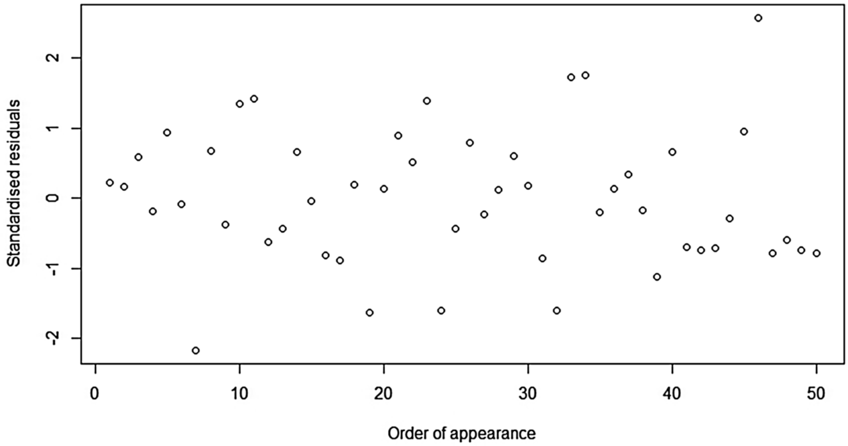

Fig. 9.9 Standardised Residuals in order of appearance in the dataset

```
summary( g$res/gs$sigma )
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.16749 -0.70628 -0.06543  0.00000  0.63850  2.56423

#countries = row.names( savings )
#identify( 1:50, g$res/gs$sigma, countries )
```

We do not identify any particular trend with respect to the sampling order.

- We define the studentised residuals r_i as:

$$r_i = \frac{\hat{\varepsilon}_i}{S \cdot \sqrt{(1 - h_{ii})}}.$$

It can be shown that r_i are distributed as $t(n-p)$. Since the distribution of r_i is known, we can calculate the p-value to test if the i-th data point is an influential point. In reality, we would like to simultaneously test if there are multiple influential points, hence it is important to adjust the significance level of the tests. There are various methods of correcting the significance level in the case of multiple tests, including the Bonferroni correction.

```
gs = summary( g )

gs$sigma
## [1] 3.802669

#manually
stud = g$residuals / ( gs$sigma * sqrt( 1 - lev ) )
```

```
#automatically
stud = rstandard( g )

watchout_ids_stud = which( abs( stud ) > 2 )
watchout_stud = stud[ watchout_ids_stud ]
watchout_stud
##     Chile    Zambia
## -2.209074  2.650915

par( xpd = T, mar = par()$mar + c(0,0,1,0))
plot( g$fitted.values, stud,
      ylab = "Studentised residuals",
      xlab = "Fitted values", pch = 16 )
points( g$fitted.values[watchout_ids_stud],
        stud[watchout_ids_stud],
        col = 'gray70', pch = 16 )
points( g$fitted.values[watchout_ids_rstd],
        stud[watchout_ids_rstd],
        col = 'gray70', pch = 17 )
points( g$fitted.values[watchout_ids_lev],
        stud[watchout_ids_lev],
        col = 'gray70', pch = 18 )
segments( 5, -2, 16, -2, lty = 2, col = 1 )
segments( 5, 2, 16, 2, lty = 2, col = 1 )
legend( "top", inset=c(0,-0.2),
        horiz = T, xpd = T, col = rep('gray70',3),
        c('Studentised Res.',
         'Standardised Res.', 'Leverages'),
       pch = c( 16, 17,18 ), bty = 'n' )
```

In Fig. 9.10, Chile and Zambia are identified as influential points.

In the graph, we do not identify any pink points (influential points according to the studentised residuals), because the studentised and standardised residuals identify the same influential points.

- Cook's distance is defined as follows:

$$C_i = \frac{r_i^2}{p} \cdot \left[\frac{h_{ii}}{1 - h_{ii}}\right];$$

  where r_i are the studentised residuals. We observe that this measure is a combination of the concept of leverage point (through h_{ii}) and the concept of influential point given by the residuals (through r_i).

Rule of Thumb A point is defined as influential, if the following inequality holds:

$$C_i > \frac{4}{n - p}.$$

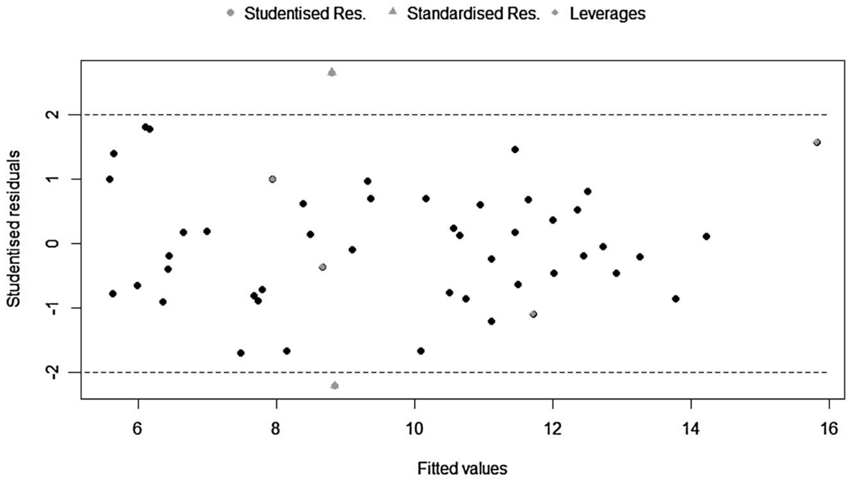

Fig. 9.10 Representation of the studentised residuals. Those data identified as influential points based on the leverages (grey diamonds), standardised residuals (grey triangles) and studentised residuals (grey circles) are highlighted

We represent in Fig. 9.11 Cook's distance for each point and highlight in light grey the points defined as influential (i.e., those exceeding the threshold y = 4/(n − p)).

```
Cdist = cooks.distance( g )

watchout_ids_Cdist = which( Cdist > 4/(n-p) )
watchout_Cdist = Cdist[ watchout_ids_Cdist ]
watchout_Cdist
##      Japan     Zambia      Libya
## 0.14281625 0.09663275 0.26807042

plot( g$fitted.values, Cdist, pch=16, xlab='Fitted values',
      ylab = 'Cook\'s distance' )
points( g$fitted.values[ watchout_ids_Cdist ],
        Cdist[ watchout_ids_Cdist ],
        col = 'gray70', pch = 16 )
abline( h = 4/(n-p), lty = 2, col = 1 )
```

In Fig. 9.11, we identify as influential points according to Cook's distance: Japan, Zambia and Libya.

(i) One way to directly and effectively evaluate influential points in the dataset is given by the command `influence.Plot`. The graph depicts the studentised residuals on the y-axis, the leverages (h_{ii}) on the x-axis and each point of the graph is depicted as a circle, the radius of which is proportional to Cook's distance.

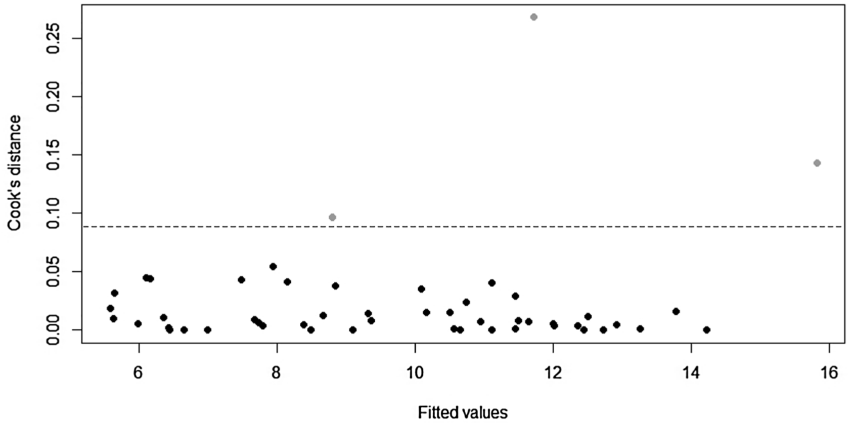

Fig. 9.11 Representation of Cook's distance for each statistical unit. The dashed line is y = 4/(n-p). The grey points are influential points according to Cook's distance

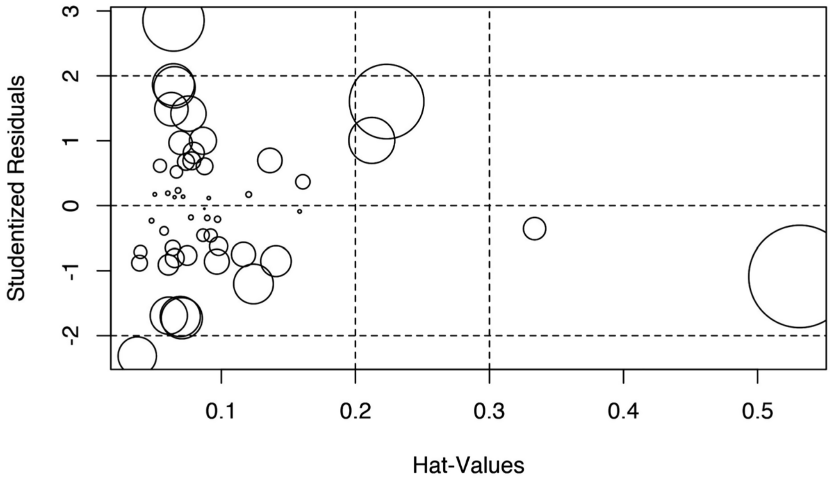

Fig. 9.12 Influence plot

```
influencePlot( g, id=list(method="identify"))
```

In Fig. 9.12 the influence plot of the dataset under examination is represented and Zambia, Japan, USA, Libya and Chile are highlighted as influential points.

Another technique to get an immediate idea of the influential points present in the graph, is to apply the command `influential.measures`, which represents in matrix form various methods of diagnosing influential points (such as h_{ii} and Cook's distance).

The DFBETAs (first r columns of the matrix) represent the impact of the single statistical unit in the estimation of $\boldsymbol{\beta}$. In particular, the DFBETA associated with the regressor j is:

$$\frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{\hat{\sigma}^2_{(i)}(X^T X)^{-1}_{jj}}};$$

where the subscript (i) indicates that we are neglecting the i-th observation.

The DFFITs (column $r + 1$) represent the impact of the single statistical unit in the estimation of $\hat{y}$. In particular, the DFFIT associated with observation i is:

$$\frac{\hat{y}_i - \hat{y}_{i(i)}}{\hat{\sigma}^2_{(i)}\sqrt{h_{ii}}}.$$

The larger the values of DFBETAs and DFFITs associated with the i-th observation, the more inclined we are to declare the i-th observation as an influential point.

Data that are anomalous according to all criteria are marked with an asterisk (Chile, USA, Zambia and Libya in this case).

```
infl_point_overview = influence.measures( g )
summary( infl_point_overview )
## Potentially influential observations of
## lm(formula = sr~pop15 + pop75 + dpi + ddpi,data = savings):
##
##                dfb.1_ dfb.pp15 dfb.pp75 dfb.dpi dfb.ddpi
## Chile          -0.20   0.13     0.22    -0.02    0.12
## United States  0.07  -0.07     0.04    -0.23   -0.03
## Zambia          0.16  -0.08    -0.34     0.09    0.23
## Libya           0.55  -0.48    -0.38    -0.02   -1.02_*
##                dffit   cov.r  cook.d   hat
## Chile          -0.46   0.65_*   0.04   0.04
## United States -0.25   1.66_*   0.01   0.33_*
## Zambia          0.75   0.51_*   0.10   0.06
## Libya          -1.16_* 2.09_*   0.27   0.53_*
```

(j) To evaluate the effect of influential points on the outcome of the model, two quantities can be considered:

- The variation of $\hat{\beta}$ in the case of evaluating a model using the entire dataset and in the case of evaluating a model using the entire dataset minus the $i-$th observation:

$$\left|\frac{\hat{\beta} - \hat{\beta}_{(i)}}{\hat{\beta}}\right|.$$

- The variation of the estimated responses $\hat{\mathbf{y}}$ in the case of evaluating a model using the entire dataset and in the case of evaluating a model using the entire dataset minus the i-th observation:

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)} = X^T(\hat{\beta} - \hat{\beta}_{(i)}).$$

Let's now evaluate how the model coefficients vary, in the case where the influential points are removed from the dataset according to the values of h_{ii} and Cook's distance.

- *Leverage points.*

```
gl = lm( sr ~ pop15 + pop75 + dpi + ddpi, savings,
         subset = ( lev < 0.2 ) )
summary( gl )
##
## Call:
## lm(formula = sr ~ pop15 + pop75 + dpi + ddpi, data=savings,
##     subset = (lev < 0.2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.9632 -2.6323  0.1466  2.2529  9.6687
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.221e+01  9.319e+00   2.384   0.0218 *
## pop15       -3.403e-01  1.798e-01  -1.893   0.0655 .
## pop75       -1.124e+00  1.398e+00  -0.804   0.4258
## dpi         -4.499e-05  1.160e-03  -0.039   0.9692
## ddpi         5.273e-01  2.775e-01   1.900   0.0644 .
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 3.805 on 41 degrees of freedom
## Multiple R-squared:  0.2959, Adjusted R-squared:  0.2272
## F-statistic: 4.308 on 4 and 41 DF,  p-value: 0.005315

abs( ( g$coefficients - gl$coefficients ) / g$coefficients )
## (Intercept)       pop15       pop75         dpi        ddpi
##   0.2223914   0.2622274   0.3353998   0.8664714   0.2871002
```

The leverage points significantly influence the estimates, in fact, a variation of at least 22% (relative variation to $\hat{\beta}_0$) is recorded.

- *Cook's distance.*

```
#id_to_keep = (1:n)[ - watchout_ids_Cdist ]
id_to_keep = !( 1:n %in% watchout_ids_Cdist )

gl = lm( sr ~ pop15 + pop75 + dpi + ddpi,
```

```
    savings[ id_to_keep, ]  )

abs( ( gl$coef - g$coef )/g$coef )
## (Intercept)       pop15       pop75         dpi        ddpi
## 0.305743704 0.339320881 0.820854095 0.642906116 0.009976742
```

In this case too, there is a strong variation of the estimated coefficients, except for ddpi.

(k) We evaluate the homoscedasticity of the residuals through scatterplot analysis.

We start by evaluating the homoscedasticity through a scatterplot, Fig. 9.13, where $\hat{\varepsilon}$ are reported on the y-axis and $\hat{\mathbf{y}}$ are reported on the x-axis.

```
plot( g$fit, g$res, xlab = "Fitted values",
      ylab = "Residuals",
      pch = 16 )
abline( h = 0, lwd = 2, lty = 2, col = 1 )
```

In Fig. 9.13 we observe that the residuals are quite scattered around 0, but there are extreme points in the graph. It would be appropriate to redo this analysis after setting the model on a subset of the dataset that does not contain leverage points.

(l) We evaluate the normality of the residuals through:

- QQ-plot.
- Shapiro-Wilk test.

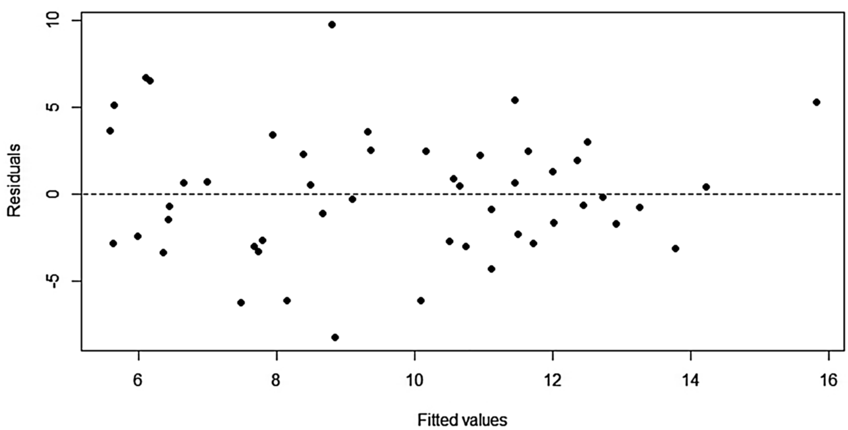

Fig. 9.13 Scatterplot of residuals

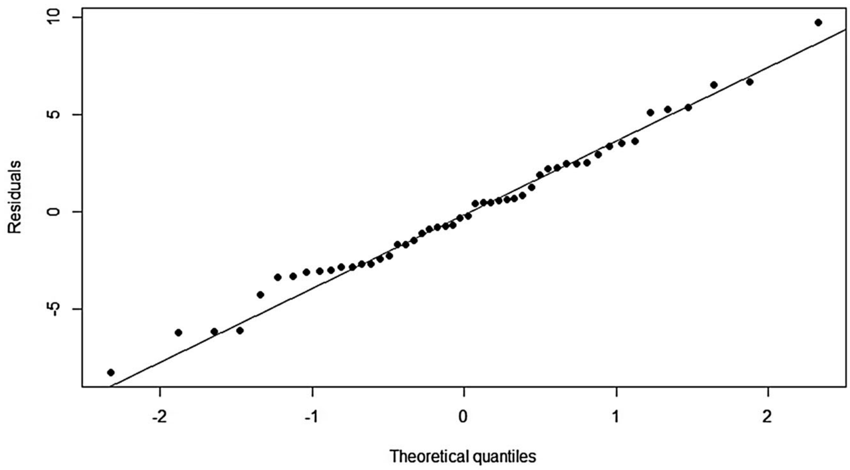

Fig. 9.14 QQ-plot of residuals

```
qqnorm( g$res, ylab = "Residuals",
        xlab = "Theoretical quantiles",
        main = NULL, pch = 16 )
qqline( g$res )
```

```
shapiro.test( g$res )
##
##  Shapiro-Wilk normality test
##
## data:  g$res
## W = 0.98698, p-value = 0.8524
```

From the QQ-plot in Fig. 9.14 we observe that the empirical quantiles of the residuals (reported on the y-axis) are well approximated by the theoretical quantiles of a standard Gaussian (reported on the x-axis).

From the Shapiro test we obtain a p-value of 0.8524, so we can accept the null hypothesis, i.e. the normality of the residuals.

9.6

(a) We import the dataset and visualise it in Fig. 9.15.

```
load("data_es2.RData")
pairs(data_es2)
```

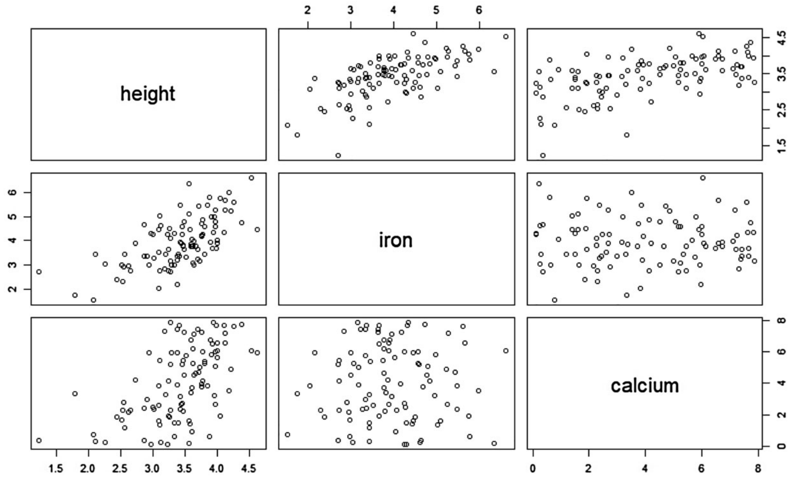

Fig. 9.15 Data visualisation

From the `pairs` plot we infer a linear relationship between height and iron and between height and calcium. There seems to be no correlation between iron and calcium.

(b) We evaluate a multiple linear regression model to answer the researchers:

```
mod = lm(altezza ~ ferro + calcio, data = data_es2)

summary(mod)
##
## Call:
## lm(formula = altezza ~ ferro + calcio, data = data_es2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2277 -0.2160  0.0025  0.2415  0.7597
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.40856    0.13889   10.14   <2e-16 ***
## ferro        0.36849    0.03148   11.71   <2e-16 ***
## calcio       0.13818    0.01353   10.21   <2e-16 ***
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.3148 on 97 degrees of freedom
## Multiple R-squared:  0.7156, Adjusted R-squared:  0.7098
## F-statistic: 122.1 on 2 and 97 DF,  p-value: < 2.2e-16
```

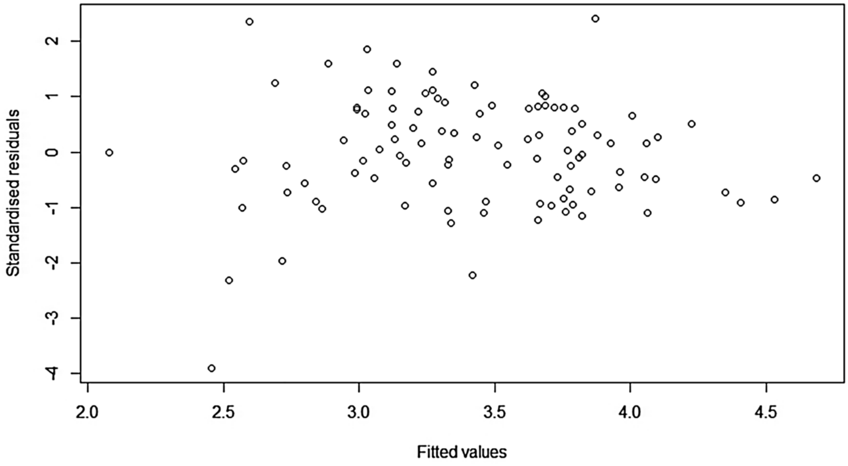

Fig. 9.16 Standardised residuals

The model is quite good ($R^2 = 71.56\%$) and both predictors are significant. Moreover, the estimated βs are positive, so as the percentages of iron and calcium increase, there is an increase in the height of the rock.

(c) We evaluate the validity of the model assumptions:

- Homoscedasticity.
- Normality.

```
mod_res = mod$residuals/summary(mod)$sigma
plot( mod$fitted, mod_res,
      xlab = 'Fitted values',
      ylab = 'Standardised residuals'  )
```

From the scatterplot of residuals in Fig. 9.16, we infer that The assumption of homoscedasticity is respected, although there seem to be some leverage points. Before investigating these extreme points, we proceed to verify normality.

```
qqnorm( mod_res, ylab = "Standardised residuals",
        xlab = "Theoretical quantiles",
        main = NULL, pch = 16  )
qqline( mod_res, col = 1, lty = 2 )
```

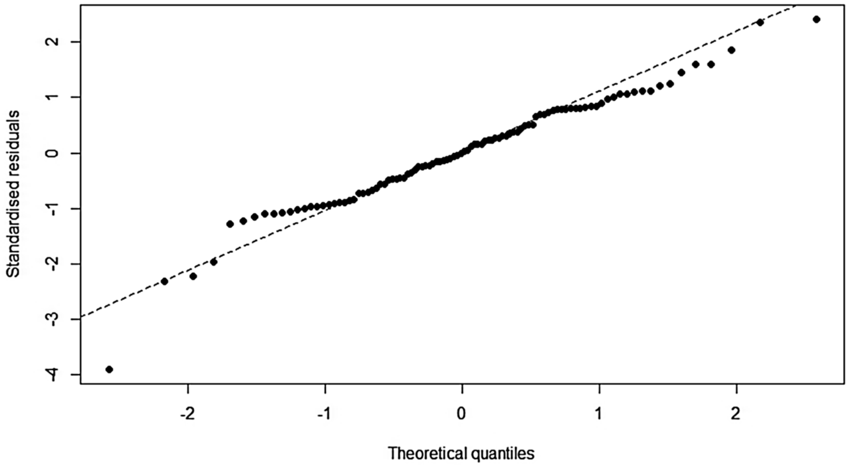

Fig. 9.17 QQ-plot of residuals

```
shapiro.test( mod_res )
##
##  Shapiro-Wilk normality test
##
## data:  mod_res
## W = 0.97051, p-value = 0.0242
```

Observing the residuals in Fig. 9.17, we notice that the assumption of normality is violated, as there are heavy and negative tails. Furthermore, the p-value of the Shapiro test is less than 5%. Therefore we conclude that the assumption of normality is violated.

(d) Given that the assumption of normality is violated and that the response variable can only take positive values, we evaluate the Box-Cox transformation of the response variable.

```
b = boxcox( height ~ iron + calcium,
            lambda = seq(0.3, 5, by=0.01), data = data_es2)
```

```
names(b)
## [1] "x" "y"
#y likelihood evaluation
#x lambda evaluated
best_lambda_ind = which.max( b$y )
best_lambda = b$x[ best_lambda_ind ]
best_lambda
## [1] 2.28
```

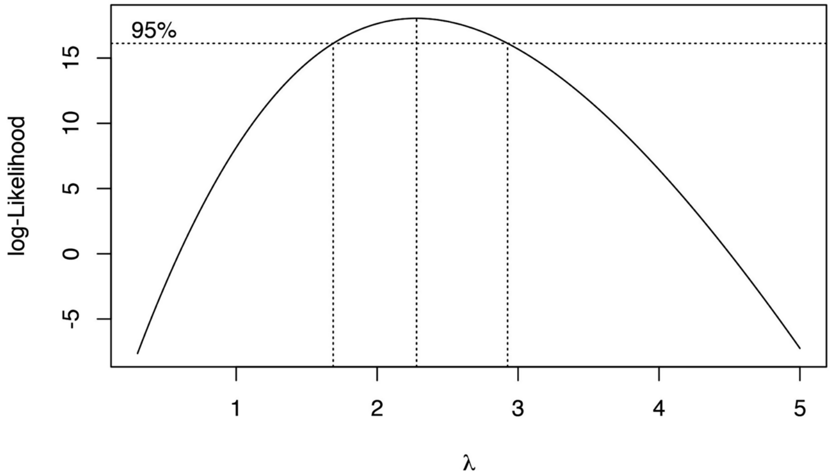

Fig. 9.18 Box-Cox type transformation

The best transformation that emerges from Fig. 9.18 is the one associated with the maximum of the curve. The estimates are obtained through maximum likelihood. According to this method, the best transformation is associated with $\lambda = 2.28$. Despite this, we would like an interpretable transformation, so we opt for $\lambda = 2$ and calculate the square of the $\mathbf{y}$.

We retrace the analyses.

```
mod1 = lm( ( height )^2 ~ iron + calcium, data = data_es2 )
summary(mod1)
##
## Call:
## lm(formula = (height)^2 ~ iron + calcium, data = data_es2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6364 -1.3311 -0.0558  1.3655  6.4870
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.01712    0.84971  -1.197    0.234
## iron         2.39617    0.19256  12.443   <2e-16 ***
## calcium      0.89183    0.08276  10.776   <2e-16 ***
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 1.926 on 97 degrees of freedom
## Multiple R-squared:  0.7385, Adjusted R-squared:  0.7332
## F-statistic:   137 on 2 and 97 DF,  p-value: < 2.2e-16

mod1_res = mod1$residuals/summary( mod1 )$sigma
```

```
plot( mod1$fitted, mod1_res,
      xlab = 'Fitted values',
      ylab = 'Standardised residuals'  )
```

In Fig. 9.19 we can see that the residuals have a cloud-like behaviour around zero, so the assumption of homoscedasticity is valid. However, there remains an influential point that should be further investigated.

```
qqnorm( mod1_res,  ylab = "Standardised residuals",
        xlab = "Theoretical quantiles",
        main = NULL, pch = 16  )
abline( 0, 1, col = 1, lty = 2)
```

```
shapiro.test( residuals( mod1 ) )
##
##  Shapiro-Wilk normality test
##
## data:  residuals(mod1)
## W = 0.98644, p-value = 0.401
```

The QQ-plot in Fig. 9.20 and the Shapiro test confirm the normality of the residuals.

9.7

(a) We represent the data using the `pairs` command in Fig. 9.21.

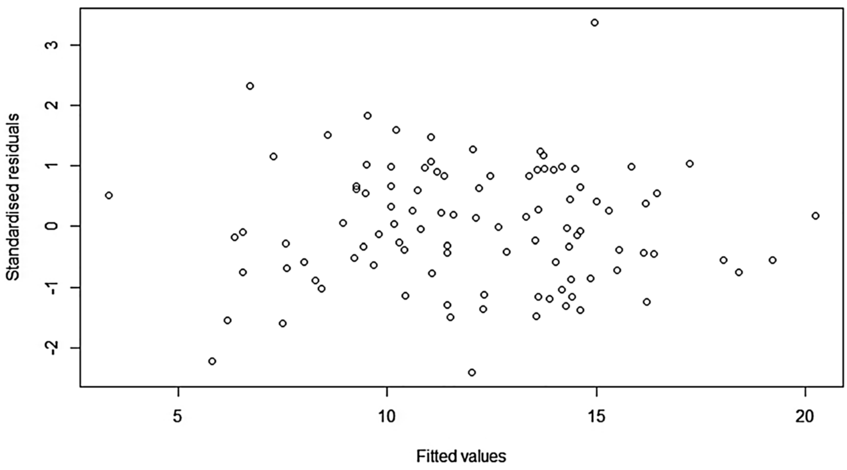

Fig. 9.19 Scatterplot of standardised residuals

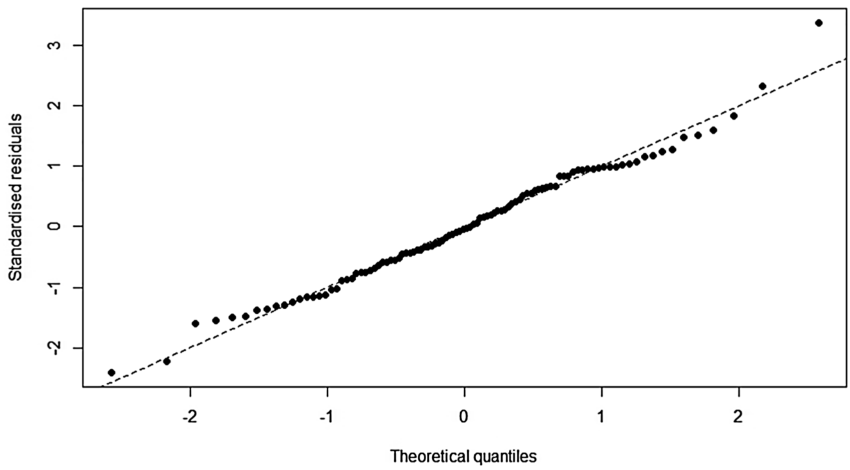

Fig. 9.20 QQ-plot of standardised residuals

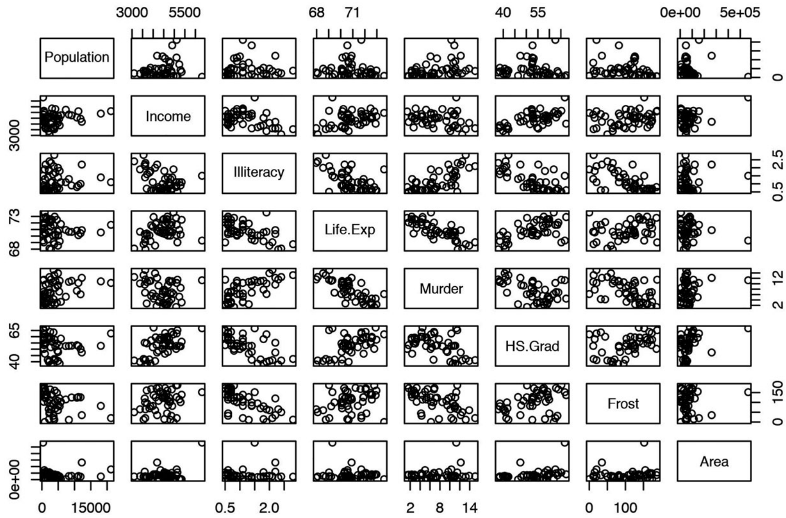

Fig. 9.21 Data visualisation

```
data( state )
statedata = data.frame( state.x77, row.names = state.abb,
                        check.names = T )

#head( statedata )

pairs( statedata )
```

```
X = statedata [ , -4 ] #we do not consider the response variable
cor( X )
##              Population     Income  Illiteracy     Murder
## Population  1.00000000  0.2082276  0.10762237  0.3436428
## Income      0.20822756  1.0000000 -0.43707519 -0.2300776
## Illiteracy  0.10762237 -0.4370752  1.00000000  0.7029752
## Murder      0.34364275 -0.2300776  0.70297520  1.0000000
## HS.Grad    -0.09848975  0.6199323 -0.65718861 -0.4879710
## Frost      -0.33215245  0.2262822 -0.67194697 -0.5388834
## Area        0.02254384  0.3633154  0.07726113  0.2283902
##                 HS.Grad     Frost        Area
## Population -0.09848975 -0.3321525 0.02254384
## Income      0.61993232  0.2262822 0.36331544
## Illiteracy -0.65718861 -0.6719470 0.07726113
## Murder     -0.48797102 -0.5388834 0.22839021
## HS.Grad     1.00000000  0.3667797 0.33354187
## Frost       0.36677970 1.0000000 0.05922910
## Area        0.33354187 0.0592291 1.00000000
```

Observation It is important to pay attention to *spurious* correlations between two variables, i.e., entirely random correlations, in which there is no plausible logical-causal mechanism that relates them. On these sites, you can find some amusing examples of spurious correlations:

http://www.tylervigen.com/spurious-correlations
http://guessthecorrelation.com

From the graphical representation of the data, we intuit an evident positive linear dependence between the outcome variable `Life Exp` and `HS.Grad`, `Frost` and `Income` (although the last two less evidently). There is also a negative linear dependence between `Life Exp` and the variables `Murder` and `Illiteracy`.

There may be some influential points present.

(b) We investigate the complete model.

```
g = lm( Life.Exp ~ ., data = statedata )
summary( g )
##
## Call:
## lm(formula = Life.Exp ~ ., data = statedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.48895 -0.51232 -0.02747  0.57002  1.49447
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.094e+01  1.748e+00  40.586  < 2e-16 ***
## Population   5.180e-05  2.919e-05   1.775   0.0832 .
## Income      -2.180e-05  2.444e-04  -0.089   0.9293
```

```
## Illiteracy   3.382e-02  3.663e-01   0.092   0.9269
## Murder      -3.011e-01  4.662e-02  -6.459 8.68e-08 ***
## HS.Grad      4.893e-02  2.332e-02   2.098   0.0420 *
## Frost       -5.735e-03  3.143e-03  -1.825   0.0752 .
## Area        -7.383e-08  1.668e-06  -0.044   0.9649
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.7448 on 42 degrees of freedom
## Multiple R-squared:  0.7362, Adjusted R-squared:  0.6922
## F-statistic: 16.74 on 7 and 42 DF,  p-value: 2.534e-10
```

We observe that the model represents the data well (R^2 equal to 0.7362), although only the variables `murder` and `HS.grad` seem to be significant. An increase in the murder rate (`murder`) leads to a decrease in life expectancy (`Life Exp`). This statement is motivated by the fact that $\hat{\beta}_{murder} = -0.3$ is negative. On the contrary $\hat{\beta}_{HS.grad} = 0.048$ is positive, so an increase in the percentage of high school graduates (`HS.grad`) leads to an increase in `Life Exp`.

(c) We verify the hypothesis of homoscedasticity.

```
plot( g$fitted, g$residuals/summary(g)$sigma,
      xlab = 'Fitted values',
      ylab = 'Standardised residuals' )
```

In Fig. 9.22 we observe that the residuals are scattered around zero, so the hypothesis of homoscedasticity seems to be respected.

We evaluate the normality of the residuals.

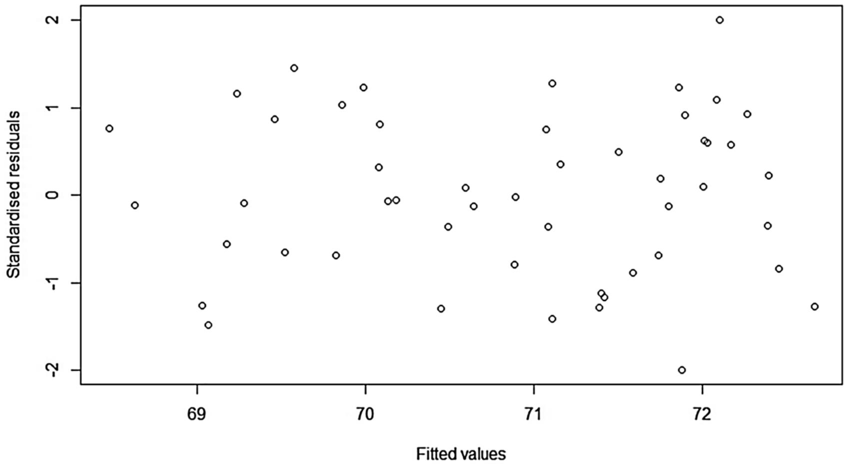

Fig. 9.22 Scatterplot of residuals

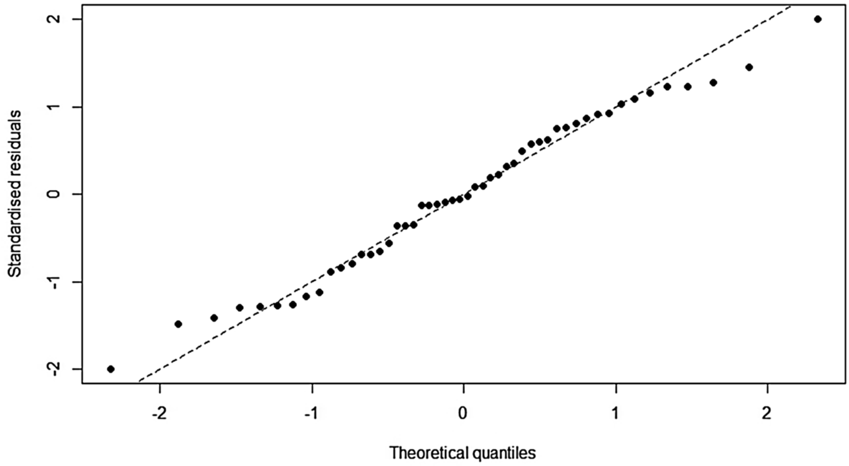

Fig. 9.23 QQ-plot of residuals

```
qqnorm( g$residuals/summary(g)$sigma,
        ylab = "Standardised residuals",
        xlab = "Theoretical quantiles",
        main = NULL, pch = 16 )
abline( 0, 1, col = 1, lty = 2 )
```

```
shapiro.test( residuals( g ) )
##
##  Shapiro-Wilk normality test
##
## data:  residuals(g)
## W = 0.97926, p-value = 0.5212
```

From the QQ-plot in Fig. 9.23, we observe that the empirical quantiles of the standardised residuals are very close to the theoretical quantiles of a standard normal, moreover the p-value of the Shapiro test is much higher than 5%, therefore we conclude that the residuals are normal.

(d) We proceed with a selection of the variables in the model with:

- Manual backward selection.
- Automatic selection.

Manual Backward Selection

At each step, we remove the predictor associated with the lowest significance (i.e., highest p-value).

We select the model that has all predictors with p-value below 5%.

We therefore start by removing Area.

```
# remove Area
g1 = update( g, . ~ . - Area )
summary( g1 )
##
## Call:
## lm(formula = Life.Exp ~ Population + Income + Illiteracy +
##     + Murder + HS.Grad + Frost, data = statedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49047 -0.52533 -0.02546  0.57160  1.50374
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.099e+01  1.387e+00  51.165  < 2e-16 ***
## Population   5.188e-05  2.879e-05   1.802   0.0785 .
## Income      -2.444e-05  2.343e-04  -0.104   0.9174
## Illiteracy   2.846e-02  3.416e-01   0.083   0.9340
## Murder      -3.018e-01  4.334e-02  -6.963 1.45e-08 ***
## HS.Grad      4.847e-02  2.067e-02   2.345   0.0237 *
## Frost       -5.776e-03  2.970e-03  -1.945   0.0584 .
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.7361 on 43 degrees of freedom
## Multiple R-squared:  0.7361, Adjusted R-squared:  0.6993
## F-statistic: 19.99 on 6 and 43 DF,  p-value: 5.362e-11
#help('update')
#help('update.formula')
```

We remove Illiteracy.

```
# remove Illiteracy
g2 = update( g1, . ~ . - Illiteracy )
summary( g2 )
##
## Call:
## lm(formula = Life.Exp ~ Population + Income + Murder +
##     + HS.Grad + Frost, data = statedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4892 -0.5122 -0.0329  0.5645  1.5166
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.107e+01  1.029e+00  69.067  < 2e-16 ***
## Population   5.115e-05  2.709e-05   1.888   0.0657 .
## Income      -2.477e-05  2.316e-04  -0.107   0.9153
## Murder      -3.000e-01  3.704e-02  -8.099 2.91e-10 ***
## HS.Grad      4.776e-02  1.859e-02   2.569   0.0137 *
```

```
## Frost       -5.910e-03  2.468e-03  -2.395   0.0210 *
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.7277 on 44 degrees of freedom
## Multiple R-squared:  0.7361, Adjusted R-squared:  0.7061
## F-statistic: 24.55 on 5 and 44 DF,  p-value: 1.019e-11
```

We remove Income.

```
# Remove Income
g3 = update( g2, . ~ . - Income )
summary( g3 )
##
## Call:
## lm(formula = Life.Exp ~ Population + Murder + HS.Grad +
##     + Frost, data = statedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47095 -0.53464 -0.03701  0.57621  1.50683
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.103e+01  9.529e-01  74.542  < 2e-16 ***
## Population   5.014e-05  2.512e-05   1.996  0.05201 .
## Murder      -3.001e-01  3.661e-02  -8.199 1.77e-10 ***
## HS.Grad      4.658e-02  1.483e-02   3.142  0.00297 **
## Frost       -5.943e-03  2.421e-03  -2.455  0.01802 *
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.7197 on 45 degrees of freedom
## Multiple R-squared:  0.736,  Adjusted R-squared:  0.7126
## F-statistic: 31.37 on 4 and 45 DF,  p-value: 1.696e-12
```

We remove Population.

```
# remove Population
g4 = update( g3, . ~ . - Population )
summary( g4 )
##
## Call:
## lm(formula = Life.Exp ~ Murder + HS.Grad + Frost,
##         data = statedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5015 -0.5391  0.1014  0.5921  1.2268
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 71.036379   0.983262  72.246  < 2e-16 ***
```

```
## Murder      -0.283065   0.036731  -7.706 8.04e-10 ***
## HS.Grad      0.049949   0.015201   3.286  0.00195 **
## Frost       -0.006912   0.002447  -2.824  0.00699 **
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.7427 on 46 degrees of freedom
## Multiple R-squared:  0.7127, Adjusted R-squared:  0.6939
## F-statistic: 38.03 on 3 and 46 DF,  p-value: 1.634e-12
```

The decision to remove or retain `Population` should also be guided by the interpretation and importance of the variable. Without additional information, we can remove it since this leads to a slight decrease in R^2 (from 0.736 to 0.713).

Automatic Selection

To perform an automatic model selection, the command `step` is used. Various criteria can be used to proceed in the selection:

- AIC.
- BIC.
- R^2_{adj}.

Furthermore, a selection can be used:

- backward (start from the complete model and reduce);
- forward (start from the model with only intercept and add variables).

The default criterion used is AIC and the method is backward.

```
#help( step )
g = lm( Life.Exp ~ ., data = statedata )

step( g )
## Start:  AIC=-22.18
## Life.Exp ~ Population + Income + Illiteracy + Murder +
##     + HS.Grad + Frost + Area
##
##              Df Sum of Sq    RSS     AIC
## - Area        1    0.0011 23.298 -24.182
## - Income      1    0.0044 23.302 -24.175
## - Illiteracy  1    0.0047 23.302 -24.174
## <none>                    23.297 -22.185
## - Population  1    1.7472 25.044 -20.569
## - Frost       1    1.8466 25.144 -20.371
## - HS.Grad     1    2.4413 25.738 -19.202
## - Murder      1   23.1411 46.438  10.305
##
## Step:  AIC=-24.18
## Life.Exp ~ Population + Income + Illiteracy + Murder +
##     + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Illiteracy  1    0.0038 23.302 -26.174
```

```
## - Income      1    0.0059 23.304 -26.170
## <none>                    23.298 -24.182
## - Population  1    1.7599 25.058 -22.541
## - Frost       1    2.0488 25.347 -21.968
## - HS.Grad     1    2.9804 26.279 -20.163
## - Murder      1   26.2721 49.570  11.569
##
## Step:  AIC=-26.17
## Life.Exp ~ Population + Income + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Income      1     0.006 23.308 -28.161
## <none>                    23.302 -26.174
## - Population  1     1.887 25.189 -24.280
## - Frost       1     3.037 26.339 -22.048
## - HS.Grad     1     3.495 26.797 -21.187
## - Murder      1    34.739 58.041  17.456
##
## Step:  AIC=-28.16
## Life.Exp ~ Population + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## <none>                    23.308 -28.161
## - Population  1     2.064 25.372 -25.920
## - Frost       1     3.122 26.430 -23.877
## - HS.Grad     1     5.112 28.420 -20.246
## - Murder      1    34.816 58.124  15.528
##
## Call:
## lm(formula = Life.Exp ~ Population + Murder + HS.Grad +
##       + Frost,  data = statedata)
##
##
## Coefficients:
## (Intercept)  Population     Murder    HS.Grad      Frost
##   7.103e+01  5.014e-05 -3.001e-01  4.658e-02  -5.943e-03

AIC( g1 )
## [1] 119.7116
AIC( g2 )
## [1] 117.7196
AIC( g3 )
## [1] 115.7326
AIC( g4 )
## [1] 117.9743
```

With the backward selection method based on AIC, the best model is $g3$, which includes `Population` + `Murder` + `HS.Grad` + `Frost`. The algorithm starts from the AIC relative to the complete model and removes at each step the variable associated with the smallest increase in AIC.

We now apply a backward model selection based on BIC.

```
g = lm( Life.Exp ~ ., data = statedata )

AIC( g )
## [1] 121.7092
BIC( g )
## [1] 138.9174

g_AIC_back = step( g, direction = "backward", k = 2 )
## Start:  AIC=-22.18
## Life.Exp ~ Population + Income + Illiteracy + Murder +
##     + HS.Grad + Frost + Area
##
##              Df Sum of Sq    RSS     AIC
## - Area        1    0.0011 23.298 -24.182
## - Income      1    0.0044 23.302 -24.175
## - Illiteracy  1    0.0047 23.302 -24.174
## <none>                    23.297 -22.185
## - Population  1    1.7472 25.044 -20.569
## - Frost       1    1.8466 25.144 -20.371
## - HS.Grad     1    2.4413 25.738 -19.202
## - Murder      1   23.1411 46.438  10.305
##
## Step:  AIC=-24.18
## Life.Exp ~ Population + Income + Illiteracy + Murder +
##     + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Illiteracy  1    0.0038 23.302 -26.174
## - Income      1    0.0059 23.304 -26.170
## <none>                    23.298 -24.182
## - Population  1    1.7599 25.058 -22.541
## - Frost       1    2.0488 25.347 -21.968
## - HS.Grad     1    2.9804 26.279 -20.163
## - Murder      1   26.2721 49.570  11.569
##
## Step:  AIC=-26.17
## Life.Exp ~ Population + Income + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Income      1     0.006 23.308 -28.161
## <none>                    23.302 -26.174
## - Population  1     1.887 25.189 -24.280
## - Frost       1     3.037 26.339 -22.048
## - HS.Grad     1     3.495 26.797 -21.187
## - Murder      1    34.739 58.041  17.456
##
## Step:  AIC=-28.16
## Life.Exp ~ Population + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## <none>                    23.308 -28.161
## - Population  1     2.064 25.372 -25.920
## - Frost       1     3.122 26.430 -23.877
```

```
## - HS.Grad     1     5.112 28.420 -20.246
## - Murder      1    34.816 58.124  15.528
g_BIC_back = step( g, direction = "backward", k = log(n) )
## Start:  AIC=-6.89
## Life.Exp ~ Population + Income + Illiteracy + Murder +
##         HS.Grad + Frost + Area
##
##              Df Sum of Sq    RSS      AIC
## - Area        1    0.0011 23.298 -10.7981
## - Income      1    0.0044 23.302 -10.7910
## - Illiteracy  1    0.0047 23.302 -10.7903
## - Population  1    1.7472 25.044  -7.1846
## - Frost       1    1.8466 25.144  -6.9866
## <none>                    23.297  -6.8884
## - HS.Grad     1    2.4413 25.738  -5.8178
## - Murder      1   23.1411 46.438  23.6891
##
## Step:  AIC=-10.8
## Life.Exp ~ Population + Income + Illiteracy + Murder
##      + HS.Grad + Frost
##
##              Df Sum of Sq    RSS      AIC
## - Illiteracy  1    0.0038 23.302 -14.7021
## - Income      1    0.0059 23.304 -14.6975
## - Population  1    1.7599 25.058 -11.0691
## <none>                    23.298 -10.7981
## - Frost       1    2.0488 25.347 -10.4960
## - HS.Grad     1    2.9804 26.279  -8.6912
## - Murder      1   26.2721 49.570  23.0406
##
## Step:  AIC=-14.7
## Life.Exp ~ Population + Income + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Income      1     0.006 23.308 -18.601
## - Population  1     1.887 25.189 -14.720
## <none>                    23.302 -14.702
## - Frost       1     3.037 26.339 -12.488
## - HS.Grad     1     3.495 26.797 -11.627
## - Murder      1    34.739 58.041  27.017
##
## Step:  AIC=-18.6
## Life.Exp ~ Population + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## <none>                    23.308 -18.601
## - Population  1     2.064 25.372 -18.271
## - Frost       1     3.122 26.430 -16.228
## - HS.Grad     1     5.112 28.420 -12.598
## - Murder      1    34.816 58.124  23.176

BIC(g1)
## [1] 135.0077
BIC(g2)
```

```
## [1] 131.1038
BIC(g3)
## [1] 127.2048
BIC(g4)
## [1] 127.5344
```

Even using a selection method based on the BIC, the best model turns out to be $g3$.

Finally, we evaluate R^2 and R^2_{adj} as selection criteria.

```
help( leaps )

# only matrix of predictors without column of 1
x = model.matrix( g ) [ , -1 ]
y = statedata$Life

adjr = leaps( x, y, method = "adjr2" )
names( adjr )
## [1] "which" "label" "size"  "adjr2"

bestmodel_adjr2_ind = which.max( adjr$adjr2 )
g$coef[ which( adjr$which[ bestmodel_adjr2_ind, ] ) + 1 ]
##    Population        Murder       HS.Grad         Frost
##  5.180036e-05 -3.011232e-01  4.892948e-02 -5.735001e-03

help( maxadjr )
maxadjr( adjr, 5 )
##     1,4,5,6   1,2,4,5,6   1,3,4,5,6   1,4,5,6,7 1,2,3,4,5,6
##       0.713       0.706       0.706       0.706       0.699
```

Even considering R^2_{adj} as a selection criterion, $g3$ turns out to be the best model, with the highest R^2_{adj} (71.26%).

```
R2 = leaps( x, y, method = "r2" )

bestmodel_R2_ind = which.max( R2$r2 )
R2$which[ bestmodel_R2_ind, ]
##    1    2    3    4    5    6    7
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

As expected, using R^2 as the selection criterion, the best model turns out to be the complete one.

Observation The variable selection process can be contaminated by the presence of influential points.

9.8

(a) We graphically represent the data in Fig. 9.24. Since there is only one predictive variable, it is not necessary to use the `pairs` command.

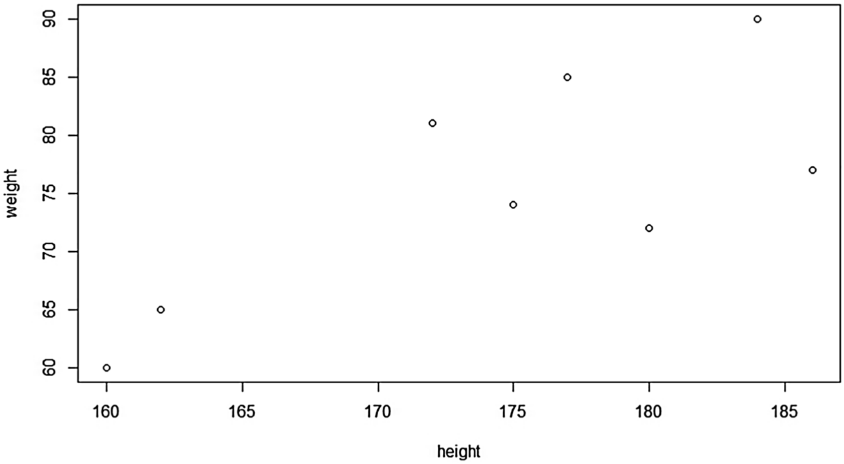

Fig. 9.24 Data visualisation

```
plot( altezza, peso )
```

The data are very few, however, a linear trend of weight with respect to height can be inferred.

(b) We set up a simple linear regression model.

```
mod = lm( peso ~ altezza )
summary( mod )
##
## Call:
## lm(formula = peso ~ altezza)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.860 -4.908 -1.244  7.097  7.518
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -62.8299    49.2149  -1.277   0.2489
## altezza       0.7927     0.2817   2.814   0.0306 *
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 7.081 on 6 degrees of freedom
## Multiple R-squared:  0.569,  Adjusted R-squared:  0.4972
## F-statistic: 7.921 on 1 and 6 DF,  p-value: 0.03058
```

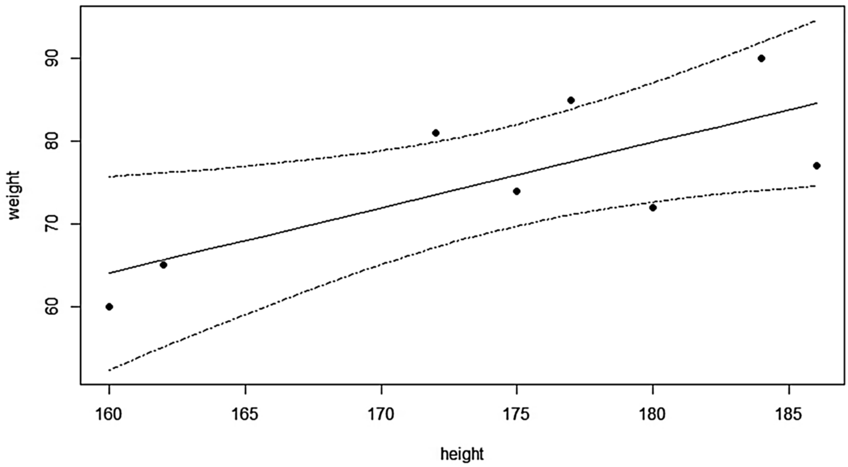

Fig. 9.25 Confidence intervals for the mean response. The continuous black line represents the values estimated through the model under consideration. The dashed black lines represent the 95% confidence bands for the mean response

The model seems mediocre, since R^2 is equal to 56.9%. Height seems significant in predicting the average weight of tomatoes (p-value of 3%). Further information is missing to better define the model.

(c) To answer the question we define a grid of values in the range of the available data (in order to have reliable estimates).

We calculate the predicted values:

$$\hat{y}_{new} = x_{new}\hat{\beta};$$

and the relative standard errors:

$$se(\mathbb{E}[y_{new}]) = \hat{S} \cdot \sqrt{x_{new}^T (X^T X)^{-1} x_{new}}.$$

We construct the graph shown in Fig. 9.25.

```
point_grid = 15
grid = seq( min( altezza ), max( altezza ),
   length.out = point_grid )

#automatically
y.pred = predict( mod, data.frame( altezza = grid ),
                  interval = "confidence", se = T )

names( y.pred )
## [1] "fit"    "se.fit"    "df"     "residual.scale"
```

```
y.pred$fit[ ,1 ] # predicted values $\hat{y}_{new}$.

y.pred$fit[ ,2 ] # LI confidence interval for $y_{new}$.


y.pred$fit[ ,3 ] # LS confidence interval for $y_{new}$.


# manually
ndata = cbind( rep( 1, length( grid ) ), grid )
y.pred_fit = ndata %*% mod$coefficients
y.pred_fit
##           [,1]
##  [1,] 64.00554
##  [2,] 65.47774
##  [3,] 66.94993
##  [4,] 68.42213
##  [5,] 69.89433
##  [6,] 71.36652
##  [7,] 72.83872
##  [8,] 74.31092
##  [9,] 75.78311
## [10,] 77.25531
## [11,] 78.72751
## [12,] 80.19971
## [13,] 81.67190
## [14,] 83.14410
## [15,] 84.61630

#standard error
y.pred$se

y.pred_se = rep( 0, point_grid )
X = model.matrix( mod )
for( i in 1:point_grid )
{
 y.pred_se[ i ] = summary( mod )$sigma * sqrt( t( ndata[i,] )
        %*% solve( t(X) %*% X ) %*% ndata[i,] )
}
y.pred_se


# n - p = 8 - 2 = 6
y.pred$df
## [1] 6

tc    = qt( 0.975, length( altezza ) - 2 )
y     = y.pred$fit[ ,1 ]
y.sup = y.pred$fit[ ,1 ] + tc * y.pred$se
y.inf = y.pred$fit[ ,1 ] - tc * y.pred$se

IC = cbind( y, y.inf, y.sup )

IC
```

```
##            y    y.inf    y.sup
## 1  64.00554 52.28376 75.72731
## 2  65.47774 54.82621 76.12926
## 3  66.94993 57.31735 76.58252
## 4  68.42213 59.73909 77.10517
## 5  69.89433 62.06616 77.72249
## 6  71.36652 64.26427 78.46877
## 7  72.83872 66.29041 79.38703
## 8  74.31092 68.09839 80.52345
## 9  75.78311 69.65227 81.91396
## 10 77.25531 70.94217 83.56845
## 11 78.72751 71.98949 85.46553
## 12 80.19971 72.83610 87.56332
## 13 81.67190 73.52812 89.81569
## 14 83.14410 74.10549 92.18271
## 15 84.61630 74.59890 94.63370
y.pred$fit
##         fit      lwr      upr
## 1  64.00554 52.28376 75.72731
## 2  65.47774 54.82621 76.12926
## 3  66.94993 57.31735 76.58252
## 4  68.42213 59.73909 77.10517
## 5  69.89433 62.06616 77.72249
## 6  71.36652 64.26427 78.46877
## 7  72.83872 66.29041 79.38703
## 8  74.31092 68.09839 80.52345
## 9  75.78311 69.65227 81.91396
## 10 77.25531 70.94217 83.56845
## 11 78.72751 71.98949 85.46553
## 12 80.19971 72.83610 87.56332
## 13 81.67190 73.52812 89.81569
## 14 83.14410 74.10549 92.18271
## 15 84.61630 74.59890 94.63370

matplot( grid, cbind( y, y.inf, y.sup ), lty = c( 1, 4, 4 ),
         col = rep( "black", 3 ), type = "l", xlab = "height",
         ylab = "weight")
points( height, weight, col = "black", pch = 16 )
```

Observation The `predict` command expects as input the data for which you want to calculate the forecast (x_{new}) in the form of a data.frame that has as column names, the same names of the predictors used in the model.

(d) Let's calculate the prediction interval for the `grid` values considered in the previous point.

In this case the standard errors are:

$$se(y_{new}) = \hat{S} \cdot \sqrt{1 + x_{new}^T (X^T X)^{-1} x_{new}}.$$

We represent the calculated intervals in Fig. 9.26.

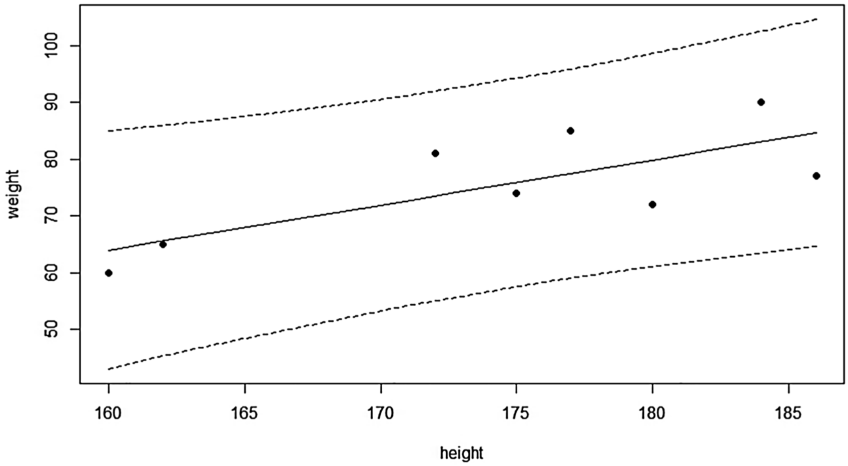

Fig. 9.26 Prediction intervals for individual observations

```
y.pred2 = predict( mod, data.frame( height = grid ),
                   interval = "prediction", se = T )

y.pred2$fit[ ,1 ] # predicted values $\hat{y}_{new}$.
y.pred2$fit[ ,2 ] # LI prediction interval for $y_{new}$.
y.pred2$fit[ ,3 ] # LS prediction interval for $y_{new}$.



#manually
ndata = cbind( rep( 1, length( grid ) ), grid )
y.pred_fit = ndata %*% mod$coefficients
y.pred_fit
##           [,1]
##  [1,] 64.00554
##  [2,] 65.47774
##  [3,] 66.94993
##  [4,] 68.42213
##  [5,] 69.89433
##  [6,] 71.36652
##  [7,] 72.83872
##  [8,] 74.31092
##  [9,] 75.78311
## [10,] 77.25531
## [11,] 78.72751
## [12,] 80.19971
## [13,] 81.67190
## [14,] 83.14410
## [15,] 84.61630
```

```
# standard error
y.pred2$se.fit

#manually
y.pred2_se = rep( 0, point_grid )

for( i in 1:point_grid )
{
 y.pred2_se[ i ] = summary( mod )$sigma *
      sqrt(  1 + t( ndata[i,] ) %*%
      solve( t(X) %*% X ) %*% ndata[i,] )
}
y.pred2_se

#In this case y.pred2_se != y.pred2$se.fit

tc    = qt( 0.975, length( height ) - 2 )
y     = y.pred2$fit[,1]
y.sup = y.pred2$fit[,1] + tc * y.pred2_se
y.inf = y.pred2$fit[,1] - tc * y.pred2_se

IP = cbind( y, y.inf, y.sup )
y.pred2$fit
##         fit      lwr       upr
## 1  64.00554 43.08632  84.92475
## 2  65.47774 45.13889  85.81658
## 3  66.94993 47.12570  86.77417
## 4  68.42213 49.04150  87.80276
## 5  69.89433 50.88134  88.90732
## 6  71.36652 52.64072  90.09232
## 7  72.83872 54.31592  91.36152
## 8  74.31092 55.90415  92.71769
## 9  75.78311 57.40375  94.16248
## 10 77.25531 58.81434  95.69628
## 11 78.72751 60.13680  97.31822
## 12 80.19971 61.37323  99.02619
## 13 81.67190 62.52680 100.81700
## 14 83.14410 63.60158 102.68662
## 15 84.61630 64.60225 104.63034

matplot( grid, y.pred2$fit, lty = c( 1, 2, 2 ),
         col = rep('black', 3),
         type = "l", xlab = "height", ylab = "weight")
points( height, weight, col = "black", pch = 16 )
```

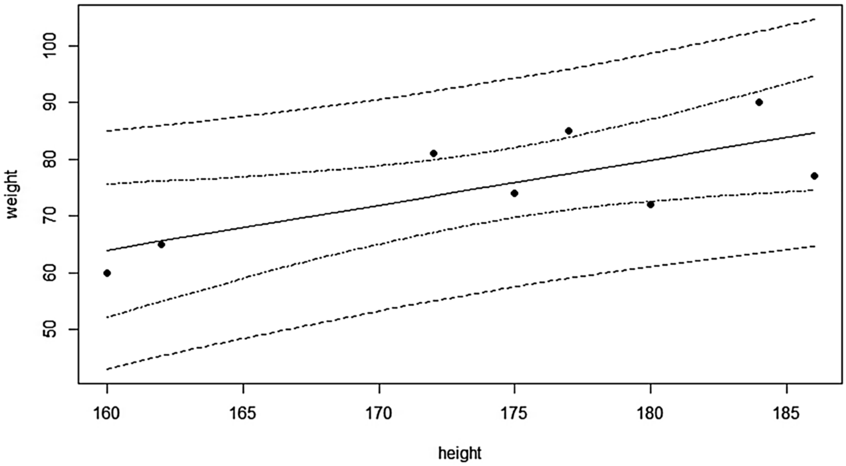

Fig. 9.27 95% confidence intervals for the mean (inner dashed line) and 95% prediction intervals for individual observations (outer dashed line)

(e) Let's compare the intervals obtained at points (c) and (d) in Fig. 9.27.

```
matplot( grid, y.pred2$fit, lty = c( 1, 2, 2 ),
         col = rep( "black", 3 ),
         type = "l", xlab = "height", ylab = "weight")
lines( grid, y.pred$fit[ , 2 ] , col = "black", lty = 4 )
lines( grid, y.pred$fit[ , 3 ] , col = "black", lty = 4 )
points( height, weight, col = "black", pch = 16 )
```

As predicted by the theory, the prediction interval is wider than the confidence interval (compare the standard errors). Moreover, all the points of the dataset fall within the prediction interval, but only some also fall within the confidence interval.

Chapter 10
Generalised Linear Models

10.1 Theory Recap

We extend regression models to the case where the dependent variable does not follow a normal distribution but belongs to the exponential family.

These models are characterised by three components:

- Y: random response variable, of which we observe N realisations $\{y_1, .., y_N\}$, whose distribution falls within the exponential family:

$$f_Y(y_i; \theta_i) = a(\theta_i)b(y_i)\exp\{y_i Q(\theta_i)\}, \qquad i \in \{1, .., N\};$$

  where θ_i is the parameter that characterises the distribution, and $Q(\theta_i)$ is called the natural parameter.
- $\eta_i = \sum_{j=1}^{r} \beta_j x_{ij}$: linear predictor.
- g: link function that connects the random response with the linear predictors. Given $\mu_i = \mathrm{e}[Y_i], \quad i = 1, .., N$, the model predicts that:

$$g(\mu_i) = \eta_i \qquad \Longrightarrow \qquad g(\mu_i) = \sum_{j=1} \beta_j x_{ij}.$$

  If $g(\mu) = \mu$, then we say that the link function g is the identity and we find the linear regression model shown in Chap. 9.

  If $g = Q$, the natural parameter then we say that g is the canonical link function because it transforms the mean of the random variable into the natural parameter of its distribution.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_10

10.1.1 Logistic Model for Binary Outcomes

Consider the case where the response variable is binary, i.e. $Y \sim Be(\pi)$. The Bernoulli distribution is part of the exponential family, in fact:

$$f_Y(y;\pi) = \pi^y(1-\pi)^{1-y} = (1-\pi)\exp\left\{y\log\left(\frac{\pi}{1-\pi}\right)\right\}\mathbb{I}_{(0,1)}(y);$$

where $\theta = \pi$, $a(\theta) = 1-\pi$, $b(y) = 1$, $Q(\theta) = \log\left(\frac{\pi}{1-\pi}\right) = logit(\pi)$. The logit is the canonical link function.

10.1.2 Models for Count Outcomes

To model count data, the Poisson distribution is generally used, $Y \sim Poisson(\mu)$. The Poisson distribution is part of the exponential family, in fact:

$$f_Y(y;\mu) = \frac{e^{-\mu}\mu^y}{y!} = \exp\{-\mu\}\frac{1}{y!}\exp\{y\log(\mu)\}\mathbb{I}_{\mathbb{N}}(y);$$

where $\theta = \mu$, $a(\theta) = e^{-\mu}$, $b(y) = 1/y!$, $Q(\theta) = \log(\mu)$.

In the case where a certain dispersion of the response variable is observed, **Y** can be modelled as a Negative Binomial:

$$\begin{aligned}
f_Y(y;k,\mu) &= \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)}\left(\frac{k}{\mu+k}\right)^k\left(1-\frac{k}{\mu+k}\right)^y;\\
\mathrm{e}[Y] &= \mu;\\
Var(Y) &= \mu + \frac{\mu^2}{k};\\
\frac{1}{k} \to 0 &\implies Var(Y) \to \mu \quad Y \xrightarrow{d} Poisson;
\end{aligned}$$

where $1/k$ is a dispersion parameter.

10.1.3 Other Link Functions

Other common link functions are:

- $\pi(\mathbf{x}) = F(\mathbf{x}) \implies F^{-1}(\pi(\mathbf{x})) = \boldsymbol{\beta}\mathbf{x}$; F generic distribution function.
- $\pi(\mathbf{x}) = \phi(\mathbf{x}) \implies \phi^{-1}(\pi(\mathbf{x})) = \boldsymbol{\beta}\mathbf{x}$; *probit* link function.

10.1.4 Interpretation of Parameters

The sign of β_j determines whether $\pi(\mathbf{x})$ increases or decreases as x increases. Let's focus on a single covariate x and define the odds ratio as:

$$\frac{\frac{\pi(x+1)}{1-\pi(x+1)}}{\frac{\pi(x)}{1-\pi(x)}} = \frac{\exp\{\beta_0 + \beta_1(x+1)\}}{\exp\{\beta_0 + \beta_1 x\}} = \exp\{\beta_1\}.$$

If $Y \sim Be(\pi)$, the logistic model is $logit(\pi) = \eta_i$, from which:

$$\pi(\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}\mathbf{x}\}}{1 + \exp\{\boldsymbol{\beta}\mathbf{x}\}}.$$

Typically, numerical methods are used to identify the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$. We will denote:

$$\hat{\mu}_i = g^{-1}(\hat{\eta}_i) = g^{-1}\left(\sum_{j=0}^{r} \hat{\beta}_j x_{ij}\right).$$

The quantity that is usually studied is the log-odds ratio, i.e. $\log\exp\{\beta_1\} = \beta_1$, which measures the relative risk increase (ratio between positive outcome and negative outcome) corresponding to a unit increase in the regressor.

10.1.5 Inference for Regression Parameters

Consider the following test, related to the parameter β_j:

$$H_0 : \beta_j = \beta_0 \qquad vs \qquad H_1 : \beta_j \neq \beta_0.$$

It can be shown that asymptotically:

$$Z = \frac{\hat{\beta}_j - \beta_0}{s.e.(\hat{\beta})} \sim N(0, 1);$$

Z is defined as the Wald statistic.

10.1.6 Model Selection

There are several methods to evaluate the optimal model. The two most well-known approaches in the literature are based respectively on deviance and on the AIC (Chapter 6 [1]).

Definition 10.1 Deviance Let $l(\hat{\boldsymbol{\mu}}; \mathbf{y})$ be the log-likelihood of the estimated model. Among all possible models, the maximum of the log-likelihood is reached at $l(\mathbf{y}; \mathbf{y})$, where we consider a parameter for each observation of the model. The model associated with $l(\mathbf{y}; \mathbf{y})$ is called the saturated model [1].

We define the deviance as:

$$-2\left[l(\hat{\boldsymbol{\mu}}; \mathbf{y}) - l(\mathbf{y}; \mathbf{y})\right].$$

The deviance is the statistic derived from the likelihood ratio to evaluate whether the model characterised by $l(\hat{\boldsymbol{\mu}}; \mathbf{y})$ is better than the saturated model. The deviance is asymptotically distributed as a $\chi^2_{(N-p)}$, where N is the sample size (which coincides with the number of parameters in the case of a saturated model) and p is the number of parameters of the model.

The deviance is used for model selection. In particular, it is possible to compare two models, characterised respectively by p_1 and p_2 parameters ($p_1 > p_2$), by performing a Chi-square test with $p_1 - p_2$ degrees of freedom.

A second approach to model selection involves evaluating the model with the lowest AIC, in line with linear regression models, Chap. 9.

10.1.7 Model Goodness

To evaluate the goodness of the model (Goodness Of Fit), a comparison is made between observed values (y_i) and values predicted ($\hat{y}_i$) by the model. To define the $\hat{y}_i$ we compare the estimated values $\hat{\pi}_i$ with a limit value π_0 (generally equal to 0.5).

Table 10.1 is defined as the misclassification table.

We define sensitivity (or sensibility) as:

$$\mathbb{P}\{\hat{Y} = 1 | Y = 1\} = \frac{a}{a+b}.$$

$$\mathbb{P}\{\hat{Y} = 0 | Y = 0\} = \frac{d}{c+d}.$$

Table 10.1 Misclassification table

	$\hat{y} = 1$	$\hat{y} = 0$
$y = 1$	a	b
$y = 0$	c	d

Sensitivity and sensibility are generally represented together in the ROC curve. To represent the ROC curve we report the sensibility on the y-axis, and 1− specificity on the x-axis. A good model is associated with a 'sharp' ROC curve, with high levels of sensitivity and specificity.

Online Supplementary Material

```
An online supplement to this chapter is available,
containing data, further insights and exercises.
```

10.1.8 Libraries

```
library( rms )
## Warning: package 'rms' was built under R version 3.5.2
## Loading required package: Hmisc
## Warning: package 'Hmisc' was built under R version 3.5.2
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 3.5.2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units
## Loading required package: SparseM
##
## Attaching package: 'SparseM'
## The following object is masked from 'package:base':
##
##     backsolve

library( ResourceSelection )
## Warning: package 'ResourceSelection' was built
## under R version 3.5.2
## ResourceSelection 0.3-4   2019-01-08
```

10.2 Exercises

Exercise 10.1 Consider the dataset related to a clinical study on patients suffering from coronary disorders (`CHDAGE_data.txt` in the online supplementary material). In particular, the aim of the study is to explain the presence or absence of significant coronary disorders based on the patients' age. The data refer to 100 patients. The variables in the database are:

- CHD binary variable: 1 if the coronary disorder is present, 0 if the disorder is absent.
- AGE continuous variable.

These data are taken from the site: http://www.umass.edu/statdata/statdata/
Answer the following questions:

(a) Graphically represent the dataset and comment on it.
(b) In order to have a better intuition of the relationship that binds CHD and AGE, transform the AGE variable into a categorical variable with 8 levels. The levels are: [20,29); [29,34); [34,39); [39,44); [44,49); [49,54); [54,59); [59,70]. Calculate the mean of CHD for each level and represent the 8 new pairs of values in the graph constructed in point a).
(c) Identify the most suitable model to describe the data and apply it. Also write the estimated model.
(d) Extract the `linear.predictors` and the `fitted.values` from the model. What is the relationship between these quantities?
(e) Represent the model used, using the graph produced at point (a).
(f) Give the definition of Odds Ratio in the simplest case of simple logistic regression with a binary dependent variable. Therefore, calculate the Odds Ratio corresponding to an age increase of 10 years.
(g) Calculate the 95% confidence interval for the Odds Ratio for a 10-year age increase.
(h) Calculate and represent the 95% confidence bands for each age value from 29 to 69.
(i) Evaluate the goodness of the model.

Exercise 10.2 In this exercise, we will analyse a clinical dataset related to the weight of newborns. The aim of the study is to identify the risk factors associated with giving birth to children weighing less than 2500 grams (low birth weight). The data refers to a sample of n = 189 women.

The variables in the database are described in the file `LOWBWTdata.txt` (see the online supplementary material):

- LOW: binary dependent variable (1 if the newborn weighs less than 2500 grams, 0 otherwise).
- AGE: mother's age in years.
- LWT: mother's weight in pounds before the start of pregnancy.
- FTV: number of medical visits during the last trimester of pregnancy.
- RACE discrete independent variable with 3 levels.

This dataset was investigated in [4].

Please answer the following questions:

(a) Graphically represent the data and comment on the graphs.
(b) Evaluate the most appropriate model to explain the binary variable `LOW`.
(c) Calculate the Odds Ratios related to the different levels of the variable `RACE` and comment on them.
(d) Evaluate the goodness of the chosen model at point b).
(e) Calculate the misclassification table and report the percentage of misclassified, using a threshold of 50%.
(f) Calculate the sensitivity and specificity of the model.
(g) Calculate the ROC curve to evaluate the GOF of the model.

Exercise 10.3 A group of financial engineers wants to investigate the factors that can influence the detection of bank fraud. In a preliminary analysis, the following variables are considered:

- `update_sito`: average annual time the site has been in maintenance.
- `media_mov_mens` average of monthly movements of the individual customer.
- `type_client`: type of client, 0 if standard client, 1 if silver client and 2 if gold client.

The event of interest is the fraud recorded by the individual customer (`fraud` equal to 1 if a fraud has been recorded in the last year and 0 otherwise). Therefore, answer the following questions after loading the file `fraud.txt` (see the online supplementary material).

(a) Graphically explore the relationship between `media_mov_mens` and `fraud`. Fit a suitable model to estimate the probability that a generic customer will suffer a fraud, using all available information. Comment on the fitted model.
(b) If deemed appropriate, propose a reduced model and/or with transformation. Compare the two models and justify the choice made.
(c) Explicitly write the chosen fitted model among the two proposed.
(d) Provide an interpretation of the odds ratio related to an increase in the average of monthly movements equal to 1.
(e) Compare the predictions that can be obtained through this model with the actual data (misclassification table, misclassification error, sensitivity, specificity).

Exercise 10.4 The dataset `TITANIC.txt` (see the online supplementary material) contains data related to the Titanic disaster, which sank on the night between 14 and 15 April 1912. For 1046 passengers, the following information is reported:

- Sex (`sex`, categorical variable with levels male and female).
- Age (`age`).
- Class (`pclass`, variable that takes the values 1,2,3) in which they were travelling.

The outcome of interest is whether the passengers survived or not the disaster, information reported within the binary variable `survived` (= 1 if the passenger survived, = 0 otherwise). A statistical investigation is to be carried out in order to assess how and to what extent the previously described covariates have influenced the survival probability of the Titanic passengers.

Perform the data analysis highlighting the following steps:

(a) Preliminary descriptive analysis: examine the contingency table of survival with respect to sex and comment on the result; also perform the boxplot of age with respect to survival and comment on the result.
(b) Fit a logistic regression model to explain the survival of passengers based on all available covariates and comment on the regression output: are the signs of the coefficients consistent with what was reasonably expected?
(c) Fit the previous logistic regression model without using the age regressor and compare the two models.
(d) Calculate the Odds Ratio of the survival probability of women compared to men.
(e) Calculate the survival probability (with its prediction interval) of a 76-year-old woman travelling in first, second and third class (specify in the command type = `response`).
(f) Calculate the misclassification table related to the model and the corresponding sensitivity.

10.3 Solutions

10.1

(a) Import the data.

```
chd = read.table( "CHDAGE_data.txt", head = TRUE )

str( chd )
## 'data.frame':    100 obs. of  3 variables:
##  $ ID : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ AGE: int  20 23 24 25 25 26 26 28 28 29 ...
##  $ CHD: int  0 0 0 0 1 0 0 0 0 0 ...

head( chd )
##   ID AGE CHD
## 1  1  20   0
## 2  2  23   0
## 3  3  24   0
## 4  4  25   0
## 5  5  25   1
## 6  6  26   0

attach( chd )
```

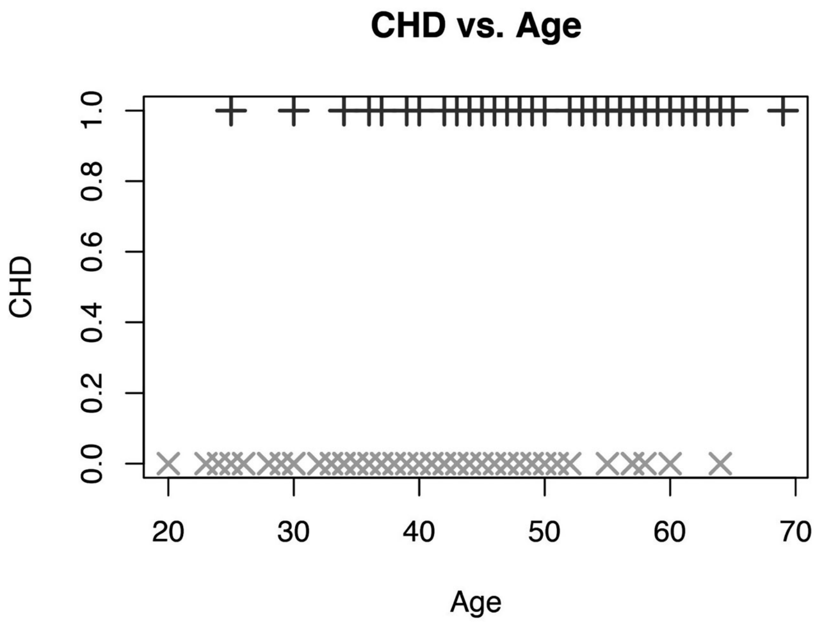

Fig. 10.1 Data visualisation. For each statistical unit, we represent the age on the x-axis, while on the y-axis the presence or absence of coronary disorders

Visualise the data in Fig. 10.1.

```
plot( AGE, CHD, pch = ifelse( CHD == 1, 3, 4 ),
      col = ifelse( CHD == 1, 'gray30', 'gray70' ),
      xlab = 'Age', ylab = 'CHD', main = 'CHD vs. Age',
      lwd = 2, cex = 1.5 )
```

From this graph, it can already be observed that as age increases, a higher number of patients suffering from coronary diseases seem to be recorded.

(a) Transform the `AGE` variable into a categorical variable with 8 levels. The levels are: [20,29); [29,34); [34,39); [39,44); [44,49); [49,54); [54,59); [59,70].

The choice of these classes is not random, but has been proposed based on the distribution of the `AGE` variable.

Insert in the x vector the limits of the age classes that you want to create (this step is arbitrary, and should be executed with good sense).

```
min( AGE )
## [1] 20
max( AGE )
## [1] 69

x  = c( 20, 29, 34, 39, 44, 49, 54, 59, 70 )
```

```
# Calculate the midpoints of the intervals we have created
mid = c( ( x [ 2:9 ] + x [ 1:8 ] )/2 )

# Divide the data into the classes we have created
GRAGE = cut( AGE, breaks = x, include.lowest = TRUE,
                       right = FALSE )
#GRAGE
```

We then calculate the average of coronary disorders with respect to each layer of the AGE variable and represent the obtained values in Fig. 10.2.

```
y = tapply( CHD, GRAGE, mean )
#y

plot( AGE, CHD, pch = ifelse( CHD == 1, 3, 4 ),
      col = ifelse( CHD == 1, 'gray30', 'gray70'),
      xlab = 'Age', ylab = 'CHD', main = 'CHD vs. Age',
      lwd = 2, cex = 1.5 )
points( mid, y, col = 1, pch = 16 )
```

Dividing patients into age classes and calculating the average of the dependent variable in each class, helps us to understand more clearly the nature of the relationship between AGE and CHD.

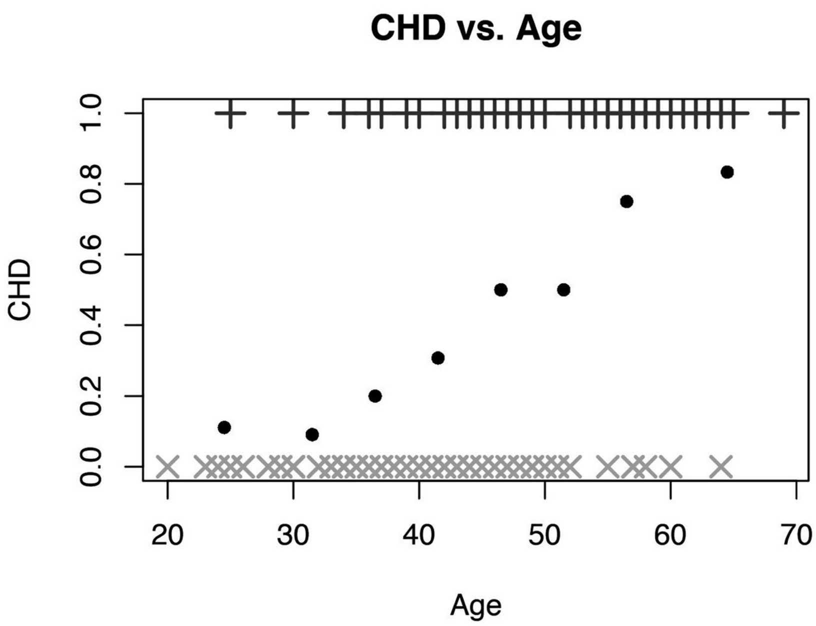

Fig. 10.2 Visualisation of the dataset with light and dark grey crosses. The black points represent the percentages of coronary disorders observed for each layer of the AGE variable

(b) We identify a model that adequately describes our data. The most suitable model is a generalised linear model with a `logit` link function.

```
help( glm )

mod = glm( CHD ~ AGE, family = binomial( link = logit ) )
summary( mod )
##
## Call:
## glm(formula = CHD ~ AGE, family = binomial(link = logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9718  -0.8456  -0.4576   0.8253   2.2859
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.30945    1.13365  -4.683 2.82e-06 ***
## AGE          0.11092    0.02406   4.610 4.02e-06 ***
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 136.66  on 99  degrees of freedom
## Residual deviance: 107.35  on 98  degrees of freedom
## AIC: 111.35
##
## Number of Fisher Scoring iterations: 4
```

The estimated model is therefore:

$$\text{logit}(\pi) = -5.30945 + 0.11092 \cdot \text{AGE};$$

where π is the probability that CHD equals 1. From the estimates obtained, we deduce that an increase in age leads to an increased risk of coronary disorders, as we had guessed graphically in the previous points.

(c) We investigate the `linear.predictors` and the `fitted.values`. First of all, the `linear.predictors` are the estimated values for the logit of the probability of having coronary disorders, $\text{logit}(\hat{\pi}_i)$. These values take values in $\mathbb{R}$.

```
mod$linear.predictors
```

The `fitted.values` are the estimated values for the probability of having coronary disorders, $\hat{\pi}_i$. These values take values in [0, 1].

```
mod$fitted.values
```

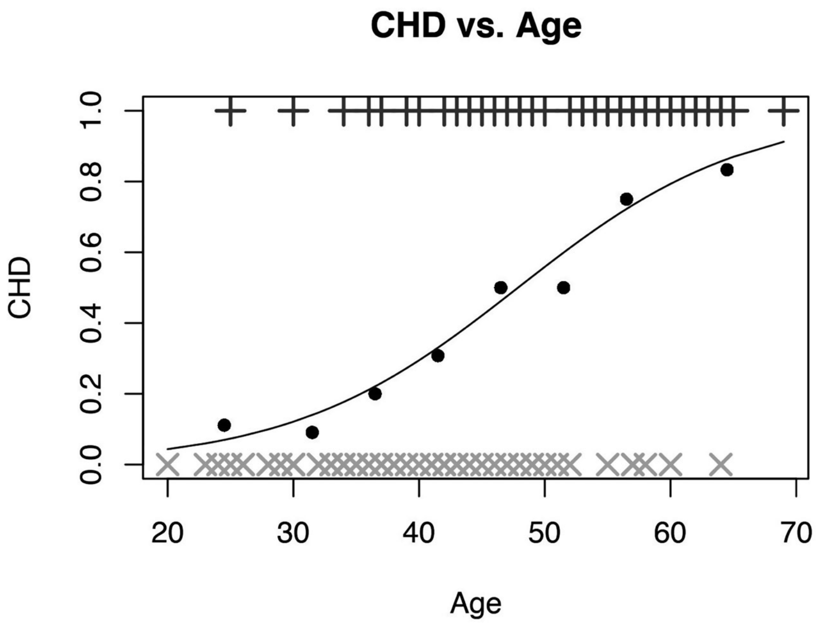

Fig. 10.3 Visualisation of the dataset (with light and dark grey crosses) and prediction obtained from the model (grey line). The black circles represent the percentages of CHD observed in relation to the different strata of AGE, calculated in point (b)

The two quantities are linked by the `logit` function.

(d) In Fig. 10.3 we represent the prediction of the model, starting from the graph proposed in point (a).

```
plot( AGE, CHD, pch = ifelse( CHD == 1, 3, 4 ),
      col = ifelse( CHD == 1, 'gray30', 'gray70'),
      xlab = 'Age', ylab = 'CHD', main = 'CHD vs. Age',
      lwd = 2, cex = 1.5 )
points( mid, y, col = 1, pch = 16 )
lines( AGE, mod$fitted, col = 'gray10' )
```

The estimated sigmoid is monotonically increasing, as we could guess from the estimate of β_{AGE}.

(e) One of the reasons why logistic regression technique is widely used, especially in the clinical field, is that the coefficients of the model have a natural interpretation in terms of *odds ratio* (hereafter *OR*).

Consider a dichotomous predictor x at levels 0 and 1. The odds that y = 1 among individuals with x = 0 is defined as:

$$\frac{\mathbb{P}(y=1|x=0)}{1-\mathbb{P}(y=1|x=0)}.$$

Similarly for subjects with x = 1, the odds that y = 1 is:

$$\frac{\mathbb{P}(y=1|x=1)}{1-\mathbb{P}(y=1|x=1)}.$$

The OR is defined as the ratio of the odds for x = 1 and x = 0.
Given that:

$$\mathbb{P}(y=1|x=1)=\frac{\exp(\beta_0+\beta_1\cdot x)}{1+\exp(\beta_0+\beta_1\cdot x)}$$

$$\mathbb{P}(y=1|x=0)=\frac{\exp(\beta_0)}{1+\exp(\beta_0)}$$

This implies:

$$\text{OR}=\exp(\beta_1)$$

Confidence intervals and generalisations to the case of variable x with more categories can be constructed immediately.

We therefore calculate the OR relative to AGE.

```
summary( mod )
##
## Call:
## glm(formula = CHD ~ AGE, family = binomial(link = logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9718  -0.8456  -0.4576   0.8253   2.2859
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.30945    1.13365  -4.683 2.82e-06 ***
## AGE          0.11092    0.02406   4.610 4.02e-06 ***
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 136.66  on 99  degrees of freedom
## Residual deviance: 107.35  on 98  degrees of freedom
```

```
## AIC: 111.35
##
## Number of Fisher Scoring iterations: 4
```

The coefficient of the AGE variable is 0.111. Therefore, the OR for a 10-year increase in age is:

```
exp( 10 * coef( mod ) [ 2 ] )
##      AGE
## 3.031967
```

for every 10-year increase in age, the risk of coronary disorder increases by about 3 times.

Observation The model assumes that the logit is linear in the age variable, i.e., the OR between people aged 20 versus 30 years is the same as between individuals aged 40 versus 50 years.

(f) We calculate a 95% confidence interval for the OR for a 10-year increase in age.

```
alpha = 0.05
qalpha =  qnorm( 1 - alpha/2 )
qalpha
## [1] 1.959964

IC.sup = exp( 10 * coef( mod ) [ 2 ]  + qalpha * 10 *
                summary( mod )$coefficients[ 2, 2 ] )
IC.inf = exp( 10 * coef( mod ) [ 2 ] - qalpha * 10 *
                summary( mod )$coefficients[ 2, 2 ] )
c( IC.inf, IC.sup )
##      AGE      AGE
## 1.892025 4.858721
```

(g) First, we set a grid of points from 29 to 69. Then, we calculate and represent in Fig. 10.4 the 95% confidence bands for each age value from 29 to 69.

```
# grid of x values at which to evaluate the regression
grid = ( 20:69 )

se = predict( mod, data.frame( AGE = grid ), se = TRUE )
# standard errors corresponding to the grid values

help( binomial )
gl = binomial( link = logit )  # link function used


plot( mid, y, col = 1, pch = 3, ylim = c( 0, 1 ),
      ylab = "Probability of CHD",
      xlab = "AGE", main = "IC for Logistic Regression" )
lines( grid, gl$linkinv( se$fit ) )
```

```
lines( grid, gl$linkinv( se$fit - qnorm( 1-0.025 ) * se$se ),
       col = 1, lty = 2 )
lines( grid, gl$linkinv( se$fit + qnorm( 1-0.025 ) * se$se ),
       col = 1, lty = 2 )
```

Observation The function `gl$linkinv` allows obtaining the value of probabilities from the link function (logit).

(h) In order to evaluate the goodness of the model, we calculate sensitivity and specificity.

```
threshold = 0.5
real.values = CHD
estimated.values = as.numeric(  mod$fitted.values > 0.5 )
tab = table( real.values, estimated.values )
tab
##              estimated.values
## real.values  0  1
##            0 45 12
##            1 14 29

sensitivity  = tab[ 2, 2 ] / ( tab [ 2, 1 ] + tab [ 2, 2 ] )
sensitivity
## [1] 0.6744186

specificity = tab [ 1, 1 ] /( tab [ 1, 2 ] + tab [ 1, 1 ] )
specificity
## [1] 0.7894737
```

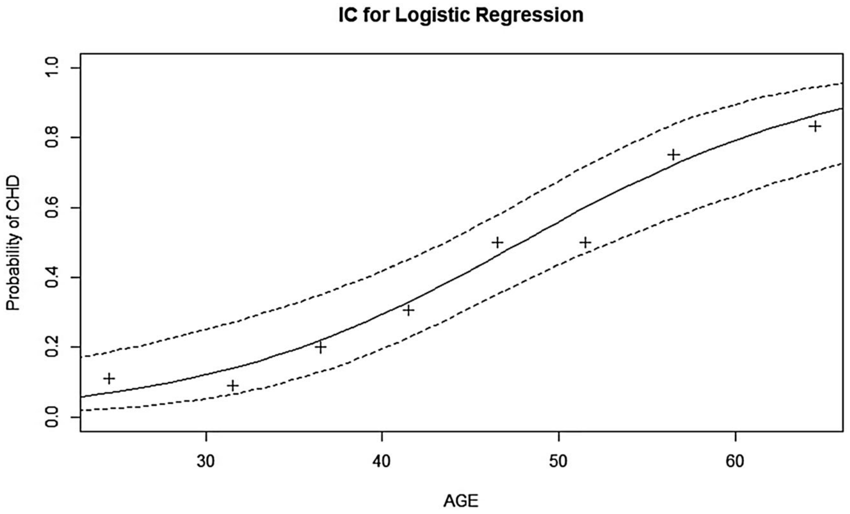

Fig. 10.4 Confidence intervals calculated for each new predicted point

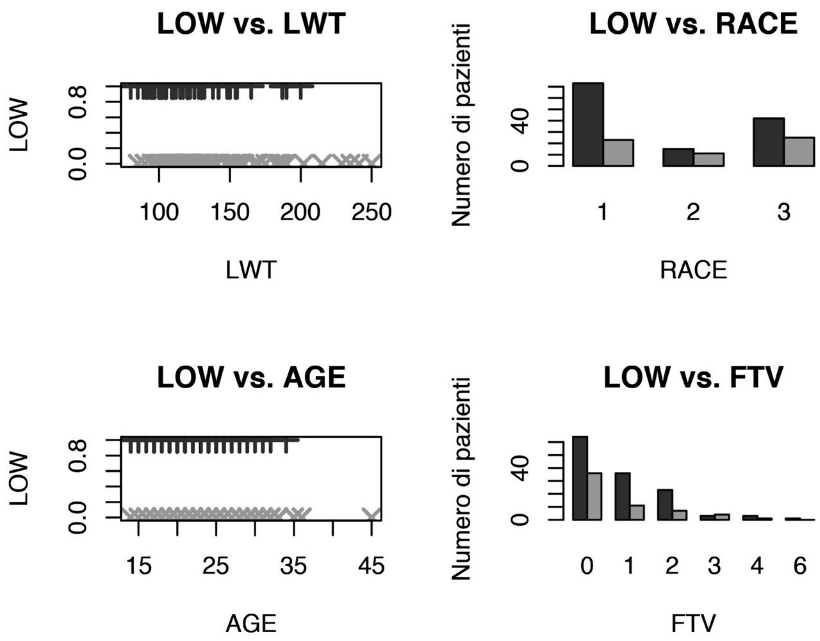

Fig. 10.5 Visualisation of the dataset

We conclude that it is a good model, given the high values of sensitivity and specificity.

10.2

(a) We import the data.

```
lw = read.table( "LOWBWTdata.txt", head = TRUE )
attach( lw )
## The following objects are masked from chd:
##
##     AGE, ID
```

We visualise the data in Fig. 10.5.

```
# treat the RACE variable as categorical
RACE  = factor( RACE )

par( mfrow = c( 2, 2 ) )
plot( LWT, LOW, pch = ifelse( LOW == 1, 3, 4 ),
      col = ifelse( LOW == 1, 'gray30', 'gray70' ),
      xlab = 'LWT', ylab = 'LOW', main = 'LOW vs. LWT',
      lwd = 2, cex = 1.5 )

counts_race <- table( LOW, RACE )
```

```
barplot( counts_race, col = c( 'gray30', 'gray70' ),
      xlab = 'RACE', ylab = 'Number of patients',
      main = 'LOW vs. RACE', beside = T)

plot( AGE, LOW, pch = ifelse( LOW == 1, 3, 4 ),
      col = ifelse( LOW == 1, 'gray30', 'gray70' ),
      xlab = 'AGE', ylab = 'LOW', main = 'LOW vs. AGE',
      lwd = 2, cex = 1.5 )

counts_FTV <- table( LOW, FTV )
barplot( counts_FTV, c( 'gray30', 'gray70' ),
      xlab = 'FTV', ylab = 'Number of patients',
      main = 'LOW vs. FTV', beside = T)
```

From the graphs, we assume that LWT could be significant, with a negative regression coefficient. AGE does not appear to be significant, nor does FTV. The variable RACE could be significant, as in the white race (RACE = 1) there is a strong presence of normal weight newborns, while in the other two categories there is a higher percentage of underweight newborns (LOW = 1, dark grey column).

(b) We set up a multiple logistic regression model to explain the variable LOW, including all available variables.

```
mod.low = glm( LOW ~ LWT + RACE + AGE + FTV,
               family = binomial( link = logit ) )
summary( mod.low )
##
## Call:
## glm(formula = LOW ~ LWT + RACE + AGE + FTV,
##        family = binomial(link = logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4163  -0.8931  -0.7113   1.2454   2.0755
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.295366   1.071443   1.209   0.2267
## LWT         -0.014245   0.006541  -2.178   0.0294 *
## RACE2        1.003898   0.497859   2.016   0.0438 *
## RACE3        0.433108   0.362240   1.196   0.2318
## AGE         -0.023823   0.033730  -0.706   0.4800
## FTV         -0.049308   0.167239  -0.295   0.7681
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.67  on 188  degrees of freedom
## Residual deviance: 222.57  on 183  degrees of freedom
## AIC: 234.57
##
## Number of Fisher Scoring iterations: 4
```

From this initial analysis, we conclude that the variables LWT and RACE are influential. In particular, a higher mother's weight is associated with a lower risk of underweight newborns and mothers of colour have a higher risk of having underweight children compared to white mothers.

If we stick to statistical significance alone, we conclude that it is possible to fit a reduced model, containing only the independent variable LWT. However, as in the case of multiple linear regression, the inclusion of a variable in the model can occur for different reasons. For example, in this case, the variable RACE is considered in the literature as important in predicting the effect in question, so it is included in the reduced model.

We evaluate a reduced model.

```
mod.low2 = glm( LOW ~ LWT + RACE,
                          family = binomial( link = logit ) )

summary( mod.low2 )
##
## Call:
## glm(formula = LOW ~ LWT + RACE,
## family = binomial(link = logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3491  -0.8919  -0.7196   1.2526   2.0993
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.805753   0.845167   0.953   0.3404
## LWT         -0.015223   0.006439  -2.364   0.0181 *
## RACE2        1.081066   0.488052   2.215   0.0268 *
## RACE3        0.480603   0.356674   1.347   0.1778
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.67  on 188  degrees of freedom
## Residual deviance: 223.26  on 185  degrees of freedom
## AIC: 231.26
##
## Number of Fisher Scoring iterations: 4
```

We note that AIC decreases and also RACE gains significance. We compare the two models using a Chi-square test.

```
anova( mod.low2, mod.low, test = "Chisq" )
## Analysis of Deviance Table
##
## Model 1: LOW ~ LWT + RACE
## Model 2: LOW ~ LWT + RACE + AGE + FTV
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       185     223.26
## 2       183     222.57  2  0.68618   0.7096
```

We conclude that we can consider the two tested models equally informative. Therefore, the best model is the simpler one, which contemplates LWT and RACE as variables.

(c) The RACE predictor is a 3-level discrete variable. In this case, level 1 (RACE = White) is assumed as the reference category.

```
model.matrix( mod.low2 ) [ 1:15, ]
##    (Intercept) LWT RACE2 RACE3
## 1            1 182     1     0
## 2            1 155     0     1
## 3            1 105     0     0
## 4            1 108     0     0
## 5            1 107     0     0
## 6            1 124     0     1
## 7            1 118     0     0
## 8            1 103     0     1
## 9            1 123     0     0
## 10           1 113     0     0
## 11           1  95     0     1
## 12           1 150     0     1
## 13           1  95     0     1
## 14           1 107     0     1
## 15           1 100     0     0

# OR 2 vs 1 ( Black vs White )
exp( coef( mod.low2 ) [ 3 ] )
##    RACE2
## 2.947821
```

Black women are a category with a risk of premature birth almost 3 times higher than white women.

```
# OR 3 vs 1 ( Other vs White )
exp( coef( mod.low2 ) [ 4 ] )
##   RACE3
## 1.61705
```

Women of other ethnicities are a category with a risk of premature birth about 1.5 times higher than white women.

(d) We perform tests to evaluate the GOF of the model.

```
mod.low2lrm = lrm( LOW ~ LWT + RACE, x = TRUE, y = TRUE )
residuals( mod.low2lrm, "gof" )
## Sum of squared errors     Expected value|H0        SD
##            38.2268160           38.2138614     0.1733477
##                     Z                    P
##             0.0747321            0.9404279

hoslem.test( mod.low2$y, fitted( mod.low2 ), g = 6 )
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  mod.low2$y, fitted(mod.low2)
## X-squared = 3.1072, df = 4, p-value = 0.5401
#g > 3
```

In this case too, we can conclude that the model provides a good fit of the data. For further details on the Hosmer-Lemeshow test, please refer to the online supplementary material.

(e) A frequently used way to present the results of a fit using logistic regression are classification tables. In these tables, the data are classified according to two keys:

- The value of the dichotomous dependent variable y.
- The value of a dichotomous variable y_{mod}, which is derived from the probability estimate obtained from the model. The values of this variable are obtained by comparing the value of the probability with a threshold (usual value 0.5).

We calculate y_{mod} (`predicted.values`).

```
threshold = 0.5

actual.values  = lw$LOW
predicted.values = as.numeric(mod.low2$fitted.values > threshold)
# 1 if > threshold, 0 if <= threshold
table( predicted.values )
```

We then compare the actual values with the predicted values, constructing a misclassification table.

```
tab = table( actual.values, predicted.values )

tab
##              predicted.values
## actual.values   0   1
##             0 124   6
##             1  53   6

# % of cases correctly classified:
```

```
round( sum( diag( tab ) ) / sum( tab ), 2 )
## [1] 0.69

# % of cases misclassified:
round( ( tab [ 1, 2 ] + tab [ 2, 1 ] ) / sum( tab ), 2 )
## [1] 0.31
```

31% of the data is misclassified.

(f) We calculate the sensitivity.

```
sensitivity = tab [ 2, 2 ] /( tab [ 2, 1 ] + tab [ 2, 2 ] )
sensitivity
## [1] 0.1016949
```

We calculate the specificity:

```
specificity = tab [ 1, 1 ] /( tab [ 1, 2 ] + tab [ 1, 1 ] )
specificity
## [1] 0.9538462
```

(g) We construct the ROC curve from the predicted values for the response from the `mod.low2` model of the `LOW` variable analysis.

```
fit2 = mod.low2$fitted


#sample mean of the survival probability in the sample

roc_threshold  = seq( 0, 1, length.out = 2e2 )
lens = length( roc_threshold )-1
roc_abscissa  = rep( NA, lens )
roc_ordinate = rep( NA, lens )

for ( k in 1 : lens )
{
  threshold = roc_threshold [ k ]
  classification = as.numeric( sapply( fit2,
                function( x ) ifelse( x < threshold, 0, 1 ) ) )

  #  CAUTION, I want the true on the rows
  # and the predicted on the columns
  # t.misc = table( lw$LOW, classification )

  roc_ordinate[ k ] = sum(
    classification[ which( lw$LOW == 1 ) ] == 1 )/
    length( which( lw$LOW == 1 ) )

  roc_abscissa[ k ] = sum(
    classification[ which( lw$LOW == 0 ) ] == 1 )/
```

```
    length( which( lw$LOW == 0 ) )

  #roc_ordinate[k]=t.misc[1, 1]/(t.misc [1, 1] + t.misc[1, 2])
  #
  #roc_abscissa[k]=t.misc[2, 1]/(t.misc [2, 1] + t.misc[2, 2])
}
```

We visualise the ROC curve in Fig. 10.6.

```
plot(roc_abscissa, roc_ordinate, type = "l",
      xlab = "1 - Specificity", ylab = "Sensitivity",
      main = "ROC Curve", lwd = 2, col = 'black',
     ylim = c( 0, 1 ), xlim = c( 0, 1 ) )
abline(h = c( 0, 1 ), v = c( 0, 1 ), lwd = 1, lty = 2,
          col = 'gray70')
abline(a = 0, b = 1, lty = 2, col = 'gray70' )

# we identify our levels of
# specificity and significance
abline( v = 1 - specificity,  h = sensitivity, lty = 3,
       col = 'gray30' )
points( 1 - specificity, sensitivity, pch = 4, lwd = 3,
            cex = 1.5, col = 'gray30')
```

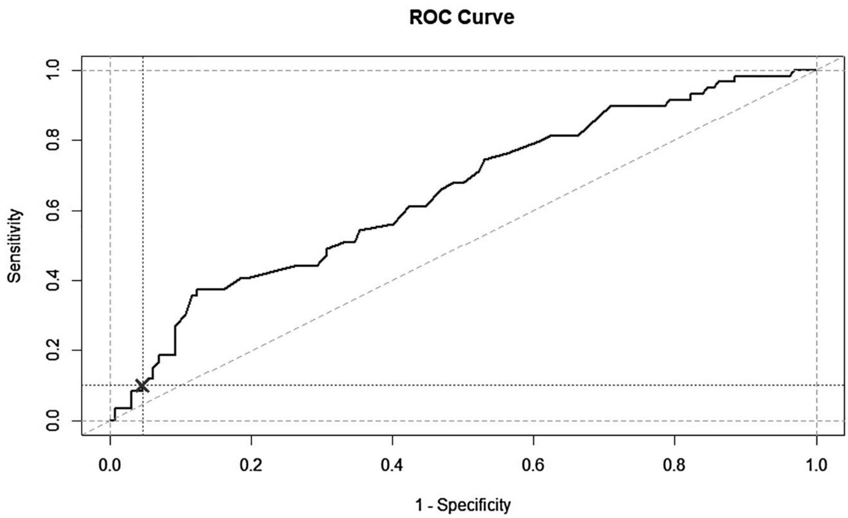

Fig. 10.6 Representation of the ROC curve, through a continuous black line. The light grey dashed lines delimit the domain and codomain of the Curve: $[0,1] \times [0,1]$. The dark grey cross and the dark grey dashed lines identify how the model under analysis is positioned within the curve

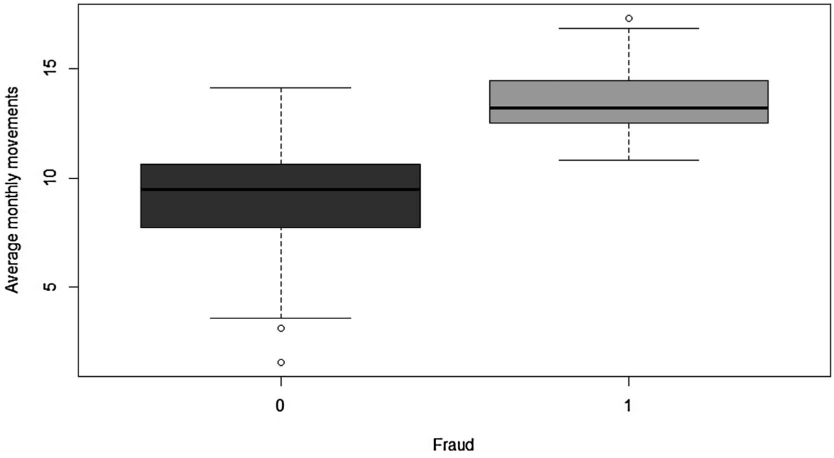

Fig. 10.7 Visualisation of the dataset through a boxplot. In dark grey are represented the movements that are the result of frauds, while in light grey those that are not the result of frauds

The ROC curve is not optimal, since it is quite flattened on the diagonal (the optimum is a curve that near zero has a positive and very high derivative).

10.3

(a) We graphically explore the relationship between `average_monthly_mov` and `fraud`.

```
data_fraud = read.table('fraud.txt', header = T)

boxplot( data_fraud$average_monthly_mov ~ data_fraud$fraud,
             col = c('gray30', 'gray70' ),
             ylab = 'Average monthly movements', xlab = 'Fraud')
```

In Fig. 10.7 there seems to be a relationship between the two variables. In particular, those who make more movements on average per month seem to have a higher risk of being a victim of fraud.

We fit a logistic regression model to explain the variable `fraud`, including all available variables.

```
mod_1 = glm( fraud ~ site_update + average_monthly_mov +
                       client_type, data = data_fraud,
                       family="binomial")

summary( mod_1 )
##
## Call:
## glm(formula = fraud ~ site_update + average_monthly_mov +
##           client_type,
##     family = "binomial", data = data_fraud)
```

```
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.93612  -0.07841  -0.00916   0.00009   1.99906
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.2459     5.3508   0.607 0.544104
## site_update      -0.3729     0.1026  -3.636 0.000277 ***
## average_monthly_mov   2.5748     0.6072   4.241 2.23e-05 ***
## client_type      -1.2787     0.8463  -1.511 0.130804
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 179.95  on 159  degrees of freedom
## Residual deviance:  43.45  on 156  degrees of freedom
## AIC: 51.45
##
## Number of Fisher Scoring iterations: 9
```

From the model, both `site_update` and `average_monthly_mov` seem to be significant (as we had guessed from the graph).

(b) We propose a reduced model that includes both `site_update` and `average_monthly_mov`.

```
mod_2 = glm( fraud ~ site_update + average_monthly_mov,
                      data = data_fraud, family="binomial")

summary( mod_2 )
##
## Call:
## glm(formula = fraud ~ site_update + average_monthly_mov,
##        family = "binomial",  data = data_fraud)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.79424  -0.09903  -0.01291   0.00013   2.02351
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.95243    5.28468   0.369 0.711791
## update_sito     -0.33912    0.09313  -3.641 0.000271 ***
## media_mov_mens   2.30649    0.51141   4.510 6.48e-06 ***
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 179.947  on 159  degrees of freedom
## Residual deviance:  45.933  on 157  degrees of freedom
```

```
## AIC: 51.933
##
## Number of Fisher Scoring iterations: 8
```

(c) Let's compare the two tested models, using a Chi-square test.

```
anova( mod_1, mod_2, test = "Chisq" )
## Analysis of Deviance Table
##
## Model 1: fraud ~ update_sito + media_mov_mens + type_client
## Model 2: fraud ~ update_sito + media_mov_mens
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       156     43.450
## 2       157     45.933 -1   -2.483   0.1151
```

From the test, there does not seem to be a significant difference between the two models, so we opt for the reduced model.

(d) Let's calculate the OR relative to an increase of one point on the average of monthly movements.

```
exp( 1*mod_2$coefficients[3] )
## media_mov_mens
##       10.03909
```

An increase of one point leads to a risk 10 times greater of suffering a fraud.

(e) Let's calculate the misclassification table, significance and specificity.

```
pred_val = ifelse( mod_2$fitted.values >= 0.5, 1, 0 )

tab = table( pred_val, data_fraud$fraud )
tab
##
## pred_val   0   1
##        0 115   6
##        1   5  34

sensitiv = tab [ 2, 2 ] /( tab [ 2, 1 ] + tab [ 2, 2 ] )
sensitiv
## [1] 0.8717949

specif = tab [ 1, 1 ] /( tab [ 1, 2 ] + tab [ 1, 1 ] )
specif
## [1] 0.9504132
```

Considering the low number of misclassified, and the high levels of specificity and significance, we can conclude that the reduced model fits well to the analysed data.

10.4

(a) Let's import the data.

```
data = read.table( 'TITANIC.txt', header = TRUE )


dim( data )
## [1] 1046    4

#str( data )

names( data )
## [1] "survived" "sex"      "age"      "pclass"
head( data )
##   survived    sex     age pclass
## 1        1 female 29.0000      1
## 2        1   male  0.9167      1
## 3        0 female  2.0000      1
## 4        0   male 30.0000      1
## 5        0 female 25.0000      1
## 6        1   male 48.0000      1
```

Let's set the survival variable as a factor.

```
data$survived = factor( data$survived )
#data$pclass = factor( data$pclass )
```

Let's calculate the contingency table of survival with respect to sex.

```
table( data$sex, data$survived )
##
##             0   1
##   female   96 292
##   male    523 135
```

From the contingency table we observe that, proportionally, more men died than women. There might be a correlation between these two variables.

Let's represent in Fig. 10.8 a boxplot to investigate the trend of survival with respect to age.

```
boxplot( data$age ~ data$survived, xlab = 'survivors',
         ylab = 'age', col = c('gray30', 'gray70' ) )
```

From the graph, there does not seem to be an effect of age on survival.

(b) Let's fit a logistic regression model to explain survival, including all the variables in the dataset.

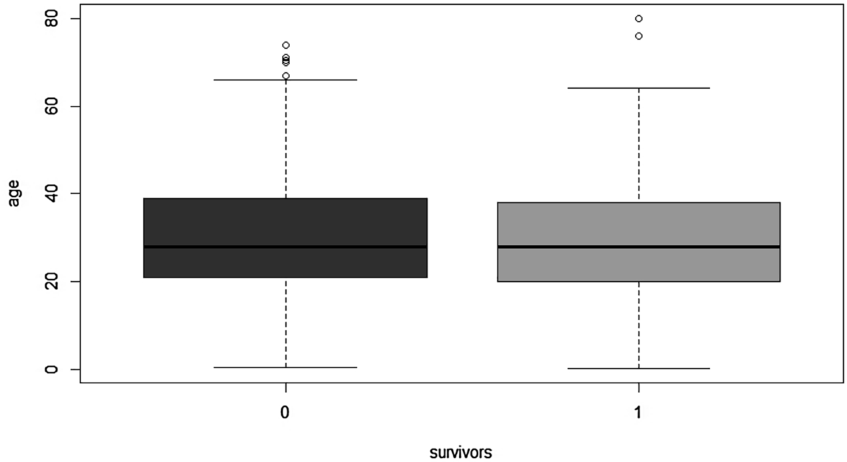

Fig. 10.8 Data visualisation through boxplot. In dark grey we represent the age of those who died, while in green we represent the age of those who survived

```
# glm model with all covariates
mod.glm = glm( survived ~ ., data = data,
               family = binomial( link = logit ) )
summary( mod.glm )
##
## Call:
## glm(formula = survived ~ ., family = binomial(link = logit),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6159  -0.7162  -0.4321   0.6572   2.4041
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.58927    0.40572  11.311  < 2e-16 ***
## sexmale     -2.49738    0.16612 -15.034  < 2e-16 ***
## age         -0.03388    0.00628  -5.395 6.84e-08 ***
## pclass      -1.13324    0.11173 -10.143  < 2e-16 ***
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1414.62  on 1045  degrees of freedom
## Residual deviance:  983.02  on 1042  degrees of freedom
## AIC: 991.02
##
## Number of Fisher Scoring iterations: 4
```

All variables appear to be significant, moreover being male, older and being in second, third class decreases the risk of survival.

(c) We fit a logistic regression model excluding the variable `age`.

```
mod.glm.red = update( mod.glm, . ~ . - age )
summary( mod.glm.red )
##
## Call:
## glm(formula = survived ~ sex + pclass,
##        family = binomial(link = logit), data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1248  -0.7134  -0.4816   0.6976   2.1033
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.00428    0.25591  11.740   <2e-16 ***
## sexmale     -2.52785    0.16326 -15.484   <2e-16 ***
## pclass      -0.85747    0.09511  -9.016   <2e-16 ***
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1414.6  on 1045  degrees of freedom
## Residual deviance: 1013.8  on 1043  degrees of freedom
## AIC: 1019.8
##
## Number of Fisher Scoring iterations: 4
```

As expected, all variables are significant.

```
anova( mod.glm, mod.glm.red, test = "Chisq" )
## Analysis of Deviance Table
##
## Model 1: survived ~ sex + age + pclass
## Model 2: survived ~ sex + pclass
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      1042     983.02
## 2      1043    1013.85 -1  -30.822 2.829e-08 ***
## ---
## Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1

1 - pchisq( 1013.85 - 983.02, 1 )
## [1] 2.816499e-08
```

Given the p-value of the test, we can conclude that the full model is more informative than the reduced model.

(d) We calculate the OR of the survival probability of men compared to women.

```
exp( -mod.glm$coefficients[ 2 ] )
##  sexmale
## 12.15057
```

Women have a survival probability 12 times higher than that of men.

(e) We calculate the survival probability of a 76-year-old woman who travelled in first, second and third class.

```
mod_pred.conf1 = predict( mod.glm,
                          data.frame(age = 76, sex = 'female',
                                     pclass = 1 ),
                          type = 'response', se.fit = T )
mod_pred.conf2 = predict( mod.glm,
                          data.frame(age = 76, sex = 'female',
                                     pclass = 2 ),
                          type = 'response', se.fit = T )
mod_pred.conf3 = predict( mod.glm,
                          data.frame(age = 76, sex = 'female',
                                     pclass = 3 ),
                          type = 'response', se.fit = T )
mod_pred.conf1$fit
##         1
## 0.7069811
mod_pred.conf2$fit
##         1
## 0.4372139
mod_pred.conf3$fit
##         1
## 0.2000918
```

As expected, the lower the class, the lower the survival probability.

(f) We calculate the misclassification table and calculate the sensitivity.

```
threshold = 0.5
actual.values = data$survived
estimated.values = as.numeric(  mod.glm$fitted.values > 0.5 )
tab = table( actual.values, estimated.values )
tab
##              estimated.values
## actual.values   0   1
##             0 523  96
##             1 126 301

# Sensitivity = True Positive Rate,
# e.g. empirical probability of classifying a
# positive as such
tab[ 2, 2 ] / sum( tab[ 2, ] )
## [1] 0.704918
```

Chapter 11
ANOVA: Analysis of Variance

11.1 Theory Recap

The analysis of variance, also known by the acronym ANOVA (ANalysis Of VAriance), is a statistical technique that aims to compare the means of a random phenomenon among different groups of statistical units. This analysis is approached through the decomposition of variance.

11.1.1 ANOVA

Consider a random variable Y_{ij} related to the statistical unit $i \in \{1, \ldots, n_j\}$ belonging to the group $j \in \{1, \ldots, g\}$. Suppose that Y_{ij} can be modelled in the following way:

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}, \qquad i = 1, .., n_j \quad j = 1, ..., g; \tag{11.1}$$

where μ is the overall mean, while τ_j represents the average deviation from μ in group g. Furthermore, it is assumed:

- Normality: $\varepsilon_{ij} \sim N(0, \sigma_j^2)$.
- Homoscedasticity: $\sigma_j^2 = \sigma^2 \quad \forall j$.
- Independence: $\varepsilon_{ij} \perp\!\!\!\perp \varepsilon_{i'j'} \quad \forall i \neq i', j \neq j'$.

The model described in Eq. (11.1) is a one-way ANOVA, as we are considering a single factor. If we were considering two factors, we would have:

$$Y_{ijk} = \mu + \tau_j + \gamma_k + \alpha_{jk} + \varepsilon_{ijk}, \qquad i = 1, .., n_{jk} \quad k = 1, .., K \quad j = 1, ..., J; \tag{11.2}$$

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_11

where one factor consists of K levels and the second factor is at J levels. In this case, we would talk about two-way ANOVA. For simplicity, we consider only the one-way ANOVA model in the theory recap.

Theorem 11.1 (Decomposition of Variance) *Denote y_{ij} the realisations of the random variable Y_{ij}, $i \in \{1, .., n_j\}$ and $j \in \{1, .., g\}$, where the total sample size is $N = \sum_{j=1}^{g} n_j$. It can be shown that:*

$$\sum_{j=1}^{g}\sum_{i=1}^{n_j}(y_{ij} - \overline{y})^2 = \sum_{j=1}^{g} n_j \cdot (\overline{y}_{j.} - \overline{y})^2 + \sum_{j=1}^{g}\sum_{i=1}^{n_j}(y_{ij} - \overline{y}_{j.})^2; \tag{11.3}$$

where:

$$\overline{y}_{j.} = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij};$$

$$\overline{y} = \frac{\sum_{j=1}^{g}\sum_{i=1}^{n_j} y_{ij}}{\sum_{j=1}^{g} n_j}.$$

Equation (11.3) can be succinctly rewritten as:

$$SS_{TOT} = SS_B + SS_W;$$

where SS_{TOT} represents the total variance, SS_B represents the variance between different groups (between groups) and SS_W represents the variance within groups (within groups).

Given the assumptions of the model, it can be shown that:

$$\frac{1}{\sigma^2}\sum_{j=1}^{g}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{j.})^2 \sim \chi^2_{N-g}.$$

$$\frac{1}{\sigma^2}\sum_{j=1}^{g} n_j \cdot (\overline{Y}_{j.} - \overline{Y})^2 \sim \chi^2_{g-1}.$$

$$\frac{1}{\sigma^2}\sum_{j=1}^{g}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y})^2 \sim \chi^2_{N-1}.$$

Table 11.1 ANOVA table

Variance	d.f.	Sum of squares	Mean	F statistic
Between	g-1	$SS_B = \sum_{j=1}^{g} n_j \cdot (\overline{y}_{j.} - \overline{y})^2$	$MS_B = \frac{SS_B}{(g-1)}$	MS_B/MS_W
Within	N-g	$SS_W = \sum_{j=1}^{g} \sum_{i=1}^{n_j} (y_{ij} - \overline{y}_{.j})^2$	$MS_W = \frac{SS_W}{(N-g)}$	
Total	N-1	$SS_{TOT} = \sum_{j=1}^{g} \sum_{i=1}^{n_j} (y_{ij} - \overline{y})^2$	$MS_{TOT} = \frac{SS_{TOT}}{(N-1)}$	

When we are interested in performing an ANOVA, we want to carry out the following hypothesis test:

$$\mathrm{H}_0 : \tau_1 = \tau_2 = .. = \tau_g \quad \text{vs} \quad \mathrm{H}_1 : \exists i \quad \text{t.c.} \quad \tau_i \neq \tau_j \quad j \in \{1, .., g\} \setminus i.$$

Under H_0 the test statistic MS_B/MS_W is distributed as a Fisher with parameters $g - 1$ and $N - g$ (see Table 11.1). A low p-value leads us to reject H_0 and therefore conclude that not all groups have the same mean.

Online Supplementary Material

```
An online supplement to this chapter is available,
containing data, further insights and exercises.
```

11.1.2 Libraries

```
library( MASS )
library( car ) #for LEVENE TEST
## Loading required package: carData
library( faraway )
##
## Attaching package: 'faraway'
## The following objects are masked from 'package:car':
##
##     logit, vif
library( Matrix )
library( RColorBrewer ) #for color palette
library( ggplot2 )
## Warning: package 'ggplot2' was built under R version 3.5.2
```

11.2 Exercises

Exercise 11.1 Write the ANOVA model, specifying the assumptions. Can the one-way ANOVA model be seen as a linear regression model? If so, specify the design matrix and the analogy between ANOVA tests and linear regression tests.

Exercise 11.2 Record the height for N individuals, coming from three different regions: a maritime region (10 individuals), a mountainous region (15 individuals) and a hilly region (12 individuals). Describe the hypothesis test that must be performed to test whether the average height is different based on the region of origin and write the necessary `R` commands to solve the problem.

Exercise 11.3 (Visualisation of Variance Decomposition) In order to visualise the theorem on variance decomposition, carry out the following points:

(a) Generate a dataset composed of three variables: 'Measures' (the quantity of interest), 'Group' (the group variable) and the 'ids' variable (which is a count of the rows). Choose $N = 200$ statistical units, divided into 4 balanced groups ('A', 'B', 'C', 'D'). Generate the dataset taking into account the assumptions of the ANOVA model.
(b) Visualise, using the 'ggplot2' library, the theorem on variance decomposition.

Exercise 11.4 (One-Way ANOVA) Analyse the `chickwts` data, available in `R`, related to the weight of chickens subjected to different diets and determine if there is a difference between the average weights of the chickens among the different groups. The `chickwts` dataset is composed of two variables:

- `weight`: response variable (Y_{ij}), weight of chicken i, under diet j.
- `feed`: categorical variable at g levels indicating the type of diet.

Import the `chickwts` data. We want to investigate whether the weight of the chickens is influenced by the type of diet.

Exercise 11.5 (Identifiability of the ANOVA Model) Provide the definition of model identifiability. Discuss two formulations of the test on the difference of means in the case where there is a single categorical variable at 7 levels, each with a count of {3,2,3,2,3,2,3}. Specify the design matrix of each model.

Exercise 11.6 (One-Way ANOVA) Consider the `coagulation` dataset [2] , in the `faraway` library:

- `coag`: coagulation time (continuous, positive variable).
- `diet`: type of diet followed (categorical variable at 4 levels).

The dataset is 24×2. Prove that the diet impacts the average coagulation time.

Exercise 11.7 (Two-Ways ANOVA) Consider the `rats` dataset in the `faraway` package. Investigate the effect of the type of poison and the type of treatment administered to 48 rats. The object of evaluation is the survival time (in tens of hours) of the 48 rats. The dataset contains the following variables:

- `time` (survival time): continuous.
- `poison` (poison): categorical at 3 levels (I, II, III).
- `treat` (treatment): categorical at 4 levels (A, B, C, D).

11.3 Solutions

11.1 The answer is affirmative: ANOVA models can be seen as linear regression models. For simplicity, let's consider a one-way ANOVA model.

On one hand, the ANOVA model can be expressed as follows:

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad i \in \{1, .., n_j\} \quad j \in \{1, .., g\}. \tag{11.4}$$

This model is analogous to:

$$Y_i = \mu + \mu_1 X_{i1} + \mu_2 X_{i2} + .. + \mu_g X_{ig} + \varepsilon_i, \qquad i \in \left\{1, .., N = \sum_{j=1}^{g} n_j\right\}; \tag{11.5}$$

a linear regression model with $g + 1$ parameters, where the covariates X_{il} are worth 1 if the statistical unit i is associated with the l-th level of the group and 0 otherwise.

In both models, the assumptions of homoscedasticity and normality must hold (among all elements, therefore also within the groups).

The model in Eq. (11.5) is not the only acceptable one, indeed in this case it can be easily proven that the design matrix is not invertible. For a more in-depth reflection on the subject, refer to Exercise 11.5.

11.2 To answer this question, an ANOVA test must be performed. In the dataset, we have a total of $N = 37$ individuals, a factor at three levels, $g = 3$. The groups are unbalanced, as they do not have the same number.

We can start with a graphical exploration of the data using the `boxplot` command and juxtaposing the boxplots relative to the different groups. If the boxplots are at different heights, we expect an effect of the factor on the response variable. If the boxplots are asymmetric, the assumption of normality might be violated. If the boxplots are of very different sizes, the assumption of homoscedasticity might be violated.

We verify that the assumption of normality is satisfied using the `shapiro.test()` command, which performs a Shapiro test. The test must be repeated on each group separately, or it can be used simultaneously using the `tapply` command. If the p-values are high, I accept H_0, i.e., I accept the assumption of normality for each group.

I verify the assumption of homoscedasticity among the groups, through the Bartlett or Levene test. If the p-value is high, I accept H_0, i.e., I accept the assumption of homogeneity of variance. For further study on these tests, refer to the online material.

If the model assumptions are respected, we proceed with our analysis with the `anova`, `aov` or `lm` command. In all three cases, we look at the p-value related to the Fisher test. If the p-value is low, I reject H_0, i.e., there is a difference in the mean of the variable of interest due to the factor (the region in this case).

11.3

(a) We generate the dataset as required by the text. Since the groups are balanced, we have 50 observations per group. Also, since the assumptions of the ANOVA model are valid, we must generate the data from a normal distribution with the same variance for each group. The mean is at the reader's discretion.

```
N = 50
set.seed(1000)
group_1 = rnorm( N, 10, 3 )
group_2 = rnorm( N, 5, 3 )
group_3 = rnorm( N, 20, 3 )
group_4 = rnorm( N, 50, 3 )

groups = rep( c('A','B','C','D'), each = N )
data_oneway_aov = cbind(
  c( group_1, group_2, group_3, group_4 ),
  groups,
  1:(N*4) )

data_oneway_aov = as.data.frame( data_oneway_aov )
colnames( data_oneway_aov ) = c('Measures', 'Group', 'ids')
```

Before proceeding, let's check that `Measures` and `ids` are variables of type `numeric`, while `Group` should be a variable of type `factor`. If not, we convert the type of these variables. This step is important for later visualisation.

```
is( data_oneway_aov$Measures )[1]
## [1] "factor"
is( data_oneway_aov$Group )[1]
## [1] "factor"

cast = data_oneway_aov$Measures
data_oneway_aov$Measures = as.numeric(levels(cast))[cast]

is( data_oneway_aov$Measures )[1]
## [1] "numeric"

cast = data_oneway_aov$ids
data_oneway_aov$ids = as.numeric(levels(cast))[cast]

is( data_oneway_aov$ids )[1]
## [1] "numeric"
```

Let's make a first graph to verify that we have correctly generated the data.

```
ggplot(data_oneway_aov, aes( x = Group, y = Measures,
                             fill = Group ) ) +
  geom_boxplot() + scale_fill_grey() +
  scale_colour_grey() + theme_bw() +
  geom_segment( aes( x = 0.5, y = 10, xend = 1.5, yend = 10,
                     colour = 'Real mean'), linetype = 2 ) +
  geom_segment( aes( x = 1.5, y = 5,  xend = 2.5, yend = 5,
                     colour = 'Real mean'), linetype = 2 ) +
  geom_segment( aes( x = 2.5, y = 20, xend = 3.5, yend = 20,
                     colour = 'Real mean'), linetype = 2 ) +
  geom_segment( aes( x = 3.5, y = 50, xend = 4.5, yend = 50,
                     colour = 'Real mean'), linetype = 2 )
```

Observing Fig. 11.1, we conclude that the dataset is correctly generated, in fact we see that the means used to generate the data (dashed lines), are very close to the medians and the boxplots are symmetric (as is correct for the Normal distribution).

(a) We then proceed to the visualisation of the theorem on the decomposition of variance.

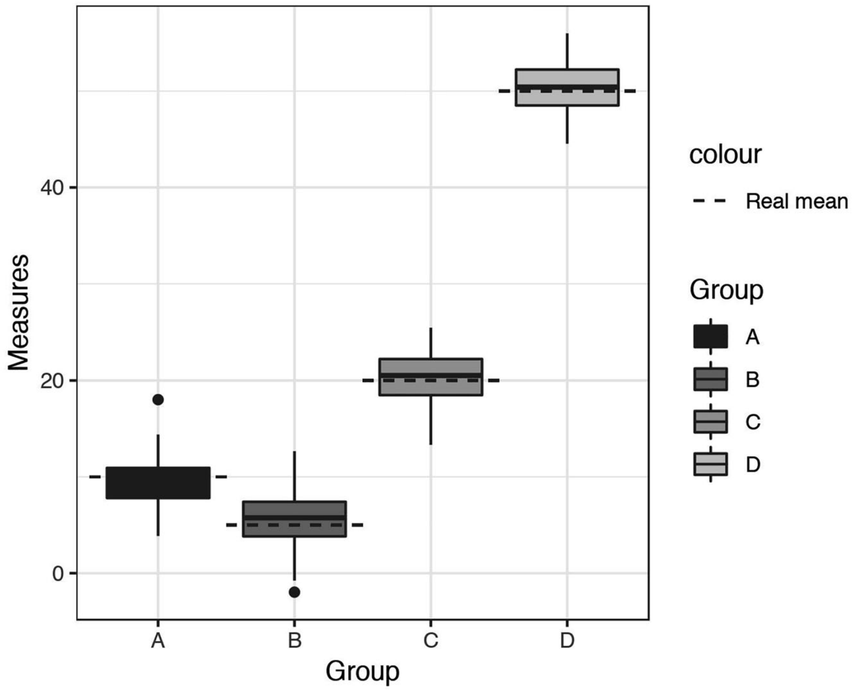

Fig. 11.1 Boxplot of the quantity of interest 'Measures', recorded in the 4 groups 'A', 'B', 'C' and 'D'

We start by calculating the quantities of interest. In particular, we create a dataset that contains the previously created dataset, the means of `Measures` for each group and the global mean of `Measures`.

```
mean_per_group = tapply( data_oneway_aov$Measures,
                         data_oneway_aov$Group, mean )
mean_tot = mean( data_oneway_aov$Measures )

data_lines = cbind( data_oneway_aov,
                              rep( mean_per_group, each = N ),
                    rep( mean_tot, 4*N) )
data_lines = as.data.frame( data_lines )
colnames( data_lines ) = c( 'Measures', 'Group', 'ids',
                            'mean_per_group', 'mean_tot' )

head( data_lines )
##    Measures Group ids mean_per_group mean_tot
## 1  8.662665     A   1        9.52411 21.42521
## 2  6.382430     A   2        9.52411 21.42521
## 3 10.123379     A   3        9.52411 21.42521
## 4 11.918165     A   4        9.52411 21.42521
## 5  7.640337     A   5        9.52411 21.42521
## 6  8.843532     A   6        9.52411 21.42521
```

We draw our first quantity of interest, the components of SS_W, in Fig. 11.2:

$$Y_{ij} - \overline{Y}_{j\cdot} \quad i \in \{1, .., n_j\} \quad j \in \{1, .., G\}.$$

```
ggplot( data_oneway_aov, aes( ids, Measures ) ) +
  geom_point( aes(color= Group)) + scale_fill_grey() +
  scale_colour_grey() + theme_bw() +
  geom_segment( x = 1, y = mean_per_group[1], xend = 50,
                yend = mean_per_group[1], colour = 1 ) +
  geom_segment( x = 51, y = mean_per_group[2], xend = 100,
                yend = mean_per_group[2], colour = 1 ) +
  geom_segment( x = 101, y = mean_per_group[3], xend = 150,
                yend = mean_per_group[3], colour = 1 ) +
  geom_segment( x = 151, y = mean_per_group[4], xend = 200,
                yend = mean_per_group[4], colour = 1 ) +
  geom_segment( data = data_lines,
                aes( x = ids, y = mean_per_group,
                        xend = ids, yend = Measures),
                colour = "gray" ) +
  theme( axis.title.x=element_blank(),
         axis.text.x=element_blank(),
         axis.ticks.x=element_blank() )
```

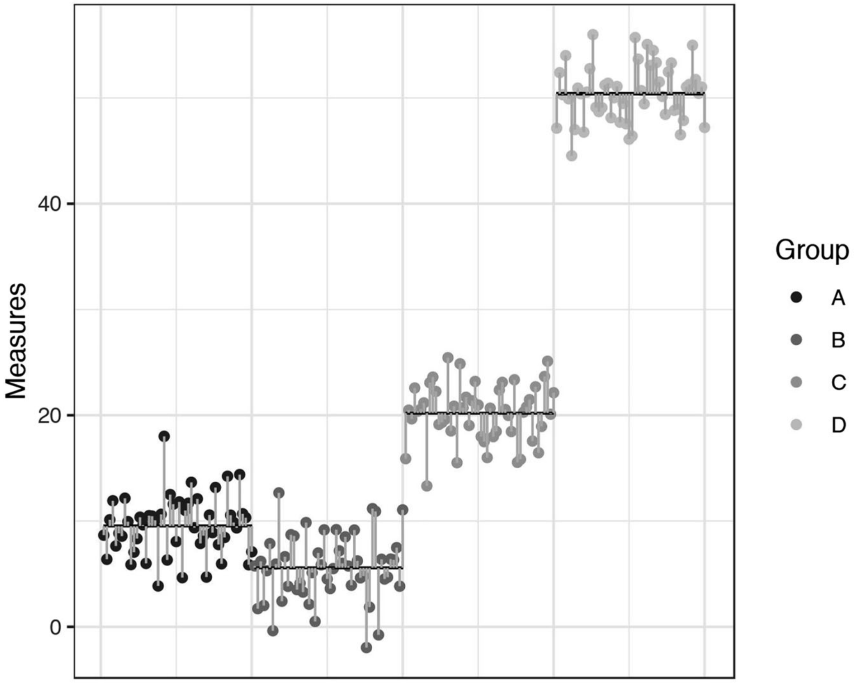

Fig. 11.2 Representation of the different components of SS_W, i.e., the distance of each point from the mean of the group to which it belongs

We draw our second quantity of interest, the components of SS_B, in Fig. 11.3:

$$\overline{Y}_{j\cdot} - \overline{Y} \quad j \in \{1, .., G\}.$$

```
ggplot( data_oneway_aov, aes( ids, Measures ) ) +
 geom_point( aes(color= Group)) + scale_fill_grey() +
 scale_colour_grey() + theme_bw() +
 geom_segment( x = 1, y = mean_per_group[1], xend = 50,
              yend = mean_per_group[1], colour = 1 ) +
 geom_segment( x = 51, y = mean_per_group[2], xend = 100,
              yend = mean_per_group[2], colour = 1 ) +
 geom_segment( x = 101, y = mean_per_group[3], xend = 150,
              yend = mean_per_group[3], colour = 1 ) +
 geom_segment( x = 151, y = mean_per_group[4], xend = 200,
              yend = mean_per_group[4], colour = 1 ) +
 geom_segment( x = 1, y = mean_tot, xend = 200,
              yend = mean_tot, colour = 1 ) +
 geom_segment( x = 25, y = mean_per_group[1], xend = 25,
              yend = mean_tot, colour = "gray" ) +
 geom_segment( x = 75, y = mean_per_group[2], xend = 75,
              yend = mean_tot, colour = "gray" ) +
```

```
geom_segment( x = 125, y = mean_per_group[3], xend = 125,
              yend = mean_tot, colour = "gray" ) +
geom_segment( x = 175, y = mean_per_group[4], xend = 175,
              yend = mean_tot, colour = "gray" ) +
theme( axis.title.x=element_blank(),
       axis.text.x=element_blank(),
       axis.ticks.x=element_blank() )
```

We draw our third quantity of interest, the components of SS_{TOT}, in Fig. 11.4:

$$Y_{ij} - \overline{Y} \quad i \in \{1, .., n_j\} \quad j \in \{1, .., G\}.$$

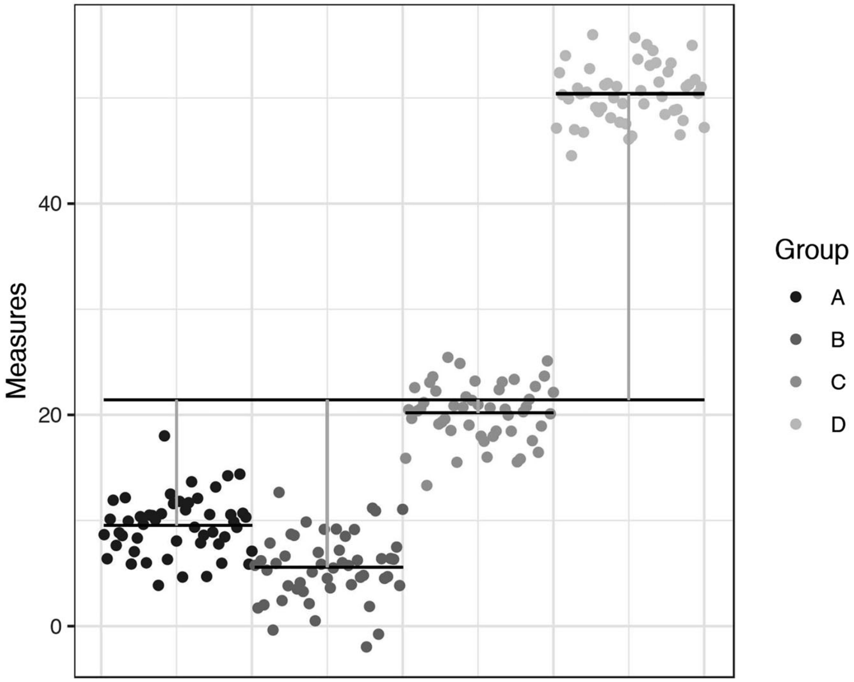

Fig. 11.3 Representation of the different components of SS_B, i.e., the distance of the mean of each group from the global mean

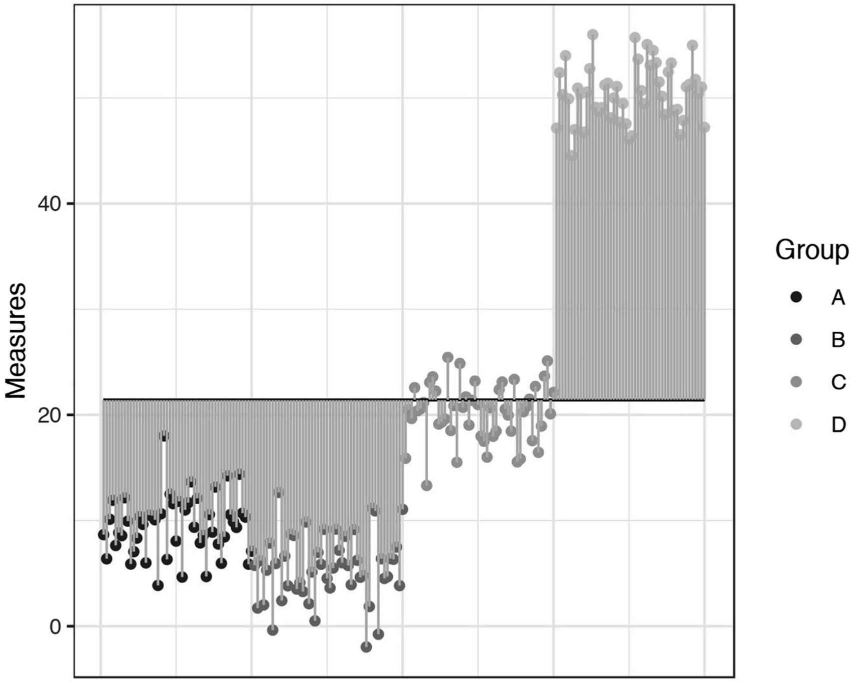

Fig. 11.4 Representation of the different components of SS_{TOT}, that is, the distance of each point from the global mean

```
ggplot( data_oneway_aov, aes( ids, Measures ) ) +
  geom_point( aes(color= Group)) + scale_fill_grey() +
  scale_colour_grey() + theme_bw() +
  geom_segment( x = 1, y = mean_tot, xend = 200,
                  yend = mean_tot,
                colour = 1 ) +
  geom_segment( data = data_lines,
                 aes( x = ids, y = mean_tot,
                        xend = ids, yend = Measures),
                colour = "gray" ) +
  theme( axis.title.x=element_blank(),
         axis.text.x=element_blank(),
         axis.ticks.x=element_blank() )
```

11.4 To solve this exercise, we need to proceed as follows:

(a) Import the dataset.
(b) Visualise the dataset.
(c) Set up the model.

(d) Verify the model's assumptions.
(e) Check the difference between the means (through testing).

(a) *Import the dataset.*

```
head( chickwts )
##   weight      feed
## 1    179 horsebean
## 2    160 horsebean
## 3    136 horsebean
## 4    227 horsebean
## 5    217 horsebean
## 6    168 horsebean
tail( chickwts )
##    weight   feed
## 66    352 casein
## 67    359 casein
## 68    216 casein
## 69    222 casein
## 70    283 casein
## 71    332 casein

attach( chickwts )
```

The dataset consists of N = 71 observations, divided into 6 groups (g = 6).

```
tapply( chickwts$weight, chickwts$feed, length )
##casein horsebean   linseed  meatmeal   soybean sunflower
##    12        10        12        11        14        12
```

The groups appear to be quite balanced.

(b) *Visualise the dataset.*
We visualise the data through boxplots, so as to get an intuition about the presence of any differences in the response between chickens that follow different diets.

```
summary( chickwts )
##      weight             feed
##  Min.   :108.0   casein   :12
##  1st Qu.:204.5   horsebean:10
##  Median :258.0   linseed  :12
##  Mean   :261.3   meatmeal :11
##  3rd Qu.:323.5   soybean  :14
##  Max.   :423.0   sunflower:12

boxplot( weight ~ feed, xlab = 'feed', ylab = 'weight',
         main = 'Chicken weight according to feed',
         col = gray.colors(6) )
abline( h = mean( weight ) )
```

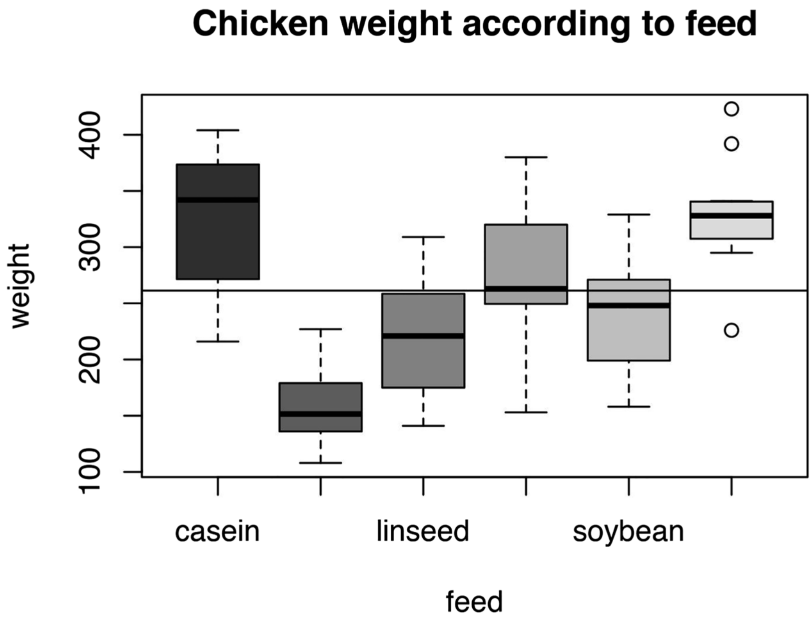

Fig. 11.5 Representation of the weight of chickens according to different diets

From the comparison of the boxplots in Fig. 11.5, it seems that there is some effect, the means appear different depending on the diet followed.

(c) *Set up the model.*
We want to investigate the following one-way ANOVA model:

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij};$$

in which $i \in \{1, .., n_j\}$ is the index of the statistical unit within the group j, while $j \in \{1, .., g\}$ is the group index.
We are interested in performing the following test:

$$H_0 : \tau_i = \tau_j \quad \forall i, j \in \{1, .., 6\} \qquad vs \qquad H_1 : \exists\,(i, j)\,|\tau_i \neq \tau_j.$$

Paraphrasing, H_0 assumes that all chickens belong to a single population, while H_1 assumes that the chickens belong to 2, 3, 4, 5 or 6 populations with different means. In Fig. 11.6 we have graphically represented what we would record if H_0 were true (the means in the different groups would coincide) in the left panel, and what we actually record in our dataset in the right panel.

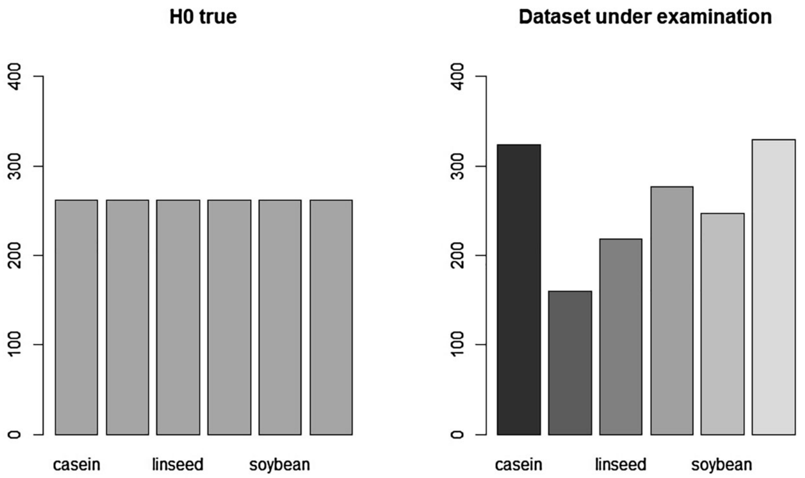

Fig. 11.6 Representation of what we would observe if H_0 were true in the left panel. Representation of the average weight of chickens in different diets in the right panel

```
par( mfrow = c( 1, 2 ) )

barplot( rep( mean( weight ), 6 ),
         names.arg = levels( feed ),
         ylim = c( 0, max( weight ) ), main = "H0 true",
         col = 'grey' )

barplot( tapply( weight, feed, mean ),
         names.arg = levels( feed ),
         ylim = c( 0, max( weight ) ),
         main = "Dataset under examination",
         col = gray.colors(6) )
```

(d) *Verify the model's assumptions.*
We verify that the ANOVA assumptions are met:

- *Normality* within the group (through Shapiro test).
- *Homoscedasticity* between the groups (through Bartlett or Levene test).

```
n  = length( feed )
ng  = table( feed )
treat  = levels( feed )
g  = length( treat )

# Normality of data in groups
Ps = c( shapiro.test( weight [ feed == treat [ 1 ] ] )$p,
        shapiro.test( weight [ feed == treat [ 2 ] ] )$p,
        shapiro.test( weight [ feed == treat [ 3 ] ] )$p,
```

```
        shapiro.test( weight [ feed == treat [ 4 ] ] )$p,
        shapiro.test( weight [ feed == treat [ 5 ] ] )$p,
        shapiro.test( weight [ feed == treat [ 6 ] ] )$p )
#Ps


# In a more compact and elegant way:
Ps = tapply( weight, feed,
                   function( x ) ( shapiro.test( x )$p ) )
Ps
##    casein horsebean   linseed  meatmeal
## 0.2591841 0.5264499 0.9034734 0.9611795
##   soybean sunflower
## 0.5063768 0.3602904
```

The p-value vector of the Shapiro test (Ps) are all high, so I accept the hypothesis of normality in all groups.
Let's verify the hypothesis of homoscedasticity.

```
Var = c( var( weight [ feed == treat [ 1 ] ] ),
         var( weight [ feed == treat [ 2 ] ] ),
         var( weight [ feed == treat [ 3 ] ] ),
         var( weight [ feed == treat [ 4 ] ] ),
         var( weight [ feed == treat [ 5 ] ] ),
         var( weight [ feed == treat [ 6 ] ] ) )
#Var


# In a more compact and elegant way:
Var = tapply( weight, feed, var )
#Var

# Uniformity test of variances
bartlett.test( weight, feed )
##
##  Bartlett test of homogeneity of variances
##
## data:  weight and feed
## Bartlett's K-squared = 3.2597, df = 5, p-value = 0.66

# Alternative: Levene-Test
leveneTest( weight, feed )
## Levene's Test for Homogeneity of Variance
## (center = median)
##       Df F value Pr(>F)
## group  5  0.7493 0.5896
##       65
```

The tests agree, we conclude that the hypothesis of variance homogeneity is respected.

(e) *Verify difference between means (through test).*

Now that we have verified that the hypotheses are satisfied we can proceed with a one-way ANOVA.

$$F_0 = \frac{SS_{TREAT}/r}{SS_{RES}/(n-p)} \sim F(r, n-p);$$

in which:

$$SS_{TREAT} = \sum_{j=1}^{g} \tau_j^2 \cdot n_j.$$

In the one-way ANOVA the number of regressors $r = g - 1$.
In R we can perform an ANOVA test in three ways:
1. Performing a manual F test.

```
Media  = mean( weight )
Mediag  = tapply( weight, feed, mean )

SStot  = var( weight ) * ( n-1 )
SStreat = sum( ng * ( Mediag-Media )^2 )
SSres  = SStot - SStreat

alpha = 0.05
Fstatistic = ( SStreat / ( g-1 ) ) / ( SSres / ( n-g ) )

# "small" values do not lead us to reject
cfr.fisher = qf( 1-alpha, g-1, n-g )
Fstatistic > cfr.fisher
## [1] TRUE
Fstatistic
## [1] 15.3648
cfr.fisher
## [1] 2.356028

P = 1-pf( Fstatistic, g-1, n-g )
P
## [1] 5.93642e-10
```

Observing our F statistic (Fstatistic), we notice that we are well beyond the 5% threshold. So I have strong evidence to reject the null hypothesis (confirmed by the p-value equal to $5.94e^{-10}$.
2. Running the aov command.

```
help( aov )

fit = aov( weight ~ feed )
# or anova( mod )
summary( fit )
#              Df Sum Sq Mean Sq F value   Pr(>F)
```

```
# feed         5 231129   46226   15.37 5.94e-10 ***
# Residuals   65 195556    3009
# ---
#Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
```

The aov command shows the decomposition of variance and the outcome of the ANOVA test. In this case $SS_B = 231129$ and $SS_W = 195556$. The p-value of the test is $5.94e^{-10}$, so we reject the null hypothesis.
3. Running the lm command.

```
mod = lm( weight ~ feed )
summary( mod )
##
## Call:
## lm(formula = weight ~ feed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -123.909  -34.413    1.571   38.170  103.091
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     323.583     15.834  20.436  < 2e-16 ***
## feedhorsebean -163.383     23.485  -6.957 2.07e-09 ***
## feedlinseed   -104.833     22.393  -4.682 1.49e-05 ***
## feedmeatmeal   -46.674     22.896  -2.039 0.045567 *
## feedsoybean    -77.155     21.578  -3.576 0.000665 ***
## feedsunflower    5.333     22.393   0.238 0.812495
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 54.85 on 65 degrees of freedom
## Multiple R-squared:  0.5417, Adjusted R-squared:  0.5064
## F-statistic: 15.36 on 5 and 65 DF,  p-value: 5.936e-10
```

Through the lm command, we model our response variable using a linear model. The test that interests us is the one related to the global significance of the model reported in the last line of the summary (see Chap. 9).
Through all three proposed approaches, we reject H_0 and conclude that there is a difference between the means of the different groups.

11.5 We are in the case of one-way ANOVA. This model can be represented as:

$$Y_{ij} = \tau + \mu_j + \varepsilon_{ij}, \qquad i \in \{1, .., n_j\} \quad j \in \{1, .., g\}.$$

Alternatively, we can consider a linear regression model with a categorical variable X (*dummy variable*) at g levels.

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Non-invertible Design Matrix

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_g X_{ig} + \varepsilon_{ij}.$$

Observation

$$\beta_0 = \mu\beta_j = \mu_j \quad \forall j = 1, .., g.$$

In this model, the design matrix X has dimension $N \times (g+1)$. Each row of X, $\mathbf{x_i}$ is a binary vector of length $g+1$, in which one appears at the first element (corresponding to the intercept) and in the $j+1$-th element, where j represents the group of membership of the element i.

Considering 7 groups with sizes {3,2,3,2,3,2,3} respectively, the design matrix X described above is:

$$\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

This design matrix (X.full in the code) is singular, i.e. not invertible (to invert it manually we must resort to the Moore-Penrose pseudoinverse).

In other words, the model described in Eq. (11.5) is not identifiable.

Invertible Design Matrix

Alternatively, we can consider the following model:

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{g-1} X_{ig-1} + \varepsilon_{ij}.$$

In this model, the design matrix X has dimension $N \times g$, the elements of which are $\{-1, 0, 1\}$. The first column, as in the previous case, is made up of all 1s (elements related to the intercept). While, the rows related to the statistical units of the first $g - 1$ groups are composed of all zeros, except the first element and the j-th element, where j represents the group of membership. Finally, the rows related to the statistical units belonging to the group g are made up of all -1, except the first element which is 1. This design matrix is also called a contrast matrix.

This design matrix X in our case becomes:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

We immediately notice that this matrix is similar to the previous one but it is non-singular and therefore invertible.

Observation If we execute the command `tapply( feed, feed, length )`, R calculates the group sizes and reorders them in alphabetical order by group name. If we want to use the group sizes in the order in which they appear in the dataset (`feed`), we must do:

```
n
## [1] 71

group_names = unique( as.character( feed ) )
ng = tapply( feed, feed, length )[ group_names ]
```

Non-invertible Design Matrix in R

We construct the matrix `X.full`, that is a design matrix where we consider all the groups (dimension = $N \times (g + 1)$). In particular, we create $g + 1$ columns and assemble them using the `cbind` command.

```
# group 1 (in the order of the data in ( weight,feed )
x1.full = c( rep( 1, ng [ 1 ] ),
            rep( 0, n - ng [ 1 ] ) )

# group 2 (in the order of the data in ( weight,feed )
x2.full = c( rep( 0, ng [ 1 ] ),
            rep( 1, ng [ 2 ] ),
            rep( 0, n - ng [ 1 ] - ng [ 2 ] ) )

# group 3 (in the order of the data in ( weight,feed )
x3.full = c( rep( 0, ng [ 1 ] + ng [ 2 ] ),
            rep( 1, ng [ 3 ] ),
            rep( 0, n - ng [ 1 ] - ng [ 2 ] - ng [ 3 ] ) )

# group 4 (in the order of the data in ( weight,feed )
x4.full = c( rep( 0, n - ng [ 6 ] - ng [ 5 ] - ng [ 4 ] ),
            rep( 1, ng [ 4 ] ),
            rep( 0, ng [ 5 ]  + ng [ 6 ] ) )

# group 5 (in the order of the data in ( weight,feed )
x5.full = c( rep( 0, n - ng [ 6 ] - ng [ 5 ] ),
            rep( 1, ng [ 5 ] ),
            rep( 0, ng [ 6 ] ) )

# group 6 (in the order of the data in ( weight,feed )
x6.full = c( rep( 0, n - ng [ 6 ] ),
            rep( 1, ng [ 6 ] ) )

X.full = cbind( rep( 1, n ),
                   x1.full,
                   x2.full,
                   x3.full,
                   x4.full,
                   x5.full,
                   x6.full )
```

To prove that `X.full` does not have full rank, we observe that one column is a linear combination of other columns of the matrix.

```
stopifnot( length(
  which( X.full[ , 1 ] - rowSums( X.full[ , - 1 ] ) == 0 ) )
  == dim( X.full )[1]  )
```

We now estimate the $\hat{\boldsymbol{\beta}}$. Remember that X is singular, so H will be calculated as follows:

$$H = X \cdot (X^T \cdot X)^{\dagger} \cdot X^T$$

where $(X^T \cdot X)^{\dagger}$ indicates the Moore-Penrose pseudo-inverse. And the $\hat{\boldsymbol{\beta}}$ will be calculated as:

$$\hat{\boldsymbol{\beta}} = (X^T \cdot X)^{\dagger} \cdot X^T \cdot \mathbf{y}$$

```
# H.full = X.full%*%solve(t(X.full)%*%X.full)%*%t(X.full)
# R gives an error, because it's singular!

H.full = X.full%*%ginv(t(X.full)%*%X.full)%*%t(X.full)

y  = weight

betas.full = as.numeric(
  ginv(t(X.full)%*%X.full)%*%t(X.full) %*% y)
```

The mean in the j-th group is:

$$\mathbb{E}[Y_j] = \mu + \tau_j = \beta_0 + \beta_j, \qquad j = \{1, .., g\}.$$

```
means_by_group = betas.full[ 1 ] +
  betas.full[ 2:length( betas.full ) ]
names( means_by_group ) = group_names

means_by_group
## horsebean   linseed   soybean sunflower  meatmeal    casein
##  160.2000  218.7500  246.4286  328.9167  276.9091  323.5833

tapply( weight, feed, mean )[ unique( as.character( feed ) ) ]
## horsebean   linseed   soybean sunflower  meatmeal    casein
##  160.2000  218.7500  246.4286  328.9167  276.9091  323.5833
```

The global mean is:

$$\mu = \sum_{j=1}^{g} \frac{n_j \cdot \mu_j}{N}.$$

```
global_mean = ng %*% means_by_group / n
global_mean
##          [,1]
## [1,] 261.3099

mean( weight )
## [1] 261.3099
```

Invertible Design Matrix in R

```
x1.red = c( rep( 1, ng [ 1 ] ),
            rep( 0, n - ng [ 1 ] - ng [ 6 ] ),
            rep( -1, ng [ 6 ] ) )
stopifnot( sum( x1.red - ( x1.full - x6.full ) ) == 0 )

x2.red = c( rep( 0, ng [ 1 ] ),
            rep( 1, ng [ 2 ] ),
            rep( 0,
             n - ng [ 1 ] - ng [ 2 ] - ng [ 6 ] ),
            rep( -1, ng [ 6 ] ) )
stopifnot( sum( x2.red - ( x2.full - x6.full ) ) == 0 )

x3.red = c( rep( 0, ng [ 1 ] + ng [ 2 ] ),
            rep( 1, ng [ 3 ] ),
\begin{Verbatim}[frame = single,fontsize=\small\ttfamily]
rep( 0,
     n - ng [ 1 ] - ng [ 2 ] - ng [ 3 ] - ng [ 6 ] ),
    rep( -1, ng [ 6 ] ) )
stopifnot( sum( x3.red - ( x3.full - x6.full ) ) == 0 )

x4.red = c( rep( 0, n - ng [ 6 ] - ng [ 5 ] - ng [ 4 ] ),
            rep( 1, ng [ 4 ] ),
            rep( 0, ng [ 5 ] ),
            rep( -1, ng [ 6 ] ) )
stopifnot( sum( x4.red - ( x4.full - x6.full ) ) == 0 )

x5.red = c( rep( 0, n - ng [ 6 ] - ng [ 5 ] ),
            rep( 1, ng [ 5 ] ),
            rep( -1, ng [ 6 ] ) )
stopifnot( sum( x5.red - ( x5.full - x6.full ) ) == 0 )


X.red  = cbind( rep( 1, n ),
                x1.red,
                x2.red,
                x3.red,
```

```
                 x4.red,
                 x5.red )
```

We now estimate the $\hat{\boldsymbol{\beta}}$.

```
H.red  = X.red %*% solve( t( X.red ) %*% X.red ) %*% t(X.red)

betas.red = as.numeric(
  solve( t( X.red ) %*% X.red ) %*% t( X.red ) %*% y )
```

The mean in the j-th group is obtained as follows:

$$\mu_j = \beta_0 + \beta_i \qquad i = 1, \ldots, g-1.$$

$$\mu_g = \beta_0 - (\beta_1 + ... + \beta_{g-1}).$$

```
means_by_group = betas.red[ 1 ] + betas.red[ -1 ]
means_by_group = c( means_by_group, betas.red[ 1 ] -
                      sum( betas.red[ -1 ] ) )

names( means_by_group ) = group_names
means_by_group
## horsebean   linseed   soybean sunflower  meatmeal    casein
##  160.2000  218.7500  246.4286  328.9167  276.9091  323.5833
tapply( weight, feed, mean )[ group_names ]
## horsebean   linseed   soybean sunflower  meatmeal    casein
##  160.2000  218.7500  246.4286  328.9167  276.9091  323.5833
```

Therefore, with both the singular and non-singular design matrix, we arrive at the same result.

Observation What does R do automatically?

To answer this question, let's revisit the previous exercise on the `chickwts` data and extract, using the `model.matrix` command, the design matrix of the tested models.

```
mod_aov = aov( weight ~ feed )
X_aov = model.matrix( mod_aov )
```

We see that the design matrix created by the ANOVA is of dimensions $N \times g$. We note that the variable (level) `casein` is missing and is used as a baseline. The considered regression model then becomes:

$$Y_{ij} = \beta_0 + \beta_2 X_{i2} + \cdots + \beta_g X_{ig} + \varepsilon_{ij}.$$

```
mod_lm = lm( weight ~ feed )
X_lm = model.matrix( mod_lm )
```

The same applies in the case of a linear model.

Observation The *first group* (in the alphanumeric order of the levels of the stratification variable, `feed`, and NOT in the order of appearance of the data) is suppressed and taken as a reference (baseline).

We now calculate the $\hat{\boldsymbol{\beta}}$.

```
betas.lm = coefficients( mod_lm )
```

The mean in the j-th group is obtained as follows:

$$\mu_{baseline} = \beta_0.$$

$$\mu_j = \beta_0 + \beta_j, \qquad j \neq baseline.$$

```
means_by_group = c( betas.lm[ 1 ],
                             betas.lm[ 1 ] + betas.lm[ -1 ] )
names( means_by_group ) = levels( feed )

means_by_group
##    casein horsebean   linseed  meatmeal   soybean sunflower
##  323.5833  160.2000  218.7500  276.9091  246.4286  328.9167
tapply( weight, feed, mean )
##    casein horsebean   linseed  meatmeal   soybean sunflower
##  323.5833  160.2000  218.7500  276.9091  246.4286  328.9167
```

11.6 We import the dataset.

```
data( coagulation )

dim( coagulation )
## [1] 24  2
names( coagulation )
## [1] "coag" "diet"
head( coagulation )
##   coag diet
## 1   62    A
## 2   60    A
## 3   63    A
## 4   59    A
## 5   63    B
## 6   67    B
```

To answer the exercise question, we would like to set up a one-way ANOVA. Before carrying out the analyses, we check the assumptions of the model:

1. Normality;
2. Homoscedasticity.

Elements that violate the assumptions, are:

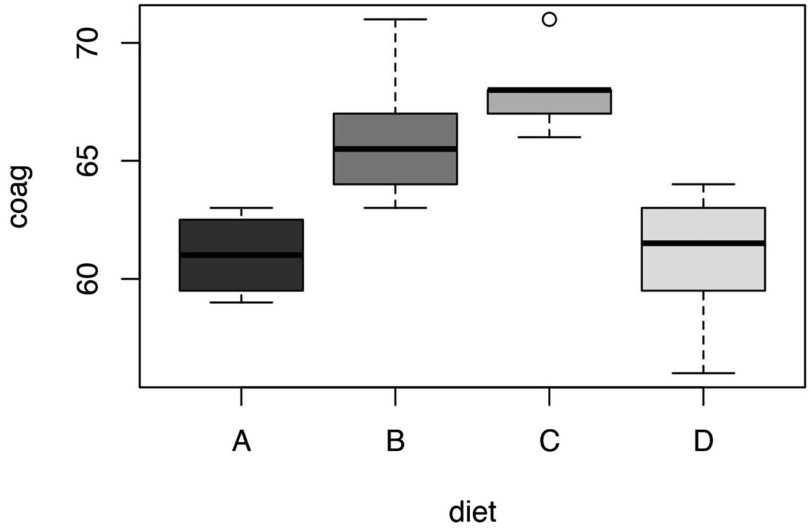

Fig. 11.7 Boxplot of coagulation time depending on the type of diet

1. Skewness (see asymmetric group-specific boxplots).
2. Heteroscedasticity (see different sizes of group-specific boxplots).

Observation Graphical analyses have a purely exploratory purpose, especially when dealing with small-sized datasets (like `coagulation`). Indeed, even in the case of homogeneous variances among the groups, we can expect variability between the groups. We draw the group-specific boxplots in Fig. 11.7.

```
plot( coag ~ diet, data = coagulation, col = grey.colors(4) )
```

It seems that the assumptions are respected. We find skewness only in group C, where however only 4 observations are recorded, one of which is significantly distant from the others.

```
table( coagulation$diet )
##
## A B C D
## 4 6 6 8

coagulation$coag[ coagulation$diet == 'C' ]
## [1] 68 66 71 67 68 68
unique( coagulation$coag[ coagulation$diet == 'C' ] )
## [1] 68 66 71 67
```

The observations of coagulation in group C are all very close.
We now analyse the ANOVA model.

```
mod = lm( coag ~ diet, coagulation )
summary( mod )
##
## Call:
## lm(formula = coag ~ diet, data = coagulation)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -5.00  -1.25   0.00   1.25   5.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.100e+01  1.183e+00  51.554  < 2e-16 ***
## dietB       5.000e+00  1.528e+00   3.273 0.003803 **
## dietC       7.000e+00  1.528e+00   4.583 0.000181 ***
## dietD       2.991e-15  1.449e+00   0.000 1.000000
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 2.366 on 20 degrees of freedom
## Multiple R-squared:  0.6706, Adjusted R-squared:  0.6212
## F-statistic: 13.57 on 3 and 20 DF,  p-value: 4.658e-05
```

We see the design matrix used by R.

```
dim( model.matrix( mod ) )
## [1] 24  4
model.matrix( mod )[1:5, ]  #n x g
##   (Intercept) dietB dietC dietD
## 1           1     0     0     0
## 2           1     0     0     0
## 3           1     0     0     0
## 4           1     0     0     0
## 5           1     1     0     0
```

Group 'A' is taken as the reference (or baseline) group (first according to alphanumeric order). The effects must be interpreted as differences compared to the baseline group. We can read the model output in the following way:

- Group A: mean = 61.
- Group B: mean = 61 + 5.
- Group C: mean = 61 + 7.
- Group D: mean = 61 + 0.

Thanks to the F statistic of the analysed model, we can conclude that there is an effect of the diet on coagulation.

We now try to fit the same model, removing the intercept and see how the following quantities change: design matrix, estimates of β and p-value of the F test.

```
mod_i = lm( coag ~ diet - 1, coagulation )

summary( mod_i )
##
## Call:
## lm(formula = coag ~ diet - 1, data = coagulation)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -5.00  -1.25   0.00   1.25   5.00
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## dietA  61.0000     1.1832   51.55   <2e-16 ***
## dietB  66.0000     0.9661   68.32   <2e-16 ***
## dietC  68.0000     0.9661   70.39   <2e-16 ***
## dietD  61.0000     0.8367   72.91   <2e-16 ***
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 2.366 on 20 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9986
## F-statistic:  4399 on 4 and 20 DF,  p-value: < 2.2e-16

model.matrix( mod_i )[1:5,]
##   dietA dietB dietC dietD
## 1     1     0     0     0
## 2     1     0     0     0
## 3     1     0     0     0
## 4     1     0     0     0
## 5     0     1     0     0
```

We can immediately observe that the design matrix is still invertible, but it has changed. This leads to a different interpretation of the $\boldsymbol{\beta}$. In fact, we now have:

$$\beta_j = \tau + \mu_j, \quad j \in \{1, .., g\};$$

that is, we can directly read from the output the means of the coagulation times in the individual groups.

Observation As highlighted in Chap. 9, in models without an intercept R^2 loses its meaning.

We then proceed with the model diagnostics, i.e. the verification (quantitative) of the assumptions.

```
par( mfrow = c(1,2) )

qqnorm(mod$res, pch=16, col='black',
       main='QQ-norm of residuals')
qqline( mod$res, lwd = 2, col = 1 ,lty = 2 )
```

```
shapiro.test( mod$res )
##
##  Shapiro-Wilk normality test
##
## data:  mod$res
## W = 0.97831, p-value = 0.8629

plot( mod$fit, mod$res, xlab = "Fitted", ylab = "Residuals",
      main = "Residual-Fitted plot", pch = 16 )
```

From the Shapiro test on the residuals and the graph on the left in Fig. 11.8, we conclude that the normality assumption is respected.

```
bartlett.test( coagulation$coag, coagulation$diet )
##
##  Bartlett test of homogeneity of variances
##
## data:  coagulation$coag and coagulation$diet
## Bartlett's K-squared = 1.668, df = 3, p-value = 0.6441
leveneTest( coagulation$coag, coagulation$diet )
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.6492 0.5926
##       20
```

From the Bartlett test and the graph on the right in Fig. 11.8, we can consider the homoscedasticity assumption to be valid.

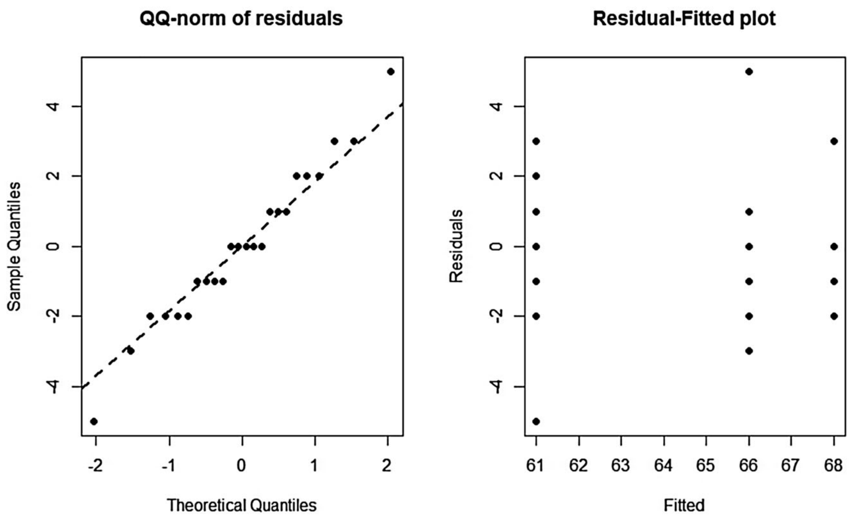

Fig. 11.8 QQ-norm of the model residuals in the left panel. Representation of residuals vs fitted values from the model in the right panel

In both graphs of Fig. 11.8 we recognise typical patterns of discrete data (both the response variable and the predictive variable are discrete). The response variable is recorded in a discrete manner (probably these are truncated values), but in itself it is a continuous variable, so it is not wrong to evaluate an ANOVA model.

11.7 To answer the question it is necessary to set up a two-way ANOVA. To do this, we try to get a graphical intuition of the effect of the two factors and their interaction on the variable of interest (survival time). We verify the model assumptions and evaluate the results obtained.

We load the `rats` data.

```
data( rats )

dim( rats )
## [1] 48  3

head( rats )
##   time poison treat
## 1 0.31      I     A
## 2 0.82      I     B
## 3 0.43      I     C
## 4 0.45      I     D
## 5 0.45      I     A
## 6 1.10      I     B
tail( rats )
##    time poison treat
## 43 0.24    III     C
## 44 0.31    III     D
## 45 0.23    III     A
## 46 0.29    III     B
## 47 0.22    III     C
## 48 0.33    III     D
names( rats )
## [1] "time"   "poison" "treat"
```

We visualise the data in Fig. 11.9 and in Fig. 11.10.

```
ggplot(rats, aes( x = treat, y = time, fill = treat ) ) +
  geom_boxplot() + scale_fill_grey() +
  scale_colour_grey() + theme_bw()
```

```
ggplot(rats, aes( x = poison, y = time, fill = poison ) ) +
  geom_boxplot() + scale_fill_grey() +
  scale_colour_grey() + theme_bw()
```

From these initial graphs, we can infer an effect of both the treatment and the poison on the survival time of the rats. It is not certain whether the assumption of homoscedasticity among the groups is respected, however appropriate tests must be performed.

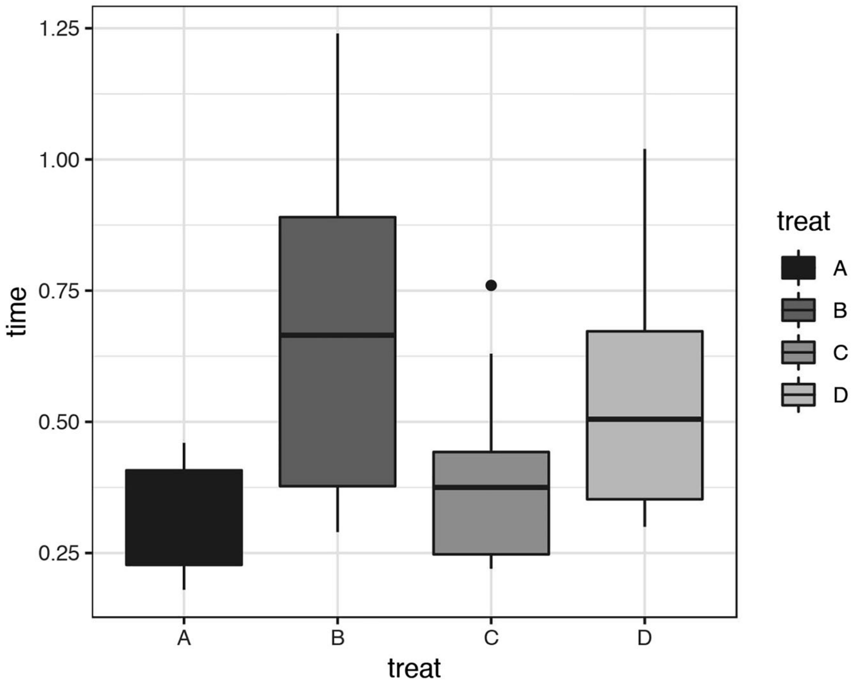

Fig. 11.9 Boxplot of the survival time of rats in relation to the type of treatment received

Since we are dealing with two factors, it is necessary to understand if it is appropriate to consider in the model also the interaction of these. A tool to investigate the possible presence of interaction between the factors is the `interaction.plot` command in Fig. 11.11 and in Fig. 11.12.

```
help(interaction.plot)
interaction.plot( rats$treat, rats$poison, rats$time )
```

```
interaction.plot( rats$poison, rats$treat, rats$time )
```

Parallel lines suggest the absence of an interaction effect between the two factors on the variable of interest (survival time of the rats). However, it is not correct to exclude the effect of the interaction of the two factors only through a graphical exploration. We therefore start from the *complete model* (which contemplates both factors and their *interaction*).

Before applying a two-way ANOVA, we must test the validity of the hypotheses of:

- Normality (in all 12 groups).
- Homogeneity of variance (between groups).

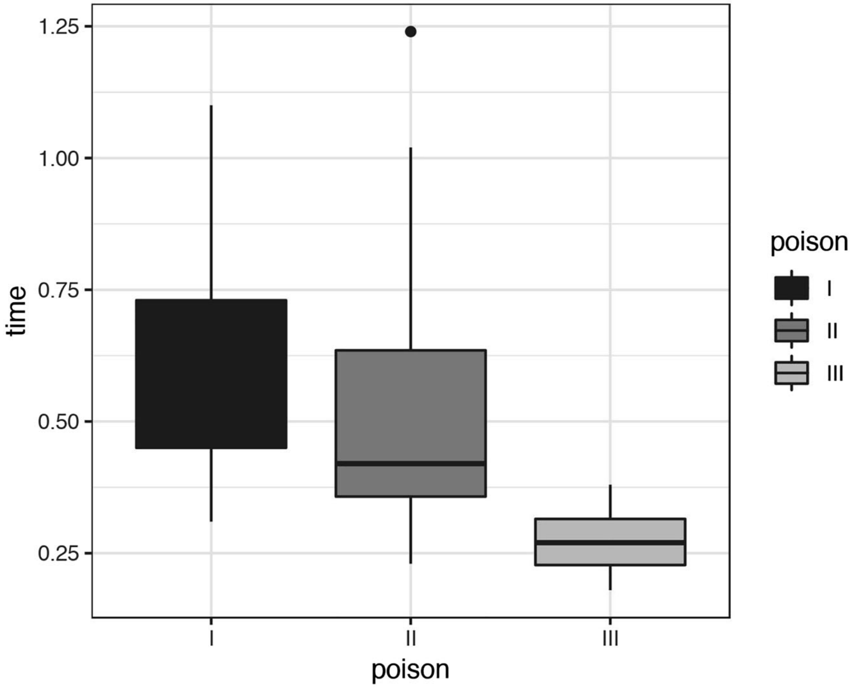

Fig. 11.10 Boxplot of the survival time of rats in relation to the type of poison administered

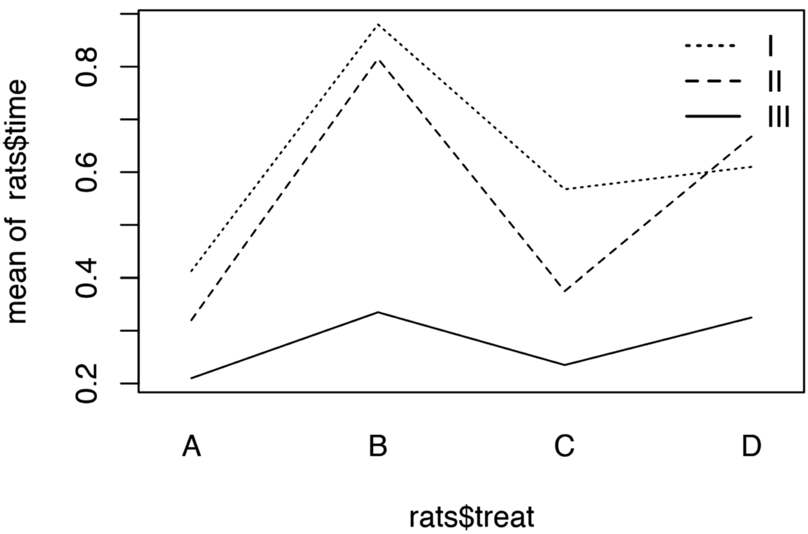

Fig. 11.11 Interaction plot to evaluate the interaction of the two factors

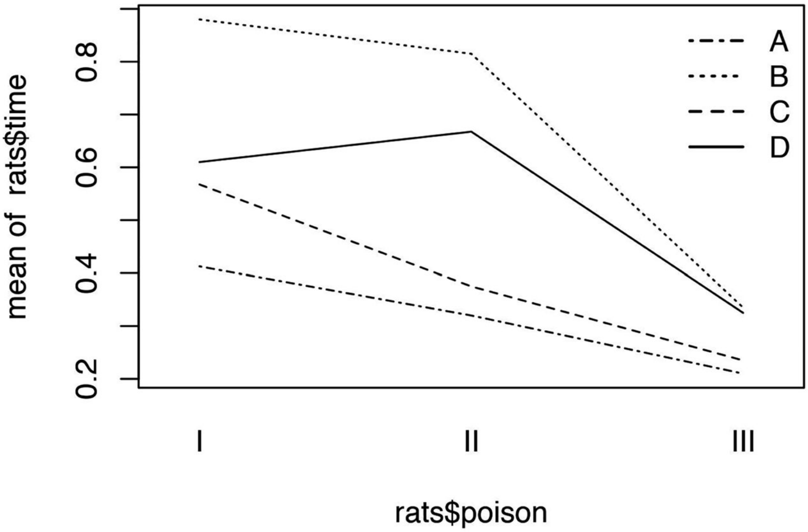

Fig. 11.12 Interaction plot to evaluate the interaction of the two factors

```
tapply( rats$time, rats$treat:rats$poison,
        function( x ) shapiro.test( x )$p )
##        A:I       A:II      A:III
## 0.07414486 0.84756406 0.57735490
##        B:I       B:II      B:III
## 0.69983383 0.70083721 0.17057001
##        C:I       C:II      C:III
## 0.40503490 0.92091109 0.97187706
##        D:I       D:II      D:III
## 0.42739119 0.90650963 0.68893644
```

By running the Shapiro test, we can conclude that the normality hypothesis is respected in all groups (although the first group, A-I, should be further investigated).

```
leveneTest( rats$time, rats$treat:rats$poison )
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value    Pr(>F)
## group 11  4.1323 0.0005833 ***
##       36
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
bartlett.test( rats$time, rats$treat:rats$poison )
##
##  Bartlett test of homogeneity of variances
##
## data:  rats$time and rats$treat:rats$poison
## Bartlett's K-squared = 45.137, df = 11, p-value = 4.59e-06
```

The hypothesis of homogeneity of variances is widely violated (observe the p-value of the Levene test and the Bartlett test).

We can consider a transformation of the variables. We opt for a `Box-Cox` type transformation, considering the complete model.

```
g = lm( time ~ poison * treat, rats )
#"*" gives the full model: linear effect AND interaction
#g = lm( time ~ poison + treat + poison : treat , rats )

b = boxcox( g, lambda = seq(-3,3,by=0.01), plotit = F )
best_lambda = b$x[ which.max( b$y ) ]
best_lambda
## [1] -0.82
```

The `boxcox` command also returns the graph in Fig. 11.13 (to obtain it, simply set `plotit = T`).

```
plot( b$x, b$y, xlab = expression(lambda),
        ylab = 'log-likelihood')
```

From Fig. 11.13 we deduce that the optimal λ is -0.82, however, as already mentioned in the chapter related to linear regression, we round λ to ensure greater interpretability. We therefore opt for $\lambda = -1$.

We then recheck the model assumptions.

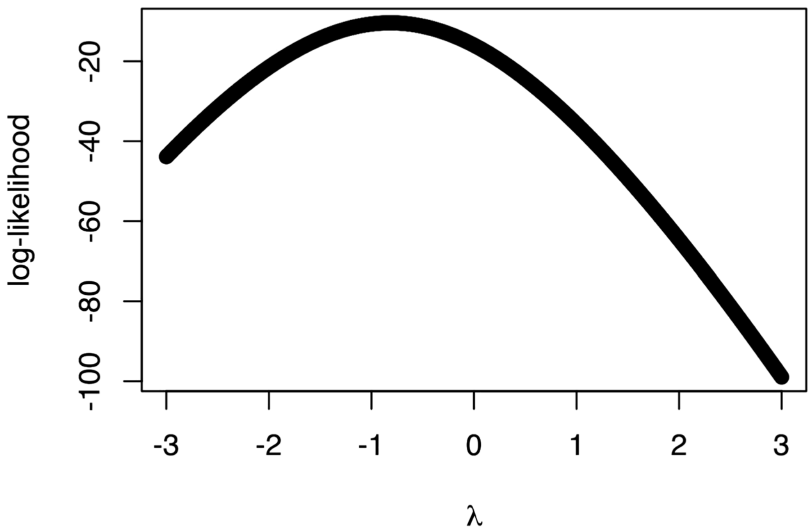

Fig. 11.13 Box-Cox type transformation: investigation of the optimal λ

```
tapply( (rats$time)^(-1), rats$treat:rats$poison,
        function( x ) shapiro.test( x )$p )
##        A:I        A:II       A:III
## 0.03115001 0.65891022 0.38884991
##        B:I        B:II       B:III
## 0.95061185 0.79724850 0.17554581
##        C:I        C:II       C:III
## 0.38264156 0.87818060 0.96578666
##        D:I        D:II       D:III
## 0.16801940 0.84342484 0.78353223

leveneTest( (rats$time)^(-1), rats$treat:rats$poison )
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group 11  1.1272 0.3698
##       36
bartlett.test( (rats$time)^(-1), rats$treat:rats$poison )
##
##  Bartlett test of homogeneity of variances
##
## data:  (rats$time)^(-1) and rats$treat:rats$poison
## Bartlett's K-squared = 9.8997, df = 11, p-value = 0.5394
```

The model assumptions are respected, apart from the normality of group A-I. We can use a two-way ANOVA model, bearing in mind that in the presence of interaction the assumptions are not fully respected.

```
g1 = lm( 1/time ~ poison * treat, data = rats )
summary( g1 )
##
## Call:
## lm(formula = 1/time ~ poison * treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76847 -0.29642 -0.06914  0.25458  1.07936
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.48688    0.24499  10.151 4.16e-12 ***
## poisonII            0.78159    0.34647   2.256 0.030252 *
## poisonIII           2.31580    0.34647   6.684 8.56e-08 ***
## treatB             -1.32342    0.34647  -3.820 0.000508 ***
## treatC             -0.62416    0.34647  -1.801 0.080010 .
## treatD             -0.79720    0.34647  -2.301 0.027297 *
## poisonII:treatB    -0.55166    0.48999  -1.126 0.267669
## poisonIII:treatB   -0.45030    0.48999  -0.919 0.364213
## poisonII:treatC     0.06961    0.48999   0.142 0.887826
## poisonIII:treatC    0.08646    0.48999   0.176 0.860928
## poisonII:treatD    -0.76974    0.48999  -1.571 0.124946
## poisonIII:treatD   -0.91368    0.48999  -1.865 0.070391 .
## ---
```

```
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.49 on 36 degrees of freedom
## Multiple R-squared:  0.8681, Adjusted R-squared:  0.8277
## F-statistic: 21.53 on 11 and 36 DF,  p-value: 1.289e-12
anova( g1 )
## Analysis of Variance Table
##
## Response: 1/time
##              Df Sum Sq Mean Sq F value    Pr(>F)
## poison        2 34.877 17.4386 72.6347 2.310e-13 ***
## treat         3 20.414  6.8048 28.3431 1.376e-09 ***
## poison:treat  6  1.571  0.2618  1.0904    0.3867
## Residuals    36  8.643  0.2401
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
```

From the model, it is clear that both the type of poison and the type of treatment influence the survival time of the mice. However, the interaction of the two factors is not significant.

We will therefore examine the reduced model and re-evaluate the hypotheses.

```
g1_red = lm( 1/time ~ poison + treat, data = rats )
summary( g1_red )
##
## Call:
## lm(formula = 1/time ~ poison + treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82757 -0.37619  0.02116  0.27568  1.18153
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.6977     0.1744  15.473  < 2e-16 ***
## poisonII      0.4686     0.1744   2.688  0.01026 *
## poisonIII     1.9964     0.1744  11.451 1.69e-14 ***
## treatB       -1.6574     0.2013  -8.233 2.66e-10 ***
## treatC       -0.5721     0.2013  -2.842  0.00689 **
## treatD       -1.3583     0.2013  -6.747 3.35e-08 ***
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
##
## Residual standard error: 0.4931 on 42 degrees of freedom
## Multiple R-squared:  0.8441, Adjusted R-squared:  0.8255
## F-statistic: 45.47 on 5 and 42 DF,  p-value: 6.974e-16
anova( g1_red )
## Analysis of Variance Table
##
## Response: 1/time
##           Df Sum Sq Mean Sq F value    Pr(>F)
## poison     2 34.877 17.4386  71.708 2.865e-14 ***
## treat      3 20.414  6.8048  27.982 4.192e-10 ***
```

```
## Residuals 42 10.214  0.2432
## ---
##Signif. codes:0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
```

We evaluate normality in three ways:

- Graphical evaluation of the residuals of the reduced model (see Fig. 11.14).
- Shapiro test on the residuals of the reduced model.
- Shapiro test on the response variable.

```
#1)
qqnorm( g1_red$res/summary( g1_red )$sigma, pch = 16,
        main = 'QQ-norm of residuals'  )
abline( 0, 1, lwd = 2, lty = 2, col = 1 )
```

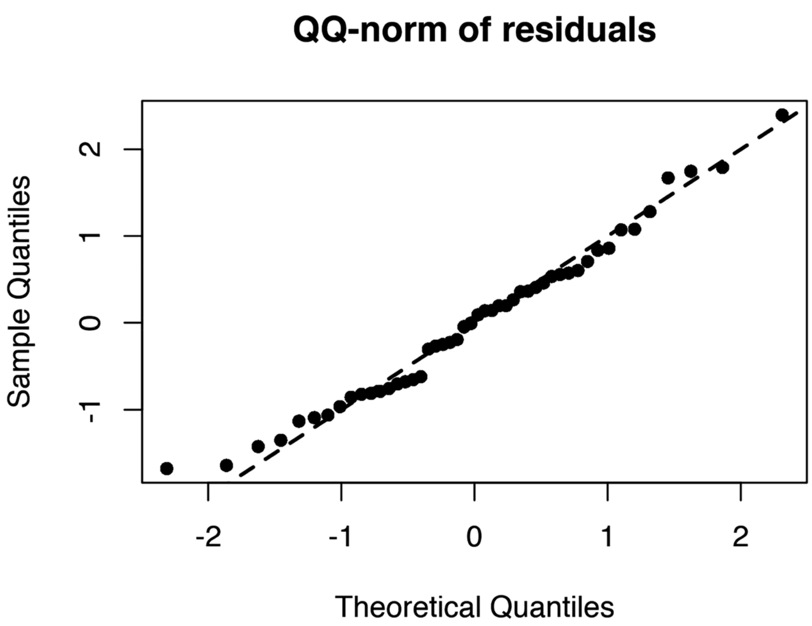

Fig. 11.14 QQ-norm of standardised residuals

```
#2)
shapiro.test( g1_red$res )
##
##  Shapiro-Wilk normality test
##
## data:  g1_red$res
## W = 0.97918, p-value = 0.5451

#3)
tapply( 1/rats$time, rats$poison,
            function( x ) shapiro.test( x )$p )
##         I        II       III
## 0.1672488 0.8944364 0.3944087

tapply( 1/rats$time, rats$treat,
            function( x ) shapiro.test( x )$p )
##         A         B         C         D
## 0.2221106 0.1021497 0.3632241 0.2712347
```

In all three ways, we arrive at the same conclusion.

Finally, we evaluate the homogeneity of variance between the groups.

```
leveneTest( 1/rats$time, rats$poison )
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2   1.715 0.1915
##       45
leveneTest( 1/rats$time, rats$treat )
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3   0.614 0.6096
##       44

bartlett.test( 1/rats$time, rats$poison )
##
##  Bartlett test of homogeneity of variances
##
## data:  1/rats$time and rats$poison
## Bartlett's K-squared = 3.1163, df = 2, p-value = 0.2105
bartlett.test( 1/rats$time, rats$treat )
##
##  Bartlett test of homogeneity of variances
##
## data:  1/rats$time and rats$treat
## Bartlett's K-squared = 1.5477, df = 3, p-value = 0.6713
```

The Levene and Bartlett tests confirm the hypothesis.

Therefore, we conclude that rats given different poisons or treatments have different survival times.

Chapter 12
Summary Exercises

12.1 Exercises

Exercise 12.1 Let $X_1, ..., X_n$ be a random sample of size n from a Gaussian law with mean $\ln\theta$ and variance 4, with θ a positive unknown parameter.

(a) Given $U_n = c_n \sum_{i=1}^n e^{X_i}$, determine the constant c_n so that U_n is an unbiased estimator of θ.
Hint: use the fact that, if $Y \sim N(m, \sigma^2)$ then, for $t \in \mathbb{R}$, $\mathrm{E}\left[e^{tY}\right] = e^{tm+\frac{1}{2}t^2\sigma^2}$.
(b) Calculate $\mathrm{Var}[U_n]$ for c_n determined in the previous point. Determine whether the estimator U_n is consistent and asymptotically normal.
(c) Construct another unbiased estimator V_n of θ, starting from the one obtained with the method of moments. Determine whether V_n is consistent and determine its asymptotic law.
(d) Determine which of the two estimators U_n and V_n of θ is preferable and justify the choice.
(e) Construct an asymptotic confidence interval of level $1-\alpha$ for θ based on V_n.

Exercise 12.2 Let $X_1, \ldots, X_n$ be a random sample from a distribution with law:

$$f(x;\theta) = 4\frac{(x-1)^3}{(\theta-1)^4}\mathbb{I}_{(1,\theta)}(x);$$

where θ is a positive unknown parameter, $\theta > 1$.

(a) Determine a statistic T sufficient for θ.
(b) Using the definition of completeness, determine whether T is complete for θ.
(c) Find the UMVUE for θ.
(d) Construct a pivot quantity Q for θ.
(e) Construct the minimum length confidence interval of level $1-\alpha$ based on Q.

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8_12

Exercise 12.3 Let $X_1, .., X_n$ be a random sample of size $n \geq 1$ where each variable has law:

$$f(x;\theta) = \frac{\theta}{x^2} e^{-\frac{\theta}{x}} \mathbb{I}_{(0,+\infty)}(x);$$

where θ is a positive real parameter $\theta > 0$.

(a) Find W a sufficient, minimal and complete statistic for θ.
(b) Construct the maximum likelihood estimator $\hat{\theta}$ for θ.
(c) Determine whether $\hat{\theta}$ is consistent for θ.
(d) Construct the UMVUE for θ.

Now consider a sample of size $n = 1$.

(e) Construct the UMP test of level α for the verification of the hypotheses:

$$H_0: \theta = \theta_0 \qquad \text{vs} \qquad H_1: \theta = \theta_1;$$

with $\theta_1 > \theta_0$.
(f) Construct the UMP test of level α for the verification of the hypotheses:

$$H_0: \theta = \theta_0 \qquad \text{vs} \qquad H_1: \theta > \theta_0.$$

Exercise 12.4 Let $X_1, .., X_n$ be a random sample of size $n \geq 1$ where each variable has law:

$$f(x;\theta) = \frac{x}{K} \mathbb{I}_{\{1,2,\ldots,\theta\}}(x);$$

where θ is an *integer* parameter such that $\theta \geq 1$.

(a) Calculate the constant K as a function of θ.
(b) Construct the moments estimator $\bar{\theta}$ for θ and determine whether it is consistent. Does it always provide admissible estimates? *(Hint: it may be useful to remember that* $\sum_{j=1}^{n} j^2 = n(n+1)(2n+1)/6$).
(c) Construct the maximum likelihood estimator $\hat{\theta}$ for θ and determine whether it is consistent. Does it always provide admissible estimates?
(d) Construct the critical region based on the likelihood ratio for the hypothesis test:

$$H_0: \theta \leq \theta_0 \qquad \text{vs} \qquad H_1: \theta > \theta_0.$$

Exercise 12.5 Let $X_1, .., X_n$ be a random sample of size $n \geq 1$ where each variable has law

$$f(x;\theta) = \frac{2x}{\theta^2}\mathbb{I}_{(0,\theta)}(x);$$

with parameter $\theta > 0$.

(a) Find a statistic T sufficient and minimal for θ.
(b) Calculate the moments estimator $\overline{\theta}$ for θ.
(c) Calculate the mean square error of $\overline{\theta}$.
(d) Now consider the parameter $\theta = \theta_0$ known and fix $n = 1$. Construct the UMP test of level $\alpha \in (0, 1)$ for the hypothesis test:

$$H_0 : X \sim f(x;\theta_0) \qquad \text{vs} \qquad H_1 : X \sim U(0,\theta_0).$$

(e) Calculate the power of the test constructed in point (d) and determine whether the test is unbiased.

Exercise 12.6 Let $X_1, \ldots, X_n$ be a random sample from a distribution with the following probability density:

$$f(x;\theta) = 2\frac{\theta^2}{x^3}\mathbb{I}_{(\theta,\infty)}(x), \qquad \theta > 0.$$

(a) Calculate the maximum likelihood estimator $\widehat{\theta}$ for θ.
(b) Calculate the probability density of $\widehat{\theta}$.
(c) Find the critical region of level $\alpha \in (0, 1)$ based on the likelihood ratio for the hypothesis test: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0, \theta_0 > 0$.
(d) Using the critical region constructed in point (c), find a confidence interval for θ of level $(1 - \alpha)$.
(e) Using the pivotal quantity $Q = \widehat{\theta}/\theta$, find the constant $c > 0$ such that the confidence interval $(0, \widehat{\theta}c)$ for θ is of level $(1 - \alpha)$.

Exercise 12.7 Let $X_1, \ldots, X_n$ be a random sample from a distribution with law:

$$f(x;\theta,b) = \frac{\theta x^{\theta-1}}{b^\theta}\mathbb{I}_{(0,b)}(x);$$

where θ and b are two unknown positive parameters, $\theta > 0,\ b > 0$.

(a) Find a sufficient and minimal statistic for (θ, b).
(b) Assume that θ is known. Determine the maximum likelihood estimator $\hat{b}_{ML}$ for b.
(c) Determine the law of $\hat{b}_{ML}$ and study its consistency.
(d) Determine a pivotal quantity Q for b.

(e) Determine the confidence interval of level $1-\alpha$ for b based on Q of minimum length.

Exercise 12.8 Let X be a random variable with density

$$f(x;\theta) = \frac{\theta^k e^{kx} e^{-\theta e^x}}{\Gamma(k)} \mathbb{I}_{(-\infty,+\infty)}(x);$$

where k is a known positive parameter and θ is an unknown positive parameter, $\theta > 0$.

(a) Prove that $W = \sum_{i=1}^n e^{X_i}$ is a sufficient, minimal and complete statistic for θ.
(b) Calculate the law of e^X.
(c) Calculate and recognise the law of W.
(d) Construct a UMP test of level α for $H_0\colon \theta \le \theta_0$ vs $H_1\colon \theta > \theta_0$.
(e) Construct a pivotal quantity for θ based on W, and derive a Confidence Interval of level $1-\alpha$ for θ.

Exercise 12.9 Let $X_1, \ldots, X_n$ be a random sample from a distribution with law:

$$f(x;\theta) = 2\theta x \exp\{(-\theta x^2)\}\mathbb{I}_{[0,+\infty)}(x);$$

where θ is an unknown positive parameter, $\theta > 0$.

(a) Calculate the mean of X_i.
(b) Determine, using the method of moments, the estimator $\hat{\theta}_{MOM}$ for θ.
(c) Determine a statistic T sufficient minimal and complete for θ.
(d) Determine the law of T.
(e) Determine the maximum likelihood estimator $\hat{\theta}_{ML}$ for θ.
(f) Establish whether $\hat{\theta}_{MOM}$ and $\hat{\theta}_{ML}$ are consistent for θ.
(g) Determine the asymptotic law of $\hat{\theta}_{ML}$.
(h) Knowing that $\mathrm{Var}[X_i] = \frac{0.21}{\theta}$ determine, using the Delta Method 1.17, the asymptotic law of $\hat{\theta}_{MOM}$.
(i) Calculate the asymptotic relative efficiency of $\hat{\theta}_{ML}$ with respect to $\hat{\theta}_{MOM}$, i.e. ARE($\hat{\theta}_{ML}$;$\hat{\theta}_{MOM}$).

Exercise 12.10 Consider the following family of functions defined for every $\theta \in \mathbb{R}$

$$f(x;\theta) = c\left(1-(x-\theta)^2\right)\mathbf{1}_{[\theta,\theta+1]}(x).$$

(a) Determine the constant c so that the function $f_\theta(x)$ is a probability density for every $\theta \in \mathbb{R}$.

Consider a sample X of unit size with probability distribution $f(x;\theta)$ with c determined in point (a).

(b) Calculate the maximum likelihood estimator $\hat{\theta}_{ML}$ for θ.

(c) Prove that $Q = 1 - (\hat{\theta}_{ML} - \theta)^2$ is a pivotal quantity.
(d) Calculate a confidence interval for θ based on the pivotal quantity Q, using the quantiles of Q $a = 0.5$ and $b = 0.9$.
(e) Determine the confidence level $1 - \alpha$, of the interval constructed in point (d).
(f) Determine the critical region, of level α calculated in point (e), of the test H_0: $\theta = \theta_0$ vs H_1: $\theta \neq \theta_0$.

Exercise 12.11 Let $X_1, \ldots, X_n$ be a random sample from a Gamma(2,1/θ) with $\theta > 0$. We then have

$$f(x;\theta) = \theta^{-2}\, x\, \mathrm{e}^{-x/\theta} I_{(0,+\infty)}(x).$$

(a) Determine a sufficient and complete statistic for θ.
(b) Determine the maximum likelihood estimator $\hat{\theta}_n$ for θ.
(c) Show that $\hat{\theta}_n$ coincides with the estimator $\bar{\theta}_n$ obtained by the method of moments.
(d) What is the law of $\hat{\theta}_n$?
(e) Is $\hat{\theta}_n$ biased?
(f) Is $\hat{\theta}_n$ UMVUE?
(g) Propose a confidence interval for θ of level 0.99.

12.2 Solutions

12.1

(a) Let

$$U_n = c_n \sum_{i=1}^{n} \mathrm{e}^{X_i}.$$

We calculate the mean of U_n:

$$\mathbb{E}[U_n] = c_n \sum_{i=1}^{n} \mathbb{E}[e^{X_i}] = c_n \sum_{i=1}^{n} e^{\log(\theta)+2} = c_n n e^2 \theta;$$

where we have exploited the following relation:

$$\mathbb{E}\left[\mathrm{e}^{tY}\right] = \mathrm{e}^{tm + \frac{1}{2}t^2\sigma^2};$$

choosing $t = 1$, $m = \log(\theta)$ and $\sigma^2 = 4$. This relation holds because Y is a Gaussian random variable.

Imposing $\mathbb{E}[U_n] = \theta$, we get: $c_n = \frac{1}{ne^2}$.

(b)

$$\begin{aligned}
\text{Var}\,[U_n] &= \frac{1}{n^2 e^4} \sum_{i=1}^{n} \left(\mathbb{E}[e^{2X_i}] - (e^2\theta)^2 \right) = \\
&= \frac{1}{n^2 e^4} \sum_{i=1}^{n} \left(e^{2\log\theta + \frac{1}{2}4\cdot 4} - e^4\theta^2 \right) = \\
&= \frac{1}{n^2 e^4} n \left(\theta^2 (e^8 - e^4) \right) = \frac{\theta^2}{n}(e^4 - 1).
\end{aligned}$$

Therefore, since U_n is unbiased and $Var(U_n) \to 0$, we can conclude that U_n is a consistent estimator (see Theorem 8.1).
Furthermore, by the CLT:

$$\sqrt{n}\,(U_n - \theta) \to N\left(0, \theta^2(e^4 - 1)\right).$$

So U_n is asymptotically normal.

(c)

$$\mathbb{E}[X] = \log(\theta) \implies \hat{\theta}_{MOM} = e^{\overline{X}_n}.$$

Given that:

$$\overline{X}_n \sim N\left(\log(\theta), \frac{4}{n}\right);$$

then:

$$\mathbb{E}[\hat{\theta}_{MOM}] = \mathbb{E}[e^{X_n}] = e^{\log\theta + \frac{1}{2}\frac{4}{n}} = \theta e^{\frac{2}{n}}.$$

Therefore:

$$V_n = e^{-\frac{2}{n}} e^{\overline{X}_n}.$$

Given that $\overline{X}_n \xrightarrow{q.c.} \log(\theta)$ we have that $V_n \xrightarrow{q.c.} \theta$ and therefore it is consistent.
Furthermore, given that:

$$\sqrt{n}\left(\overline{X}_n - \log(\theta)\right) \xrightarrow{\mathcal{L}} N(0, 4);$$

using the Delta Method 1.17 with $g(x) = e^x$ we have that:

$$\sqrt{n}\left(e^{\overline{X}_n} - \theta\right) \xrightarrow{\mathcal{L}} N(0, 4\theta^2).$$

Now:

$$\sqrt{n}\,(V_n - \theta) = \sqrt{n} \underbrace{\underbrace{\left(V_n - e^{\overline{X}_n}\right)}_{=\sqrt{n}\left(e^{\overline{X}_n}\left(e^{-\frac{2}{n}-1}\right)\right)}}_{\xrightarrow{q.c.}\theta} + \underbrace{\sqrt{n}\left(e^{\overline{X}_n} - \theta\right)}_{\xrightarrow{\mathcal{L}} N(0,4\theta^2)}.$$

Therefore:

$$\sqrt{n}\,(V_n - \theta) \xrightarrow{\mathcal{L}} N(\theta, 4\theta^2).$$

(d) We need to compare $\theta^2(e^4 - 1)$ with $4\theta^2$. Since $4 < (e^4 - 1)$, we prefer V_n.

(e) Using the Slutsky Theorem 1.15, we can state that:

$$IC_{1-\alpha} = \left[V_n \pm \frac{2V_n}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right].$$

12.2

(a) Consider the density:

$$f(x;\theta) = 4\frac{(x-1)^3}{(\theta-1)^4}\mathbb{I}_{[1,\theta]}(x), \quad \theta > 1.$$

Given that the joint law is:

$$f(\boldsymbol{x};\theta) = \frac{4^n \prod_{i=1}^n (x_i - 1)^3}{(\theta-1)^{4n}}\mathbb{I}_{(0,\theta)}(X_{(n)});$$

we can conclude, thanks to Theorem 2.1, that the statistic $T = X_{(n)}$ is sufficient for θ.

(b)

$$F_{X_{(n)}}(t) = (\mathbb{P}\{X_i \le t\})^n = \left[\int_1^t 4\frac{(x-1)^3}{(\theta-1)^4}\,\mathrm{d}x\right]^n =$$

$$= \left(\frac{t-1}{\theta-1}\right)^{4n} \qquad t \in [1, \theta];$$

from which

$$f_{X_{(n)}}(t) = \frac{4n(t-1)^{4n-1}}{(\theta-1)^{4n}}\mathbb{I}_{[1,\theta]}(t).$$

Using the definition of completeness, we obtain:

$$0 = \mathbb{E}[g(T)] = \int_1^\theta 4n\frac{(t-1)^{4n-1}}{(\theta-1)^{4n}}g(t)\,\mathrm{d}t.$$

This holds $\forall\theta$ if and only if $g(t) = 0$. Therefore T is a complete statistic.

(c) Let's calculate $\mathbb{E}[T]$:

$$\mathbb{E}[T] = \int_1^\theta t\frac{4n(t-1)^{4n-1}}{(\theta-1)^{4n}} = \frac{t(t-1)^{4n}}{(\theta-1)^{4n}}\bigg|_1^\theta - \frac{1}{(\theta-1)^{4n}}\int_1^\theta (t-1)^{4n}\,\mathrm{d}t =$$
$$= \theta - \frac{\theta-1}{4n+1} = \frac{4n\theta+1}{4n+1};$$

therefore UMVUE will be:

$$\frac{X_{(n)}(4n+1)-1}{4n}.$$

(d) Considering:

$$Q = \frac{X_{(n)}-1}{\theta-1}.$$

$$F_Q(t) = \mathbb{P}\{X_{(n)} \le 1+t(\theta-1)\} = \left(\frac{1+t(\theta-1)-1}{(\theta-1)}\right)^{4n} = t^{4n} \qquad t \in [0,1].$$

We conclude that Q is a pivot quantity.

(e)

$$\mathbb{P}\{a \le Q \le b\} = b^{4n} - a^{4n} = 1-\alpha.$$

Now:

$$a \le \frac{X_{(n)}-1}{\theta-1} \le b \iff 1+\frac{X_{(n)}-1}{b} \le \theta \le 1+\frac{X_{(n)}-1}{a}.$$

Therefore the length, l, of the IC is proportional to $\frac{1}{a} - \frac{1}{b}$.

$$\frac{\partial l}{\partial a} = -\frac{1}{a^2} - \frac{b'(a)}{b^2} = 0.$$

Moreover, deriving the constraint, we get:

$$4nb^{4n-1}b'(a) - 4na^{n-1} = 0.$$

Therefore $b'(a) = \left(\frac{a}{b}\right)^{4n-1}$, from which:

$$\frac{\partial l}{\partial a} = -\frac{1}{a^2} + \frac{1}{b^2}\left(\frac{a}{b}\right)^{4n-1} = \frac{a^{4n+1} - b^{4n+1}}{a^2 b^{4n+1}} < 0.$$

The minimum length is obtained for maximum a i.e. $b = 1$. Hence: $1 - a^{4n} = 1 - \alpha \implies a = \sqrt[4n]{\alpha}$.

$$IC = \left[X_{(n)}; 1 + \frac{X_{(n)} - 1}{\sqrt[4n]{\alpha}}\right].$$

12.3

(a) Given that:

$$f(x;\theta) = \frac{\theta}{x^2} e^{-\frac{\theta}{x}} \mathbb{I}_{[0,+\infty)}(x)$$

belongs to the exponential family,

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

is a sufficient statistic for θ. Moreover, given that:

$$w(\theta) = -\theta : \mathbb{R}^+ \to \mathbb{R}^-$$

and $\mathbb{R}^-$ contains an open set of $\mathbb{R}$, we can conclude that $T(\boldsymbol{X})$ is a sufficient and complete statistic for θ. Consequently, it is also minimal.

(b)

$$L(\theta;\boldsymbol{x}) = \frac{\theta^n}{\prod_{i=1}^n x_i^2} e^{-\theta \sum \frac{1}{x_i}} \prod_{i=1}^{n} \mathbb{I}_{[0,+\infty)}(x_i).$$

$$l(\theta;\boldsymbol{x}) \propto n \log(\theta) - \theta \sum \frac{1}{x_i}.$$

$$\frac{\partial l(\theta;\boldsymbol{x})}{\partial \theta} \geq 0 \iff \frac{n}{\theta} \geq \sum \frac{1}{x_i} \iff \theta \leq \frac{n}{\sum \frac{1}{x_i}}.$$

Therefore:

$$\hat{\theta}_{MLE} = \frac{n}{\sum \frac{1}{X_i}}.$$

(c) Let $Y = \frac{1}{X}$, $f_Y(y) = \theta y^2 e^{-\theta y} \frac{1}{y^2} \sim \mathcal{E}(\theta)$. Therefore:

$$\frac{\sum \frac{1}{x_i}}{n} \stackrel{q.c.}{\to} \frac{1}{\theta} \implies \hat{\theta}_{MLE} \stackrel{q.c.}{\to} \theta.$$

$\hat{\theta}_{MLE}$ is a consistent estimator.

(d) $\mathbb{E}[\hat{\theta}_{MLE}]$ turns out to be $\frac{n}{n-1}\theta$ (the properties of the gamma distribution are exploited). Then $\frac{n-1}{n}\hat{\theta}_{MLE}$ is UMVUE, as it is an unbiased estimator of θ, a function of sufficient and minimal statistic.

(e) Let's consider the test:

$$H_0: \theta = \theta_0 \qquad \text{vs} \qquad H_1: \theta = \theta_1;$$

with $\theta_1 > \theta_0$. To construct the rejection region of the test, we apply the N-P Theorem 6.1.

$$\begin{aligned} R &= \left\{\theta_1 e^{-\frac{\theta_1}{x}} \frac{1}{x^2} > k\theta_0 e^{-\frac{\theta_0}{x}} \frac{1}{x^2}\right\} = \\ &= \left\{e^{-\frac{1}{x}(\theta_1 - \theta_0)} > k\frac{\theta_0}{\theta_1}\right\} = \\ &= \{x > h\}. \end{aligned}$$

The last equality is justified by the fact that $e^{-\frac{1}{x}(\theta_1-\theta_0)}$ is increasing in x. Therefore, imposing:

$$\alpha = \mathbb{P}_{\theta_0}\{X > h\} = \int_h^{+\infty} \frac{\theta_0}{x^2} e^{-\frac{\theta_0}{x}} \,\mathrm{d}x = -e^{-\frac{\theta_0}{x}}\Big|_h^{+\infty} = 1 - e^{-\frac{\theta_0}{h}};$$

we have that:

$$h = -\frac{\theta_0}{\log(1-\alpha)}.$$

Therefore the rejection region is:

$$R = \left\{x > -\frac{\theta_0}{\log(1-\alpha)}\right\}.$$

(f) From the previous point we can observe that the rejection region does not depend on θ_1, therefore the rejection region for the test at point (f), coincides with that calculated at point (e).

12.4

(a) We impose that:

$$1 = \sum_{x=1}^{\theta} \frac{x}{k} = \frac{1}{k} \sum_{x=1}^{\theta} x = \frac{1}{k} \frac{\theta(\theta+1)}{2}.$$

Therefore $k = \frac{\theta(\theta+1)}{2}$.

(b)

$$\mathbb{E}[X] = \frac{2}{\theta(\theta+1)} \sum_{x=1}^{\theta} x^2 = \frac{2}{\theta(\theta+1)} \frac{\theta(\theta+1)(2\theta+1)}{6} = \frac{2\theta+1}{3}.$$

Therefore:

$$\overline{X}_n = \frac{2\hat{\theta}_{MOM}+1}{3} \implies \hat{\theta}_{MOM} = \frac{3\overline{X}_n - 1}{2}.$$

Furthermore:

$$\overline{X}_n \overset{q.c.}{\to} \frac{2\theta+1}{3}.$$

Therefore $\hat{\theta}_{MOM}$ is a consistent estimator and the estimates are always reliable since $\theta \in \mathbb{N}$.

(c)

$$L(\theta; \boldsymbol{x}) = \frac{2^n}{\theta^n(\theta+1)^n} \prod_{i=1}^{n} x_i \prod_{i=1}^{n} \mathbb{I}_{\{1,\ldots,\theta\}}(x_i) =$$

$$= \frac{2^n}{\theta^n(\theta+1)^n} \prod_{i=1}^{n} x_i \mathbb{I}_{\{X_{(n)},+\infty\}}(\theta).$$

The likelihood is decreasing in θ, therefore:

$$\hat{\theta}_{MLE} = X_{(n)}.$$

We now evaluate the consistency of the estimator.

$$F_{X_{(n)}}(t) = \left(F_{X_i}(t)\right)^n = \left(\sum_{i=1}^{n} \frac{2i}{\theta(\theta+1)}\right)^n - \left(\frac{t(t+1)}{\theta(\theta+1)}\right)^n.$$

Therefore $F_{X_{(n)}}$ is piecewise constant and:

$$F_{X_{(n)}}(t) \to \delta_\theta(t) \implies X_{(n)} \overset{q.c.}{\to} \theta.$$

The values of $X_{(n)}$ are always admissible.

(d) We consider the following test:

$$H_0: \theta \le \theta_0 \qquad \text{vs} \qquad H_1: \theta > \theta_0.$$

We identify the rejection region of the test through LRT:

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \le \theta_0} L(\theta; \boldsymbol{x})}{L(\hat{\theta}_{MLE}; \boldsymbol{x})}.$$

$$\sup_{\theta \le \theta_0} L(\theta; \boldsymbol{x})\Big|_{\theta \le \theta_0} = \begin{cases} X_{(n)} & \text{if } X_{(n)} \le \theta_0; \\ 0 & \text{if } X_{(n)} > \theta_0. \end{cases}$$

Therefore:

$$\lambda(\boldsymbol{x}) = \begin{cases} 1 & \text{if } X_{(n)} \le \theta_0; \\ 0 & \text{if } X_{(n)} > \theta_0. \end{cases}$$

Then the rejection region is:

$$R = \left\{X_{(n)} > \theta_0\right\}.$$

12.5

(a)

$$f(\boldsymbol{x}; \theta) = \frac{2^n}{\theta^{2n}} \prod_{i=1}^{n} x_i \, \mathbb{I}_{(0,\theta)}(x_{(n)}).$$

Therefore $X_{(n)}$ is a sufficient statistic and, using the L-S Theorem 2.3, we conclude that $X_{(n)}$ is a minimal sufficient statistic.

(b)

$$\mathbb{E}[X] = \int_0^{\theta} \frac{2x^2}{\theta^2}\,\mathrm{d}x = \frac{2}{\theta^2}\frac{x^3}{3}\bigg|_0^{\theta} = \frac{2}{3}\theta.$$

Therefore $\hat{\theta}_{MOM} = \frac{3}{2}\overline{X}_n$.

(c)

$$\begin{aligned}
\mathrm{Var}(\hat{\theta}_{MOM}) &= \frac{9}{4}\mathrm{Var}(\overline{X}_n) = \frac{9}{4n}\mathrm{Var}(X_i) = \\
&= \frac{9}{4n}\left(\int_0^{\theta}\frac{2x^3}{\theta^2}\,\mathrm{d}x - \left(\frac{2}{3}\theta\right)^2\right) = \\
&= \frac{9}{4n}\left(\frac{2}{\theta^2}\frac{\theta^4}{4} - \frac{4}{9}\theta^2\right) = \frac{9}{4n}\left(\frac{\theta^2}{2} - \frac{4}{9}\theta^2\right) = \\
&= \frac{9}{4n}\left(\frac{9\theta^2 - 8\theta^2}{18}\right) = \frac{\theta^2}{8n}.
\end{aligned}$$

Therefore:

$$MSE(\hat{\theta}_{MOM}) = \frac{\theta^2}{8n}.$$

(d) Let's consider the test:

$$H_0:\ X \sim f(x;\theta_0) \qquad \text{vs} \qquad H_1: X \sim U(0,\theta_0).$$

Applying the N-P Theorem 6.1:

$$R = \left\{\frac{1}{\theta_0}\,\mathbb{I}_{(0,\theta_0)}(x) > k\frac{2x}{\theta_0^2}\,\mathbb{I}_{(0,\theta_0)}(x)\right\} \iff \left\{x < \tilde{k}\right\}.$$

By imposing:

$$\alpha = \mathbb{P}_{H_0}(X \in R) = \int_0^{\tilde{k}}\frac{2x}{\theta_0^2}\,\mathrm{d}x = \frac{\tilde{k}^2}{\theta_0^2},$$

we obtain $\tilde{k} = \theta\sqrt{\alpha}$.

(e) The power of the test is given by:

$$\mathbb{P}_{H_1}\left\{X < \sqrt{\alpha}\theta_0\right\} = \sqrt{\alpha}.$$

Since: $\sqrt{\alpha} > \alpha$, we conclude that the test is unbiased.

12.6

(a)

$$L(\theta; \boldsymbol{x}) = \frac{2^n \theta^{2n}}{\prod_{i=1}^n x_i^3} \, \mathbb{I}_{(0, x_{(1)})}(\theta).$$

Therefore:

$$\hat{\theta}_{MLE} = X_{(1)}.$$

(b)

$$F_{X_{(1)}}(t) = \mathbb{P}\left\{X_{(1)} \leq t\right\} = 1 - \mathbb{P}\left\{X_{(1)} > t\right\} = 1 - \left(\mathbb{P}\left\{X_i > t\right\}\right)^n.$$

$$\mathbb{P}\left\{X_i > t\right\} = \int_t^{+\infty} \frac{2\theta^2}{x^3} \, \mathrm{d}x = \begin{cases} 1, & t \leq \theta; \\ \left(\frac{\theta}{t}\right)^2, & t > \theta. \end{cases}$$

From which:

$$F_{X_{(1)}}(t) = \begin{cases} 0, & t \leq \theta; \\ 1 - \left(\frac{\theta}{t}\right)^{2n}, & t > \theta. \end{cases}$$

$$f_{X_{(1)}}(t) = \frac{2n\theta^{2n}}{t^{2n+1}} \, \mathbb{I}_{(\theta, +\infty)}(t).$$

(c) Let's consider the test:

$$H_0 : \theta = \theta_0 \qquad \text{vs} \qquad H_1 : \theta \neq \theta_0, \quad \theta_0 > 0.$$

We construct the LRT:

$$\lambda(\boldsymbol{x}) = \frac{2^n \theta_0^{2n}}{\prod x_i^3} \, \mathbb{I}_{(0, x_{(1)})}(\theta_0) \cdot \frac{\prod x_i^3}{2^n x_{(1)}^{2n} \, \mathbb{I}_{(0, x_{(1)})}(x_{(1)})} =$$

$$= \left(\frac{\theta_0}{x_{(1)}}\right)^{2n} \mathbb{I}_{(\theta_0, +\infty)}(x_{(1)});$$

from which:

$$R = \{\lambda(\boldsymbol{x}) \leq c\} \iff \left\{X_{(1)} \leq \theta_0\right\} \cup \left\{X_{(1)} \geq k\right\}.$$

By imposing:

$$\alpha = \mathbb{P}_{\theta_0}\{X \in R\} = \mathbb{P}\{X_{(1)} \leq \theta_0\} + \mathbb{P}_{\theta_0}\{X_{(1)} \geq k\} = \left(\frac{\theta_0}{k}\right)^{2n};$$

from which $k = \frac{\theta_0}{\sqrt[2n]{\alpha}}$.
The rejection region is therefore:

$$R = \left\{X_{(1)} \geq \frac{\theta_0}{\sqrt[2n]{\alpha}}\right\}.$$

(d) We observe that:

$$R^C = \left\{\theta_0 \leq X_{(1)} \leq \frac{\theta_0}{\sqrt[2n]{\alpha}}\right\};$$

from which:

$$IC_{(1-\alpha)}(\theta) = \left\{X_{(1)} \sqrt[2n]{\alpha} \leq \theta \leq X_{(1)}\right\}$$

is a confidence interval of level $1-\alpha$.

(e) Let $Q = \frac{X_{(1)}}{\theta}$:

$$F_Q(t) = \mathbb{P}\{X_{(1)} \leq t\} = \begin{cases} 0 & t \leq 1; \\ 1 - \left(\frac{1}{t}\right)^{2n} & t > 1. \end{cases}$$

Let:

$$IC = [0; cX_{(1)}] \implies (1-\alpha) = \inf_{\theta \geq 0} \mathbb{P}_\theta(\theta \leq cX_{(1)}) = \inf_{\theta \geq 0}\left(Q \geq \frac{1}{c}\right) = c^{2n}.$$

We impose the confidence level equal to $1-\alpha$, that is $c^{2n} = 1-\alpha \implies c = \sqrt[2n]{1-\alpha}$.
We conclude that:

$$IC_{1-\alpha}(\theta) = \left[0;\ \sqrt[2n]{1-\alpha}X_{(1)}\right].$$

12.7

(a)

$$f(\boldsymbol{x};\theta, b) = \frac{\theta^n}{b^{n\theta}}\left(\prod x_i^{\theta-1}\right)\mathbb{I}_{(0,b)}(X_{(n)});$$

hence, exploiting the L-S Theorem 2.3, we conclude that $\left(X_{(n)}; \prod X_i\right)$ is a sufficient and minimal statistic for (b, θ).

(b) Let θ be known.

$$L(b; \boldsymbol{x}, \theta) = \frac{\theta^n}{b^{n\theta}} \left(\prod x_i^{\theta-1}\right) \mathbb{I}_{(X_{(b)}, +\infty)}(b)$$

is decreasing in b, hence:

$$\hat{b}_{MLE} = X_{(n)}.$$

(c)

$$F_{X_{(n)}}(t) = \left(F_{X_i}(t)\right)^n = \begin{cases} 0 & t \in (-\infty; 0); \\ \left(\frac{t}{b}\right)^{n\theta} & t \in [0; b]; \\ 1 & t \in (b; +\infty). \end{cases}$$

Therefore $X_{(n)} \xrightarrow{\mathcal{L}} b$ and is consistent for b.

(d) Let $Q = \frac{X_{(n)}}{b}$, then:

$$\mathbb{P}\{Q \le t\} = t^{n\theta}.$$

Therefore Q is a pivotal quantity.

(e) CI for b:

$$IC(b) = \left[a \le \frac{X_n}{b} \le c\right] \implies \left[\frac{X_{(n)}}{c}; \frac{X_{(n)}}{a}\right];$$

with constraint: $1 - \alpha = c^{n\theta} - a^{n\theta}$.
The length of the interval is proportional to $\left(\frac{1}{a} - \frac{1}{c}\right)$.
Consider $c = c(a)$. Differentiating the constraint we get:

$$0 = n\theta c^{n\theta-1} c'(a) - n\theta a^{n\theta-1};$$

from which $c'(a) = \left(\frac{a}{c}\right)^{n\theta-1}$.
We differentiate the length of the interval as a function of a:

$$\frac{\partial l}{\partial a} = -\frac{1}{a^2} + \frac{c'(a)}{c^2} = \frac{a^{n\theta+1} - c^{n\theta+1}}{a^2 c^{n\theta+1}} < 0.$$

Then the minimum length is for $c = 1$, from which:

$$1 - \alpha = 1 - a^{n\theta} \implies a = \alpha^{1/(n\theta)}.$$

We conclude that:

$$IC_{(1-\alpha)} = \left[X_{(n)}; \frac{X_{(n)}}{\alpha^{1/(n\theta)}} \right].$$

12.8

(a) The density of X:

$$f(x;\theta) = \frac{\theta^k \mathrm{e}^{kx} \mathrm{e}^{-\theta \mathrm{e}^x}}{\Gamma(k)} \mathbb{I}_{\mathbb{R}}(x)$$

belongs to the exponential family. Then: $W(\boldsymbol{X}) = \sum_{i=1}^{n} e^{X_i}$ is a sufficient statistic. Moreover:

$$w(\theta) = -\theta : \mathbb{R}^+ \to \mathbb{R}^-$$

and $\mathbb{R}^-$ contains an open set of $\mathbb{R}$. Therefore, $W(\boldsymbol{X})$ is a sufficient, complete and minimal statistic for θ.

(b) $Y = \mathrm{e}^X$, $X = \log Y$.

$$f_Y(y) = \frac{\theta^k \mathrm{e}^{k \log y} \mathrm{e}^{-\theta \log y}}{\Gamma(k)} \frac{1}{y} = \frac{\theta^k y^{k-1} e^{-\theta y}}{\Gamma(k)} \sim \Gamma(k, \theta).$$

(c) $W \sim \Gamma(nk, \theta)$.

(d) Let $\theta_2 > \theta_1$:

$$\frac{\theta_2^{nk} y^{nk-1} \mathrm{e}^{-\theta_2 y}}{\theta_1^{nk} y^{nk-1} \mathrm{e}^{-\theta_1 y}} = \left(\frac{\theta_2}{\theta_1} \right)^{nk} \mathrm{e}^{-(\theta_2 - \theta_1) y};$$

which is decreasing in y. Therefore $-\sum \mathrm{e}^{X_i}$ has an increasing MLR in y. Then:

$$R = \left\{ \sum \mathrm{e}^{x_i} < t_0 \right\} \qquad \text{with} \qquad t_0 = \gamma_\alpha(nk, \theta_0).$$

(e) We know that:

$$\theta W \sim \Gamma(nk, 1).$$

$$2\theta W \sim \Gamma\left(\frac{2nk}{2}, \frac{1}{2} \right) \overset{d}{=} \chi^2(2nk).$$

Therefore:

$$IC_{(1-\alpha)}(\theta) = \left[\frac{\chi^2_{\frac{\alpha}{2}}(2nk)}{2W}; \frac{\chi^2_{1-\frac{\alpha}{2}}(2nk)}{2W} \right].$$

12.9

(a)

$$\begin{aligned}
\mathbb{E}[X] &= 2\theta \int_0^{+\infty} x^2 \exp\{-\theta x^2\}\,\mathrm{d}x = \\
&= \theta \int_{-\infty}^{+\infty} x^2 \exp\left\{-\frac{1}{2}2\theta x^2\right\}\mathrm{d}x = \\
&= \theta \frac{\sqrt{2\pi}}{\sqrt{2\theta}} \frac{\sqrt{2\theta}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \exp\left\{-\frac{1}{2}2\theta x^2\right\}\mathrm{d}x = \qquad \left[\text{law } N\left(0, \frac{1}{2\theta}\right)\right] \\
&= \theta \sqrt{\frac{\pi}{\theta}} \frac{1}{2\theta} = \frac{1}{2}\frac{\sqrt{\pi}}{\sqrt{\theta}}.
\end{aligned}$$

(b)

$$\overline{X}_n = \frac{\sqrt{\pi}}{2\sqrt{\theta}} \iff \hat{\theta}_{MOM} = \frac{1}{4}\frac{\pi}{\overline{X}_n^2}.$$

(c) Exploiting the properties of the exponential family, we have that: $T(\boldsymbol{X}) = \sum X_i^2$ is a sufficient, minimal and complete statistic for θ.

(d)

$$f_{X_i^2}(y) = 2\theta\sqrt{y}\exp\{-\theta y\}\frac{1}{2\sqrt{y}} = \theta\exp\{-\theta y\};$$

meaning that $X_i^2 \sim \mathcal{E}(\theta)$ and therefore $T \sim \Gamma(n, \theta)$.

(e)

$$L(\theta; \boldsymbol{x}) = 2^n\theta^n \prod x_i \exp\left\{-\theta \sum x_i^2\right\}.$$

$$l(\theta; \boldsymbol{x}) \propto n\log\theta - \theta\sum x_i^2.$$

$$\frac{\partial l(\theta; \boldsymbol{x})}{\partial\theta} = \frac{n}{\theta} - \sum x_i^2.$$

Then:

$$\hat{\theta}_{MLE} = \frac{n}{\sum X_i^2}.$$

(f) The SLLN guarantees us that:

$$\hat{\theta}_{MOM} \stackrel{q.c.}{\to} \theta \qquad \text{and} \qquad \hat{\theta}_{MLE} \stackrel{q.c.}{\to} \theta.$$

Therefore, both estimators are consistent.

(g)

$$I_n(\theta) = nI_1(\theta) = n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\right)^2\right] = n\text{Var}(X^2) = \frac{n}{\theta^2}.$$

Therefore:

$$\sqrt{n}\left(\hat{\theta}_{MLE} - \theta\right) \stackrel{\mathcal{L}}{\to} N(0, \theta^2).$$

(h) $\text{Var}(X_i) = \frac{0.21}{\theta}$, therefore:

$$\sqrt{n}\left(\overline{X}_n - \frac{\sqrt{\pi}}{2\sqrt{\theta}}\right) \stackrel{\mathcal{L}}{\to} N\left(0, \frac{0.21}{\theta}\right).$$

Consider the Delta Method 1.17 with:

$$g(t) = \frac{\pi}{4}\frac{1}{t^2}.$$

$$g'(t) = -2\frac{\pi}{4}\frac{1}{t^3}.$$

$$g\left(\frac{\sqrt{\pi}}{2\sqrt{\theta}}\right) = \theta.$$

So it holds:

$$\sqrt{n}\left(\hat{\theta}_{MOM} - \theta\right) \stackrel{\mathcal{L}}{\to} N\left(0, \frac{0.21}{\theta} g'\left(\sqrt{\frac{\pi}{\theta}}\frac{1}{2}\right)^2\right);$$

where:

$$g'\left(\frac{\sqrt{\pi}}{2\sqrt{\theta}}\right)^2 = \left(\frac{4}{\sqrt{\pi}}\theta\sqrt{\theta}\right)^2 = \frac{16}{\pi}\theta^3.$$

Therefore, we conclude that:

$$\sqrt{n}\left(\hat{\theta}_{MOM}-\theta\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{0.21}{\pi}16\theta^2\right).$$

(i)

$$ARE(\hat{\theta}_{ML}; \hat{\theta}_{MOM}) = \frac{0.21 \cdot 16}{\pi} = 1.07$$

Therefore, I conclude that $\hat{\theta}_{MLE}$ is better.

12.10

(a) We impose that the integral of the density is equal to 1:

$$c\int_{\theta}^{\theta+1}(1-(x-\theta)^2)\,\mathrm{d}x = c\int_{\theta}^{\theta+1}\mathrm{d}x - c\int_{\theta}^{\theta+1}(x-\theta)^2\mathrm{d}x =$$

$$c - c\frac{(x-\theta)^3}{3}\bigg|_{\theta}^{\theta+1} = c\frac{2}{3} = 1.$$

We therefore conclude that $c = 3/2$.

(b)

$$L(\theta; x) = \frac{3}{2}\left(1-(x-\theta)^2\right)\mathbb{I}_{(\theta,\theta+1)}(x) = \frac{3}{2}\left(1-(x-\theta)^2\right)\mathbb{I}_{(x-1,x)}(\theta).$$

which is an increasing function in θ, therefore:

$$\hat{\theta}_{MLE} = X.$$

(c) Let $Q = 1-(X-\theta)^2$. We observe that:

$$Q = 1-(X-\theta)^2 \iff (1-Q) = (X-\theta)^2 \iff X = \theta+\sqrt{1-Q}.$$

The density of Q is:

$$f_Q(q) = \frac{3}{2}\left(1-\left(\theta+\sqrt{1-q}-\theta\right)^2\right)\frac{1}{2}\frac{1}{\sqrt{1-q}}\mathbb{I}_{(0,1)}(q)$$
$$= \frac{3}{4}\frac{q}{\sqrt{1-q}}\mathbb{I}_{(0,1)}(q).$$

Therefore, Q is a pivotal quantity for θ.

(d)

$$
\begin{aligned}
IC_{(1-\alpha)}(\theta) &= [a \le Q \le b] = \\
&= \left[a \le 1 - (X-\theta)^2 \ \le b\right] = \\
&= \left[1 - b \le (X-\theta)^2 \le 1 - a\right] = \\
&= \left[\sqrt{1-b} \le X - \theta \le \sqrt{1-a}\right] = \\
&= \left[X - \sqrt{1-a} \le \theta \le X - \sqrt{1-b}\right].
\end{aligned}
$$

(e) To determine the confidence level of the interval constructed in point (d), I need to calculate:

$$\mathbb{P}\{0.5 \le Q \le 0.9\} = F_Q(0.9) - F_Q(0.5).$$

Specifically:

$$
\begin{aligned}
F_Q(t) &= \mathbb{P}\left\{(1-(X-\theta))^2 \le t\right\} = \\
&= \mathbb{P}\left\{(X-\theta)^2 \ge 1 - t\right\} = \\
&= \mathbb{P}\left\{X \ge \theta + \sqrt{1-t}\right\} = \\
&= 1 - \mathbb{P}\left\{X < \theta + \sqrt{1-t}\right\} = \\
&= 1 - \frac{3}{2}\int_{\theta}^{\theta+\sqrt{1-t}} (1-(x-\theta)^2)\,\mathrm{d}x = \\
&= 1 - \frac{3}{2}\sqrt{1-t} + \frac{3}{2}\frac{(x-\theta)^3}{3}\bigg|_{\theta}^{\theta+\sqrt{1-t}} = \\
&= 1 - \frac{3}{2}\sqrt{1-t} + \frac{1}{2}(1-t)^{3/2}.
\end{aligned}
$$

So:

$$F_Q(0.9) - F_Q(0.5) = \frac{3}{2}(\sqrt{0.5} - \sqrt{0.1}) + \frac{1}{2}\left(0.1^{3/2} - 0.5^{3/2}\right) = 0.425.$$

(f) Consider the test:

$$H_0 : \theta = \theta_0 \qquad \text{vs} \qquad H_1 : \theta \ne \theta_0.$$

The rejection region is:

$$R = \left\{X \le \theta_0 + \sqrt{0.1}\right\} \cup \left\{X \ge \theta_0 + \sqrt{0.5}\right\}.$$

12.11

(a)

$$f(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} x_i \theta^{-2n} \exp\left\{-\frac{1}{\theta}\sum x_i\right\} \mathbb{I}_{[0,+\infty)}(x_i).$$

The density of $\boldsymbol{X}$ belongs to the exponential family, so $\sum X_i$ is a sufficient statistic.
Moreover, since $w(\theta) = -\frac{1}{\theta} : (0, +\infty) \to (-\infty, 0)$ and $(-\infty, 0)$ contains an open set of $\mathbb{R}$, $\sum X_i$ is a sufficient and complete statistic.

(b)

$$l(\theta; \boldsymbol{x}) \propto -2n \log\theta - \frac{1}{\theta}\sum x_i.$$

$$\frac{\partial l}{\partial \theta} = -\frac{2n}{\theta} + \frac{\sum x_i}{\theta^2} \ge 0 \iff \frac{\sum x_i}{\theta^2} \ge \frac{2n}{\theta} \iff \theta \le \frac{\overline{X}_n}{2}.$$

So:

$$\hat{\theta}_n = \frac{\overline{X}_n}{2}.$$

(c)

$$\mathrm{e}[X] = 2\theta \quad \Rightarrow \quad \bar{\theta}_n = \frac{\overline{X}_n}{2}.$$

(d)

$$X_i \sim \Gamma\left(2, \frac{1}{\theta}\right) \quad \Rightarrow \quad \sum X_i \sim \Gamma\left(2n, \frac{1}{\theta}\right) \quad \Rightarrow \quad \frac{\overline{X}_n}{2} \sim \Gamma\left(2n, \frac{2n}{\theta}\right).$$

(e)

$$\mathbb{E}[\hat{\theta}_n] = \theta.$$

$\hat{\theta}_n$ is an unbiased estimator.

(f) $\hat{\theta}_n$ is UMVUE because it is unbiased and a function of a sufficient and complete statistic for θ.

(g) $Q = \frac{\hat{\theta}_n}{\theta} \sim \Gamma(2n, 2n)$ is a pivotal quantity. We therefore look for a and b such that:

$$\mathbb{P}\{a \leq Q \leq b\} = 0.99$$

We therefore propose as a 0.99 level CI:

$$IC(0.99) = \left(\frac{\hat{\theta}_n}{b}, \frac{\hat{\theta}_n}{a}\right);$$

with $b = \gamma_{0.995}(2n, 2n)$ and $a = \gamma_{0.005}(2n, 2n)$.

Appendix A
Probability Distributions

A.1 Continuous Distributions

Normal Distribution

$$X \sim N(\mu, \sigma^2), \quad \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}^+.$$

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

$$\mathbb{E}[X] = \mu.$$

$$Var(X) = \sigma^2.$$

Uniform Distribution

$$X \sim U_{[a,b]}, \quad a, b \in \mathbb{R}, a < b.$$

$$f_X(x; a, b) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x).$$

$$\mathbb{E}[X] = \frac{b+a}{2}.$$

$$Var(X) = \frac{(b-a)^2}{12}.$$

Exponential Distribution

$$X \sim \mathcal{E}(\lambda), \quad \lambda \in \mathbb{R}^+.$$

$$f_X(x; \lambda) = \lambda \exp\{-\lambda x\} \, \mathbb{I}_{[0,+\infty)}(x).$$

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8

$$\mathbb{E}[X] = \frac{1}{\lambda}.$$
$$Var(X) = \frac{1}{\lambda^2}.$$

Gamma Distribution

$$X \sim \Gamma(\alpha, \lambda), \quad \alpha, \lambda \in \mathbb{R}^+.$$
$$f_X(x; \alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1} \exp\{-\lambda x\}}{\Gamma(\alpha)} \mathbb{I}_{[0,+\infty)}(x).$$
$$\mathbb{E}[X] = \frac{\alpha}{\lambda}.$$
$$Var(X) = \frac{\alpha}{\lambda^2}.$$

Properties of the Gamma Distribution
If $X_1, \ldots, X_n$ are i.i.d. such that $X_i \sim \Gamma(\alpha, \lambda)$, then:

$$\sum_{i=1}^{n} X_i \sim \Gamma(n\alpha, \lambda).$$
$$\frac{X_i}{n} \sim \Gamma(\alpha, n\lambda).$$

χ^2 Distribution

$$X \sim \chi^2(\lambda), \quad \lambda \in \mathbb{N}\backslash\{0\}$$
$$f_X(x; \lambda) = \frac{1}{2^{\lambda/2}\Gamma(\lambda/2)} x^{\lambda/2-1} \exp\{-x/2\} \mathbb{I}_{[0,+\infty)}(x).$$
$$\mathbb{E}[X] = \lambda.$$
$$Var(X) = 2\lambda.$$

Properties of the χ^2 Distribution
If $X_1 \sim \chi^2(\lambda_1)$, $X_2 \sim \chi^2(\lambda_2)$, …, $X_n \sim \chi^2(\lambda_n)$, then:

$$\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} \lambda_i\right).$$

Relations Between Distributions

- *Normal* and χ^2: If $X_1, \ldots, X_n$ are i.i.d. such that $X_i \sim N(0, 1)$, then:

$$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n).$$

- *Gamma* and χ^2:

$$\Gamma\left(\frac{k}{2}, \frac{1}{2}\right) \overset{\mathcal{L}}{\equiv} \chi^2(k).$$

- *Exponential* and *Gamma*:

$$\mathcal{E}(\lambda) \overset{\mathcal{L}}{\equiv} \Gamma(1, \lambda).$$

A.2 Discrete Distributions

Bernoulli Distribution

$$X \sim Be(p), \quad p \in [0, 1].$$
$$f_X(x; p) = p^x (1-p)^{1-x}\, \mathbb{I}_{\{0,1\}}(x).$$
$$\mathbb{E}[X] = p.$$
$$Var(X) = p(1-p).$$

Binomial Distribution

$$X \sim Bin(n, p), \quad p \in [0, 1]\ n \in \mathbb{N}.$$
$$f_X(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}\, \mathbb{I}_{\{0,\ldots,n\}}(x).$$
$$\mathbb{E}[X] = np.$$
$$Var(X) = np(1-p).$$

Uniform Distribution

$$X \sim U_{\{a,b\}}, \quad a, b \in \mathbb{R},\ a < b.$$
$$f_X(x; a, b) = \frac{1}{n}\, \mathbb{I}_{\{a,b\}}(x).$$

$$\mathbb{E}[X] = \frac{b+a}{2}.$$

$$Var(X) = \frac{n^2-1}{12}.$$

n is the number of natural numbers between a and b.

Poisson Distribution

$$X \sim \mathcal{P}(\lambda), \quad \lambda \in \mathbb{R}^+.$$

$$f_X(x;\lambda) = \frac{\lambda^x \cdot \exp\{-\lambda\}}{x!}\, \mathbb{I}_{\mathbb{N}}(x).$$

$$\mathbb{E}[X] = \lambda.$$

$$Var(X) = \lambda.$$

Properties of the Poisson Distribution

If $X_1, \ldots, X_n$ are independent random variables such that $X_i \sim \mathcal{P}(\lambda_i)$, then:

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right).$$

References

1. Agresti, A., Kateri, M.: Categorical data analysis. In: Lovric M. (eds.) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg (2011)
2. Box G.E.P., Cox D.R.: An analysis of transformations. J. R. Stat. Soc. Ser. B (Methodol.) **26**(2), 211–243 (1964)
3. Casella, G., Berger, R.L.: Statistical Inference, vol. 2. Duxbury Pacific Grove (2002)
4. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression. Wiley, New York (2000)
5. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis, vol. 3. Pearson New International, London (2015)
6. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2014)

F. Gasperoni et al., *Exercise Book of Statistical Inference*, La Matematica per il 3+2 173, https://doi.org/10.1007/978-3-031-86670-8