Models of Ordered Random Variables and Exponential Families

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Mathematiker

Stefan Bedbur

aus Erkelenz

Berichter: Universitätsprofessor Dr. Udo Kamps
Professor Dr. Eric Beutner

Tag der mündlichen Prüfung: 19. Januar 2011

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

Acknowledgements

I am indebted to many people who supported me in different ways while I was working on this doctoral thesis.
In the first place, I am very grateful to my supervisor Professor Udo Kamps for the opportunity to work in the very interesting field of models of ordered random variables and for his continuous support and interest in my studies.
Moreover, I owe particular thanks to Professor Eric Beutner for always being interested in my work, for his inspiring comments and suggestions, and for his acceptance to be co-referee.
I am also very thankful to Professor Marco Burkschat for a lot of valuable advice.
Sincere thanks are given to my colleagues of the Institute of Statistics for a really good joint time in Aachen. Most of all, I would like to thank my friend and former fellow student, Tim Fischer, for his helpfulness and encouragement, for many interesting and fruitful discussions, and for a great atmosphere at work.
With all my heart, I thank my family, in particular, my parents, Margret and Manfred Bedbur, and my sister, Ruth Bedbur, for all their support and sympathy. Explicitly, I wish to thank my brother, Thomas Bedbur, for his assistance during the last time of my studies. Dear thanks are also given to my grandparents, Margarethe and Johannes Klein, for all their good wishes and thoughts.
Finally, I thank my love Judith Zimmermann who plays the most important role in my life. Judith, my unlimited gratefulness for all your love, understanding and support can neither be expressed within these acknowledgements nor with any other written or spoken words. This work is dedicated to you and would not have been possible without you.

Contents

Chapter 1

Introduction

1.1 Models of Ordered Random Variables

Models of ordered random variables play an important role in many statistical applications. In insurance mathematics, different models of *record values* are helpful to describe largest claims to an insurance company, whereas, in reliability theory, the models of *common order statistics* and *sequential order statistics* are of particular interest in modeling so called k-out-of-n systems. As a particular technical structure, a k-out-of-n system consists of n components of the same kind and fails if $n-k+1$ or more components of the system fail. If, additionally, the failure of a component may influence the lifetime of the remaining ones, the system is called a *sequential k-out-of-n system* and, otherwise, the denotation *common k-out-of-n system* is used. As an example for a sequential 3-out-of-4 system, imagine an air plane with four turbines where the failure of more than one turbine leads to the failure of the system. Upon failure of the first turbine, increased stress is put on the remaining three turbines. (Common) order statistics and sequential order statistics, respectively, model the corresponding systems, or, more precisely, the lifetime of the systems and their components, where the latter have been introduced by Kamps (1995*a*,*b*) in terms of a triangular scheme of independent random variables. Denoting by F^{-1} the quantile function of the distribution function F, i.e., $F^{-1}(y) = \inf\{x : F(x) \geq y\}$, $y \in (0,1)$, and $F^{-1}(0) = \lim_{y \searrow 0} F^{-1}(y)$, $F^{-1}(1) = \lim_{y \nearrow 1} F^{-1}(y)$, this definition is as follows (cf. Cramer & Kamps (2003)).

Definition 1.1.1 (Sequential order statistics)
Let $(Y_j^{(i)})_{1 \leq i \leq n, 1 \leq j \leq n-i+1}$ *be independent random variables with* $Y_j^{(i)} \sim F_i$, $1 \leq i \leq n$, $1 \leq j \leq n-i+1$, *where* $F_1, ..., F_n$ *are distribution functions with* $F_1^{-1}(1) \leq ... \leq F_n^{-1}(1)$.
Let

$$X_j^{(1)} = Y_j^{(1)},\ 1 \leq j \leq n, \quad \textit{and} \quad X_*^{(1)} = \min\{X_1^{(1)}, ..., X_n^{(1)}\},$$

and for $2 \leq i \leq n$

$$X_j^{(i)} = F_i^{-1}\{F_i(Y_j^{(i)})[1 - F_i(X_*^{(i-1)})] + F_i(X_*^{(i-1)})\},\ 1 \leq j \leq n-i+1,$$

and

$$X_*^{(i)} = \min\{X_j^{(i)},\ 1 \leq j \leq n-i+1\}.$$

Then the random variables $X_*^{(1)}, ..., X_*^{(n)}$ *are called sequential order statistics (SOSs) based on* $F_1, ..., F_n$.

For more details as well as the extended model along with distribution theory and a variety of properties, we refer to Cramer & Kamps (2001*b*, 2003) and Kamps (1995*a*,*b*). In Cramer & Kamps (2003), an alternative definition of SOSs is given, which coincides with Def. 1.1.1 provided that the distribution functions $F_1, \dots, F_n$ are continuous.

Definition 1.1.2 (Sequential order statistics)
Let $F_1, ..., F_n$ *be distribution functions with* $F_1^{-1}(1) \leq ... \leq F_n^{-1}(1)$, *and let* $V_1, ..., V_n$ *be independent random variables with* $V_i \sim Beta(n-i+1, 1)$, $1 \leq i \leq n$.
Then the random variables

$$X_*^{(i)} = F_i^{-1}(X^{(i)}) \quad \textit{with} \quad X^{(i)} = 1 - V_i \bar{F}_i(X_*^{(i-1)}), \quad 1 \leq i \leq n, \quad X_*^{(0)} = -\infty,$$

where $\bar{F}_i = 1 - F_i$, $1 \leq i \leq n$, *are called sequential order statistics (SOSs) based on* $F_1, ..., F_n$.

If $F_1, \dots, F_n$ are absolutely continuous distribution functions with corresponding density functions $f_1, \dots, f_n$, the joint density of the first r SOSs is given by

$$f^{X_*^{(1)},\dots,X_*^{(r)}}(x_1, \dots, x_r) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} \left[\left(\frac{1 - F_i(x_i)}{1 - F_i(x_{i-1})} \right)^{n-i} \frac{f_i(x_i)}{1 - F_i(x_{i-1})} \right] \tag{1.1.1}$$

on the cone $x_1 < \dots < x_r$, where $1 \leq r \leq n$ and $x_0 = -\infty$ (e.g., in Kamps (1995*b*), p. 29).

Throughout this doctoral thesis, we will restrict ourselves to a particular choice of the distribution functions $F_1, \dots, F_n$ and consider *SOSs with conditional proportional hazard rates* which result from the general sequential model by setting

$$F_j = 1 - (1 - F)^{\alpha_j}, \quad 1 \leq j \leq n,$$

where F is an absolutely continuous baseline distribution function with corresponding density function f and $\alpha_1, ..., \alpha_n$ are positive parameters. By doing so, the failure rate of F_j is given by $\alpha_j f/(1-F)$ and, thus, proportional to the failure rate of the baseline distribution. In that sense, the components of a *sequential* $(n-r+1)$*-out-of-*n *system with conditional proportional hazard rates* start operating at hazard rate $\alpha_1 f/(1-F)$ and, upon failure of the j^{th} component, $1 \leq j \leq r-1$, the hazard rate of the remaining components is supposed to change from $\alpha_j f/(1-F)$ to $\alpha_{j+1} f/(1-F)$. The r^{th} SOS describes the lifetime of the system.
In the above situation, the joint density of the first r SOSs $X_*^{(1)}, \dots, X_*^{(r)}$ with conditional proportional hazard rates is given by

$$\begin{aligned} f_{\boldsymbol{\alpha}}^{X_*^{(1)},\dots,X_*^{(r)}}(x_1, \dots, x_r) &= \tfrac{n!}{(n-r)!} \left(\prod_{j=1}^{r} \alpha_j \right) \left(\prod_{j=1}^{r-1} (1 - F(x_j))^{m_j} f(x_j) \right) \\ &\quad \times (1 - F(x_r))^{\alpha_r(n-r+1)-1} f(x_r) \end{aligned} \tag{1.1.2}$$

on the cone $F^{-1}(0) < x_1 < \cdots < x_r < F^{-1}(1)$, with $1 \leq r \leq n$, and $m_j = (n-j+1)\,\alpha_j - (n-j)\,\alpha_{j+1} - 1,\quad 1 \leq j \leq r-1$ (cf. Kamps (1995a,b) and Cramer & Kamps (2001*b*)). The index $\boldsymbol{\alpha}$ denotes the vector $(\alpha_1, \ldots, \alpha_r)'$ of model parameters. E.g., the particular case $\alpha_1 = \cdots = \alpha_r = 1$ corresponds to common order statistics (OSs) based on the distribution function F. For another approach to the stochastic modeling of reliability systems with conditional proportional hazard rates by using stochastic intensities, we refer to Hollander & Peña (1995) and, for related results concerning statistical inference on the respective model parameters, to Kim & Kvam (2004).

In the distribution theoretical sense, the model of SOSs with conditional proportional hazard rates coincides with the model of *generalized order statistics* that has been introduced by Kamps (1995*a*,*b*) to embed different models of ordered random variables within an enlarged parametric model in terms of the joint density of the respective random variables.

Definition 1.1.3 (Generalized order statistics)
Let $n \in \mathbb{N}$, $\gamma_1, \ldots, \gamma_n > 0$ be positive model parameters, and let F be an absolutely continuous distribution function with corresponding density function f. Then the ordered random variables $X_*^{(1)}, \ldots, X_*^{(n)}$ are called generalized order statistics (gOSs), if their joint density function is given by

$$f_{\boldsymbol{\alpha}}^{X_*^{(1)},\ldots,X_*^{(n)}}(x_1, \ldots, x_n) = \left(\prod_{j=1}^{n} \gamma_j\right) \left(\prod_{j=1}^{n-1} (1-F(x_j))^{\gamma_j - \gamma_{j+1} - 1} f(x_j)\right) \times (1-F(x_n))^{\gamma_n - 1} f(x_n) \tag{1.1.3}$$

on the cone $F^{-1}(0) < x_1 < \cdots < x_n < F^{-1}(1)$.

At different choices of the model parameter $\gamma_1, \ldots, \gamma_n$, well-known models of ordered random variables are included in the model of gOSs in the distribution theoretical sense. E.g., by setting $\gamma_j = n-j+1$, $1 \leq j \leq n$, we obtain the joint density of common OSs based on F, which are extensively investigated in David & Nagaraja (2003), or, by defining $\gamma_j = 1$, $1 \leq j \leq n$, (1.1.3) is the joint density of the first n record values based on the sequence $(X_i)_{i \in \mathbb{N}}$ of independent and identically distributed (iid) random variables with distribution function F. In case of other choices of the γ's, the models of k^{th} record values, progressive typ-II censoring and other models are included as well (cf. Kamps (1995a,b)).
With $r = n$ in (1.1.2), the joint density of all SOSs with conditional proportional hazard rates results from the model of gOSs by setting in (1.1.3) $\gamma_j = (n-j+1)\alpha_j$, $1 \leq j \leq n$, and vice versa, by setting $\alpha_j = \gamma_j/(n-j+1)$, $1 \leq j \leq n$. Hence, in the distribution theoretical sense, both models are the same.

Meanwhile, there are many articles dealing with theory and application of SOSs and gOSs (for structural results see, e.g., Belzunce et al. (2008), Bieniek (2008), Burkschat (2009) and Cramer (2006)). Inferential issues have also been addressed. In Cramer & Kamps (1996, 2001*b*), maximum likelihood estimators are presented along with useful properties, and several short-cut tests are proposed. More recent results may be found in Balakrishnan et al. (2008), Beutner & Kamps (2009) and Burkschat (2010). For a nonparametric approach, see Beutner (2008, 2010).

In this doctoral thesis, we point out that the joint density of the first r SOSs with conditional proportional hazard rates, or, equivalently, of the first r gOSs, forms a *multivariate exponential family* in the model parameters. This structural finding opens the wide and extensively examined field of exponential families to models of ordered random variables. As a consequence, simplified proofs of former results in literature, in particular of those related to inferential issues, can be given, and new useful properties can be shown as well.
Throughout the thesis, aiming at statistical inference for model parameters, we focus on SOSs to introduce the structure under consideration and to motivate needs for statistical methods. However, the derived results are also true for the model of gOSs, and, hence, may be applied when dealing with, e.g., Pfeifer's record model.

1.2 Summary

The outline of this work is as follows.

In Chapter 2, the concept 'exponential family' and related notions are formally defined. Based on an underlying exponential family structure, well-known theorems and properties in literature are presented and rearranged for application to models of ordered random variables, in particular, to the model of SOSs with conditional proportional hazard rates. All these results are stated mathematically sound which takes a few pages to explain.

In Chapter 3, we introduce the model of SOSs with conditional proportional hazard rates and its motivation. Throughout the whole chapter, the underlying baseline distribution is assumed to be known. By this assumption, after a little algebra, the exponential family structure of the joint density of SOSs with conditional proportional hazard rates is obvious, and, as a consequence, the results of Chapter 2 are applicable. In Section 3.1, basic properties of this exponential family are shown. The distributions of statistics which are of great importance for inferential issues are derived. Based on the structural insight, minimal sufficiency and completeness of these statistics are readily obtained, where, in general, both properties are hard to see. The computation of the Fisher information matrix, which is also closely connected to inferential statistics, is much simplified as well. Moreover, it turns out that the exponential family structure is preserved by considering the joint density of iid vectors of SOSs with conditional proportional hazard rates. For the most part, this section can be regarded as a preliminary work which enables us to inference on the model parameters in Sections 3.2 and 3.3.
In Section 3.2, the model parameters are estimated based on a sample of s iid vectors of SOSs with conditional proportional hazard rates. Maximum likelihood estimators for single model parameters and vectors of them are easily obtained. In the first case, uniformly minimum variance unbiased estimators are derived as well. Subsequently, useful properties of these estimators and the respective sequences of estimators, when s tends to infinity, are shown, e.g., consistency and (asymptotic) efficiency. Finally, maximum likelihood estimation is considered by assuming that the model parameters are simply ordered.

In Section 3.3, statistical tests on the model parameters are discussed based on a sample of s iid vectors of SOSs with conditional proportional hazard rates. Based on the underlying exponential family structure, uniformly most powerful unbiased tests for a variety of hypotheses concerning single model parameters are established. These tests are useful, e.g., in the context of model checking where the prior information that all other model parameters equal some pre-fixed value is given. In situations where such an information is not available and, hence, more than one parameter is on test, we propose different multivariate tests for model checking, i.e. the likelihood ratio test, Wald's (modified) test and Rao's score test. For different multivariate test problems with either a simple or composite null hypothesis, which can be interpreted as model tests, we compute the corresponding test statistics, where we, once again, benefit from the underlying exponential family structure. In each case, the asymptotic distributions of the test statistics under the null hypothesis are derived. Moreover, we compare the multivariate tests in terms of well-known asymptotic optimality properties, e.g., by considering the asymptotic relative efficiency of the tests in the sense of Bahadur. At the end of the chapter, for different test problems related to simply ordered model parameters, the asymptotic distribution of the likelihood ratio test statistic under the null hypothesis is discussed.

In Chapter 4, we generalize and extend the findings from Chapter 3. Section 4.1 is concerned with the case of s independent but not necessarily identically distributed (inid) vectors of SOSs with conditional proportional hazard rates based on a known underlying distribution. The section is similarly structured as Chapter 3 and contains respective generalizations of many statements that have already been shown for the iid case. At this, results and proofs are presented more briefly than in Chapter 3.
In Subsection 4.2, we consider the case where SOSs with conditional proportional hazard rates are based on a partially unknown baseline distribution, where the uncertainty of the underlying distribution is captured within an unknown rate parameter. As we will see, this situation is covered by the results of Chapter 3.

In Chapter 5, the theoretical results of Section 3.3 are illustrated by means of a simulation study, where, in Subsection 5.1, univariate statistical tests on single model parameters and, in Subsection 5.2, multivariate model tests are examined and compared in terms of their power functions.

Finally, in Chapter 6, the impact and the main contributions of this work are discussed.

Chapter 2

Exponential Families

In this chapter, an introduction into the extensive and almost exhaustively examined field of exponential families and their properties is presented (see, e.g., Barndorff-Nielsen (1978), Lehmann & Casella (1998), Lehmann & Romano (2005), Shao (2003) and Witting (1985)). At this, our main objectives are

- to define and explain related notations and concepts, and
- to prepare the tool 'exponential family' for application to models of ordered random variables, in particular, to SOSs with conditional proportional hazard rates.

Thus, in the following, well-known results from the literature are rearranged and extended to result in a concise account of exponential families and their properties. For this, the chapter is divided into three sections.
In Section 2.1, the term exponential family and furhter related terminology are formally defined. Throughout Chapter 2, the notation introduced in this section, which is mainly oriented to the by Witting (1985), will not change. In what follows, the reader gets an insight into the nature of exponential families and detects that an underlying exponential family structure brings along many pleasant properties considered either from a stochastical or a statistical point of view. As two examples, moments of certain random variables can easily be obtained, and inferential issues are considerably simplified using the fact that complete sufficient statistics are near at hand.
Section 2.2 and Section 2.3 are concerned with inferential statistical statements that result from an underlying exponential family structure. At this, in Section 2.2, point estimators of (unknown) parameters and their properties are established, whereas Section 2.3 deals with statistical tests on the respective parameters. Having said that the exponential family structure simplifies many problems and procedures related to statistical inference, e.g. the derivation of maximum likelihood estimators, it also enables us to address *optimality* properties to estimators and statistical tests as well.

2.1 Fundamentals

2.1.1 Definitions and Representations

We begin with a formal definition of an exponential family.

Definition 2.1.1 (Exponential family)
Let $(\mathfrak{X}, \mathfrak{B})$ *be a measurable space,* $\Theta \neq \emptyset$ *a set of parameters and* $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ *a family of probability measures on* $(\mathfrak{X}, \mathfrak{B})$*. If there exist a* σ*-finite measure* μ *on* $(\mathfrak{X}, \mathfrak{B})$ *that dominates* $\mathfrak{P}$ *and, moreover, an integer* $k \in \mathbb{N}$*, real-valued functions* $C, \zeta_1, ..., \zeta_k$ *on* Θ *and real-valued* $\mathfrak{B} - \mathbb{B}^1$*-measurable functions* $h, T_1, ..., T_k$ *on* $(\mathfrak{X}, \mathfrak{B})$*, where* $h \geq 0$*, in such a way that a* μ*-density of* P_ϑ*,* $\vartheta \in \Theta$*, is given by*

$$\frac{dP_\vartheta}{d\mu}(x) = C(\vartheta) \exp\left\{\sum_{j=1}^{k} \zeta_j(\vartheta) T_j(x)\right\} h(x), \quad x \in \mathfrak{X}, \tag{2.1.1}$$

then $\mathfrak{P}$ *is called a* k*-parametrical (or* k*-parameter) exponential family in* $\zeta_1, ..., \zeta_k$ *and* $T_1, ..., T_k$*.*

At first glance, it is quite evident that the integer k and the functions $\zeta_1, ..., \zeta_k$ as well as $T_1, ..., T_k$ are not uniquely determined. We will revert to that point in Subsection 2.1.2. Obviously, the expression on the right-hand side of (2.1.1) is divided into three terms. $C(\vartheta)$ depends on the parameter ϑ but not on x, whereas the function h of x is independent from the value of ϑ. The expression in the exponent is the standard scalar product of a k-dimensional function $\boldsymbol{\zeta}$ of ϑ and a $\mathfrak{B} - \mathbb{B}^k$-measurable k-dimensional function $\boldsymbol{T}$ of x, and, thus, depends on both variables. We state the role of $C(\vartheta)$ explicitly in the following remark.

Remark 2.1.2
$C(\vartheta)$ *in the right-hand side of (2.1.1) is a normalizing constant and does not depend on* x*, i.e.*

$$0 < C(\vartheta) = \left[\int \exp\left\{\sum_{j=1}^{k} \zeta_j(\vartheta) T_j(x)\right\} h(x) d\mu(x)\right]^{-1} < \infty, \quad \vartheta \in \Theta. \tag{2.1.2}$$

With regard to Rem. 2.1.2, the mapping

$$\kappa : \quad \Theta \to \mathbb{R} : \quad \vartheta \mapsto -\ln(C(\vartheta)), \tag{2.1.3}$$

is well-defined. Hence, (2.1.1) can be rewritten as

$$\frac{dP_\vartheta}{d\mu}(x) = \exp\left\{\sum_{j=1}^{k} \zeta_j(\vartheta) T_j(x) - \kappa(\vartheta)\right\} h(x), \quad x \in \mathfrak{X}. \tag{2.1.4}$$

For a short notation, we set $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_k)'$ and $\boldsymbol{T} = (T_1, ..., T_k)'$, where $\boldsymbol{v}'$ denotes the transpose of the vector $\boldsymbol{v}$. Then, (2.1.1) and (2.1.4), respectively, can be written as

$$\frac{dP_\vartheta}{d\mu}(x) = C(\vartheta) \exp\{\boldsymbol{\zeta}(\vartheta)' \boldsymbol{T}(x)\} h(x), \quad x \in \mathfrak{X}, \tag{2.1.5}$$

and

$$\frac{dP_\vartheta}{d\mu}(x) = \exp\left\{\boldsymbol{\zeta}(\vartheta)'\boldsymbol{T}(x) - \kappa(\vartheta)\right\} h(x), \quad x \in \mathfrak{X}, \tag{2.1.6}$$

respectively.

Many well-known distribution families form exponential families. If μ denotes Lebesgue measure, e.g., the class $\mathcal{N}(\mu, \sigma^2)$ of normal distributions with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, the class $\Gamma(\beta, \alpha)$ of gamma distributions with shape parameter $\beta > 0$ and scale parameter $\alpha > 0$ and the family $Beta(\alpha, \beta)$ of beta distributions with shape parameters $\alpha, \beta > 0$ form exponential families in the respective parameters and statistics. Moreover, if μ is assumed to be the counting measure on $\mathbb{N}_0$, it is easily seen that, e.g., the family $po(\lambda)$ of poisson distributions with parameter $\lambda > 0$, the geometrical distributions $geo(p)$ with parameter $p \in (0,1)$ and, for fixed $n \in \mathbb{N}$, the binomial distributions $bin(n,p)$ with parameter $p \in (0,1)$ admit μ-densities according to the above exponential family structure.

Typical examples for distribution families that do not form exponential families are the uniform distributions $U(0,b)$ on $(0,b)$ with parameter $b > 0$ and the family of (right- or left-) truncated exponential distributions $Exp(\eta, \rho)$ with location parameter $\eta \in \mathbb{R}$ and scale parameter $\rho > 0$. The reason why these distribution families do not satisfy a representation of type (2.1.1) lies in the fact that the respective probability measures do not have the same support as it is always the case in exponential families. This insight is an immediate consequence of the following.

The function h is $\mathfrak{B} - \mathbb{B}^1$-measurable and nonnegative, and, moreover, does not depend on the parameter ϑ. Hence, another representation of the exponential family $\mathfrak{P}$ can be stated according to the hereafter lemma.

Lemma 2.1.3

Let $\mathfrak{P}$ be an exponential family in $\zeta_1, \ldots, \zeta_k$ and $T_1, \ldots, T_k$ according to Def. 2.1.1 and let $\nu = h\mu$ be the measure on $(\mathfrak{X}, \mathfrak{B})$ with μ-density h. Then, the following assertions hold true:

(i) *ν is σ-finite.*

(ii) *ν and P_ϑ are equivalent measures for every $\vartheta \in \Theta$, i.e. ν dominates P_ϑ and vice versa for every $\vartheta \in \Theta$.*

(iii) *A ν-density of P_ϑ, $\vartheta \in \Theta$, is given by*

$$\frac{dP_\vartheta}{d\nu}(x) = C(\vartheta) \exp\left\{\sum_{j=1}^{k} \zeta_j(\vartheta) T_j(x)\right\}, \; x \in \mathfrak{X}. \tag{2.1.7}$$

Proof. According to Def. 2.1.1, h is real-valued and, thus, (i) follows from Bauer (1990), Thm. 17.11., p. 118. For a proof of (ii) see Witting (1985), pp. 143/144. Moreover, μ dominates ν and, hence, application of (ii) and Cor. 1.134 in Witting (1985), p. 132, yield statement (iii). ■

Remark 2.1.4

La. 2.1.3 (ii) implies that P_ϑ and $P_{\tilde{\vartheta}}$ are equivalent measures for arbitrary $\vartheta, \tilde{\vartheta} \in \Theta$. In particular, all of the probability measures P_ϑ, $\vartheta \in \Theta$, have the same support.

In the following, if some property holds P_ϑ-almost sure (μ-almost everywhere), we write $[P_\vartheta]$ or P_ϑ-a.s. ($[\mu]$ or μ-a.e.). Additionally, for later use, we introduce the notation $[\mathfrak{P}] = [P_\vartheta]$ for some arbitrary $\vartheta \in \Theta$, which is well-defined by notice of Rem. 2.1.4.
Finally, we end Subsection 2.1.1 by mentioning that the structure of an exponential family is preserved by consideration of the family of distributions of $\boldsymbol{T}$.

Lemma 2.1.5
Let $\mathfrak{P}$ be an exponential family in $\zeta_1, \ldots, \zeta_k$ and $T_1, \ldots, T_k$ according to Def. 2.1.1 and La. 2.1.3, and let $\mathfrak{P}^{\boldsymbol{T}} = \{P_\vartheta^{\boldsymbol{T}} : \vartheta \in \Theta\}$ be the family of the corresponding distributions of $\boldsymbol{T}$ on $(\mathbb{R}^k, \mathbb{B}^k)$. Then, $\mathfrak{P}^{\boldsymbol{T}}$ forms a k-parametrical exponential family in $\zeta_1, ..., \zeta_k$ and the projections $p_1, \ldots, p_k$ ($p_j(\boldsymbol{t}) = t_j$, $\boldsymbol{t} = (t_1, \ldots, t_k)' \in \mathbb{R}^k$, $1 \leq j \leq k$), and a $\nu^{\boldsymbol{T}}$-density of $P_\vartheta^{\boldsymbol{T}}$, $\vartheta \in \Theta$, is given by

$$\frac{dP_\vartheta^{\boldsymbol{T}}}{d\nu^{\boldsymbol{T}}}(\boldsymbol{t}) = C(\vartheta) \exp\left\{\sum_{j=1}^{k} \zeta_j(\vartheta) t_j\right\}, \quad \boldsymbol{t} = (t_1, ..., t_k)' \in \mathbb{R}^k. \tag{2.1.8}$$

Proof. In Witting (1985), Thm. 1.160, p. 149. ■

Remark 2.1.6
$\nu^{\boldsymbol{T}}$ is a σ-finite measure on $(\mathbb{R}^k, \mathbb{B}^k)$ which can be seen as follows. For an arbitrary $\vartheta \in \Theta$, the function $l(\boldsymbol{t}) = C(\vartheta) \exp\left\{\sum_{j=1}^{k} \zeta_j(\vartheta) t_j\right\}$, $\boldsymbol{t} \in \mathbb{R}^k$, is $\nu^{\boldsymbol{T}}$-integrable and satisfies $0 < l < \infty$. Then, application of La. 17.6, p. 112, in Bauer (1990), yields the assertion.

2.1.2 Strict Parametrization and Minimal Representation

In view of further work and an easier handling of exponential families, a representation of the form (2.1.1) is desired where the integer k is minimum.

Definition 2.1.7 (Strict parametrization; full rank)
Let $\mathfrak{P}$ be an exponential family according to Def. 2.1.1.

(i) $\mathfrak{P}$ is called a strictly k-parametrical exponential family if the integer k is minimal in the following sense: If $\mathfrak{P}$ satisfies another representation

$$\frac{dP_\vartheta}{d\tilde{\mu}}(x) = \tilde{C}(\vartheta) \exp\left\{\sum_{j=1}^{\tilde{k}} \tilde{\zeta}_j(\vartheta) \tilde{T}_j(x)\right\} \tilde{h}(x), \quad x \in \mathfrak{X}, \quad [\tilde{\mu}] \tag{2.1.9}$$

with a σ-finite measure $\tilde{\mu}$ on $(\mathfrak{X}, \mathfrak{B})$, an integer $\tilde{k}$ and real-valued functions $\tilde{C}, \tilde{\zeta}_1, ..., \tilde{\zeta}_{\tilde{k}}$ on Θ and real-valued $\mathfrak{B}-\mathbb{B}^1$-measurable functions $\tilde{h}, \tilde{T}_1, ..., \tilde{T}_{\tilde{k}}$ on $(\mathfrak{X}, \mathfrak{B})$, where $\tilde{h} \geq 0$, then follows $\tilde{k} \geq k$.

(ii) If $\mathfrak{P}$ is strictly k-parametrical and the interior $int(\Theta)$ of Θ is not empty, $\mathfrak{P}$ is said to be of full rank.

In literature, it is often said that if $\mathfrak{P}$ is strictly k-parametrical with densities according to (2.1.1), the exponential family is given by its *minimal representation*. However, as mentioned before, this representation is not uniquely determined, but the minimal integer k is. For this reason, we use a terminology where this integer is not omitted.
In order to give a characterization of the above property of an exponential family, we introduce another terminology.

Definition 2.1.8 (Affine independence)

(i) *Let $M \neq \emptyset$ be an arbitrary set, $k \in \mathbb{N}$ and $\zeta_1, ..., \zeta_k$ real-valued functions on M. $\zeta_1, ..., \zeta_k$ are called affinely independent if the following property holds true:*
If $a_0, a_1, \ldots, a_k \in \mathbb{R}$ are real numbers satisfying

$$a_0 + \sum_{j=1}^{k} a_j \zeta_j(x) = 0, \quad \forall x \in M,$$

then $a_0 = a_1 = ... = a_k = 0$.

(ii) *Let $(\mathfrak{X}, \mathfrak{B})$ be a measurable space, $k \in \mathbb{N}$ and $T_1, ..., T_k$ real-valued $\mathfrak{B} - \mathbb{B}^1$-measurable functions on $(\mathfrak{X}, \mathfrak{B})$. Furthermore, let Q be a measure on $(\mathfrak{X}, \mathfrak{B})$. $T_1, ..., T_k$ are called Q-affinely independent if the mappings $T_1 \mathbb{1}_{N^c}, ..., T_k \mathbb{1}_{N^c}$ are affinely independent for all $N \in \mathfrak{B}$ with $Q(N) = 0$, where, here and in the following, $\mathbb{1}_A$ denotes the indicator function of a measurable set A.*
Moreover, given a set $\mathfrak{M} \neq \emptyset$ of measures on $(\mathfrak{X}, \mathfrak{B})$, $T_1, ..., T_k$ are called $\mathfrak{M}$-affinely independent if $T_1, ..., T_k$ are Q-affinely independent for every measure $Q \in \mathfrak{M}$ in the sense above.

Now, we cite a well-known useful theorem.

Theorem 2.1.9
Let $\mathfrak{P}$ be an exponential family in $\zeta_1, \ldots, \zeta_k$ and $T_1, \ldots, T_k$ according to Def. 2.1.1. Then, the following statements hold true:

(i) *$\mathfrak{P}$ is strictly k-parametrical if and only if $\zeta_1, ...\zeta_k$ are affinely independent and $T_1, ..., T_k$ are $\mathfrak{P}$-affinely independent.*

(ii) *$T_1, ..., T_k$ are $\mathfrak{P}$-affinely independent if and only if there exists a $\vartheta \in \Theta$ such that $\mathbf{Cov}_\vartheta(\boldsymbol{T})$ is positive definite. In the latter case, we briefly write $\mathrm{Cov}_\vartheta(\boldsymbol{T}) > 0$.*

Proof. In Witting (1985), Thm. 1.153, p. 145. For a proof of the first statement, see also Barndorff-Nielsen (1978), Cor. 8.1., p. 113. ■

We have mentioned before that the minimal representation of an exponential family is not unique. E.g., multiplicative scalars or additive constants can easily be added to the ζ's and the T's. In fact, no other transformations lead to further minimal representations of $\mathfrak{P}$ as the following theorem shows.

Theorem 2.1.10
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family in $\zeta_1, \ldots, \zeta_k$ and $T_1, \ldots, T_k$ according to Def. 2.1.1 and Def. 2.1.7. Then, the following statements hold true:

(i) *If $\mathfrak{P}$ is a $\tilde{k}$-parametrical exponential family in $\tilde{\zeta}_1, ..., \tilde{\zeta}_{\tilde{k}}$ and $\tilde{T}_1, ..., \tilde{T}_{\tilde{k}}$ in virtue of (2.1.9), then follows $\tilde{k} \geq k$, and there exist two constant matrices $\mathrm{A}, \mathrm{B} \in \mathbb{R}^{k \times \tilde{k}}$ of rank k and two constant vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^k$ with*

$$\boldsymbol{\zeta} = \mathrm{A}\tilde{\boldsymbol{\zeta}} + \boldsymbol{a} \quad \text{and} \quad \boldsymbol{T} = \mathrm{B}\tilde{\boldsymbol{T}} + \boldsymbol{b} \quad [\mathfrak{P}],$$

where $\tilde{\boldsymbol{\zeta}} = (\tilde{\zeta}_1, ..., \tilde{\zeta}_{\tilde{k}})'$ and $\tilde{\boldsymbol{T}} = (\tilde{T}_1, ..., \tilde{T}_{\tilde{k}})'$.

(ii) *$\boldsymbol{\zeta}$ is uniquely and $\boldsymbol{T}$ $[\mathfrak{P}]$-uniquely determined up to not degenerated affine transformations (in the sense of (i) with $\tilde{k} = k$).*

Proof. A proof of statement (i) can be found in Barndorff-Nielsen (1978), La. 8.1, p. 112. For a proof of the second statement consult Witting (1985), Cor. 1.154, p. 146. ■

2.1.3 Natural Parameter Space and Regularity Properties

In Subsection 2.1.1, Rem. 2.1.2, we mentioned that $C(\vartheta)$ in (2.1.1) is just a normalizing constant. Hence, we can think about an extension of Θ and $\mathfrak{P}$ in a natural way.

Definition 2.1.11 (Natural parameter space; natural extension)
Let $\mathfrak{P}$ be an exponential family according to Def. 2.1.1. We define

$$\Theta^* = \{\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_k)' \in \mathbb{R}^k : \ 0 < \int \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x) d\mu(x) < \infty\},$$

and for $\boldsymbol{\zeta} \in \Theta^*$

$$C^*(\boldsymbol{\zeta}) = \left[\int \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x) d\mu(x)\right]^{-1},$$

$$f^*_{\boldsymbol{\zeta}}(x) = C^*(\boldsymbol{\zeta}) \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x), \quad x \in \mathfrak{X}, \tag{2.1.10}$$

$$\text{and} \quad P^*_{\boldsymbol{\zeta}} = f^*_{\boldsymbol{\zeta}} \mu.$$

Then, Θ^* is called the natural parameter space of $\mathfrak{P}$ and $\mathfrak{P}^* = \{P^*_{\boldsymbol{\zeta}} : \ \boldsymbol{\zeta} \in \Theta^*\}$ the natural extension of $\mathfrak{P}$.

Remark 2.1.12
Obviously, $\boldsymbol{\zeta}(\Theta) \subseteq \Theta^*$ and $\mathfrak{P} \subseteq \mathfrak{P}^*$.

Working with the natural parameter space Θ^* and the *natural parameters* $\zeta_1, ..., \zeta_k$ of an exponential family brings along a variety of pleasant regularity properties which play an important role, e.g., in statistical inference. In Def. 2.1.1, no further assumptions on the functions $\zeta_1, ..., \zeta_k$ are stated and, thus, they are possibly not differentiable at some points of the interior $int(\Theta)$ of Θ. If the *natural representation* (2.1.10) of the exponential family is used, this problem vanishes, and it turns out that

C^* and $f^*_{\boldsymbol{\zeta}}$ (considered as a function of $\boldsymbol{\zeta}$) are infinitely often differentiable with respect to $\boldsymbol{\zeta} \in int(\Theta^*)$. In fact, this regularity property is even true for every function $\beta(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[\varphi]$ with some $\mathfrak{P}^*$-integrable function φ on $(\mathfrak{X}, \mathfrak{B})$, where the derivatives of β can be obtained by differentiating under the integral sign. As a consequence, moments of the statistics $T_1, ..., T_k$ can readily be computed.
To see all this, firstly, we are concerned with the properties of the natural parameter space itself.

Theorem 2.1.13
Let $\mathfrak{P}$ be an exponential family in $\zeta_1, \ldots, \zeta_k$ and $T_1, \ldots, T_k$ according to Def. 2.1.1. With the denotations of Def. 2.1.11, the following statements hold true:

(i) *Θ^* is a convex subset of $\mathbb{R}^k$ and $\kappa^* : \Theta^* \to \mathbb{R} : \boldsymbol{\zeta} \mapsto -\ln(C^*(\boldsymbol{\zeta}))$ a convex function on Θ^*.*

(ii) *If $\mathfrak{P}$ is strictly k-parametrical, then the interior of Θ^* is not empty, and κ^* is even strictly convex on Θ^*.*

Proof. A proof of the first part of (i) can be found in Lehmann & Romano (2005), La. 2.7.1, p. 48. The second part of assertion (i) follows from Thm. 7.1., p. 103, in Barndorff-Nielsen (1978) by consideration of the Laplace transform of $\nu^{\boldsymbol{T}}$. For statement (ii) see Witting (1985), Thm. 1.161, p. 150. ∎

Statement (i) of Thm. 2.1.13 is true, whether $\mathfrak{P}$ is strictly k-parametrical or not. In the first case, Θ^* is a convex subset of $\mathbb{R}^k$ containing a non-empty interior. In the latter case, Θ^* may lie in a linear subspace of $\mathbb{R}^k$ with dimension $\tilde{k} < k$ (if the ζ's in (2.1.1) satisfy a linear constraint), and the natural parameters $\zeta_1, \ldots, \zeta_k$ might be *unidentifiable* (if the T's in (2.1.1) satisfy a linear constraint), i.e., the mapping $\boldsymbol{\zeta} \mapsto P_{\boldsymbol{\zeta}}$, $\boldsymbol{\zeta} \in \Theta^*$, might not be injective (see Lehmann & Casella (1998), p. 24).
However, by means of an appropriate reduction of the number of involved parameters, it is always possible to obtain a strictly parametrical representation of an exponential family which can subsequently be written in the form (2.1.10) in virtue of a reparametrization of the parameters. Hence, from now on, we will many a time assume that $\mathfrak{P}$ is strictly k-parametrical and given by its natural representation, where, additionally, the parameter space Θ is assumed to be the natural parameter space Θ^*. Moreover, as it is often done in literature, we will assume that $\Theta(= \Theta^*)$ is open as it is the case in most applications (Lehmann & Casella (1998), p. 24). Summarizing, we have the assumptions that $\mathfrak{P} = \{P_{\boldsymbol{\zeta}} = f_{\boldsymbol{\zeta}}\mu : \boldsymbol{\zeta} \in \Theta\}$, $\Theta = \Theta^* \subseteq \mathbb{R}^k$ open, forms a strictly k-parametrical exponential family in the natural parameters $\zeta_1, \ldots, \zeta_k$, more precisely, in the projections $p_1, \ldots, p_k$ with $p_j(\boldsymbol{\zeta}) = \zeta_j$ for $1 \leq j \leq k$, and statistics $T_1, \ldots, T_k$ which is of full rank, where for $\boldsymbol{\zeta} \in \Theta$

$$f_{\boldsymbol{\zeta}}(x) = C(\boldsymbol{\zeta}) \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x) = \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x) - \kappa(\boldsymbol{\zeta})\right\} h(x), \quad x \in \mathfrak{X}, \tag{2.1.11}$$

and $\kappa(\boldsymbol{\zeta}) = -\ln(C(\boldsymbol{\zeta}))$. In what follows, we will frequently refer to situation (2.1.11).
Now, as an immediate consequence of the dominated convergence theorem, we obtain the following regularity properties.

Lemma 2.1.14
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11), and let $\varphi : (\mathfrak{X}, \mathfrak{B}) \to (\mathbb{R}^1, \mathbb{B}^1)$ be $P_{\boldsymbol{\zeta}}$-integrable for all $\boldsymbol{\zeta} \in \Theta$. Then, the mapping

$$\tilde{\beta} : \quad \Theta \to \mathbb{R} : \quad \boldsymbol{\zeta} \mapsto \int \varphi(x) \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x) d\mu(x) \tag{2.1.12}$$

is continuous and has derivatives of all orders with respect to $\zeta_1, ..., \zeta_k$ which can be obtained by differentiating under the integral sign, i.e.,

$$\nabla_1^{l_1} ... \nabla_k^{l_k} \tilde{\beta}(\boldsymbol{\zeta}) = \int \varphi(x) T_1^{l_1}(x) ... T_k^{l_k}(x) \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x) d\mu(x), \quad \boldsymbol{\zeta} \in \Theta. \tag{2.1.13}$$

Proof. See Witting (1985), Cor. 1.163, p. 152, and Lehmann & Casella (1998), Thm. 5.8, p. 27. ■

As a consequence of La. 2.1.14, regularity properties of C and κ follow as well as a variety of equations concerning the relationship between the derivatives of these functions and the moments of the statistic $\boldsymbol{T}$. For example, if we set $\varphi \equiv 1$ in (2.1.12), we obtain that $1/C$ and, thus, C and κ are infinitely often differentiable in $\boldsymbol{\zeta} \in \Theta$. Then, (2.1.13) yields that, e.g., $\nabla_1(1/C)(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[T_1]/C(\boldsymbol{\zeta})$ and, hence, $E_{\boldsymbol{\zeta}}[T_1] = -\nabla_1(\ln(C))(\boldsymbol{\zeta}) = \nabla_1 \kappa(\boldsymbol{\zeta})$.
This and more results are presented in the following theorem of Witting (1985).

Theorem 2.1.15
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11). Then, we obtain the following statements:

(i) For every $\boldsymbol{\zeta} \in \Theta$, the statistic $\boldsymbol{T} = (T_1, ..., T_k)'$ has finite moments of any order with respect to $P_{\boldsymbol{\zeta}}$, and the functions C, κ and $\tilde{\boldsymbol{\zeta}} \mapsto E_{\tilde{\boldsymbol{\zeta}}}[T_1^{l_1} ... T_k^{l_k}]$, $\tilde{\boldsymbol{\zeta}} \in \Theta$, are infinitely often differentiable in $\boldsymbol{\zeta}$ and fulfill

$$\nabla \kappa(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[\boldsymbol{T}], \tag{2.1.14}$$

$$\mathbf{H}_\kappa(\boldsymbol{\zeta}) = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}) \tag{2.1.15}$$

$$\text{and} \quad E_{\boldsymbol{\zeta}}[T_1^{l_1} ... T_k^{l_k}] = C(\boldsymbol{\zeta}) \nabla_1^{l_1} ... \nabla_k^{l_k} \int \exp\left\{\sum_{j=1}^{k} \zeta_j T_j(x)\right\} h(x) d\mu(x).$$

At this, $\mathbf{H}_\kappa(\boldsymbol{\zeta}) = [\nabla_i \nabla_j \kappa(\boldsymbol{\zeta})]_{1 \le i,j \le k}$ denotes the Hessian matrix of κ at $\boldsymbol{\zeta} \in \Theta$.

(ii) Let $\varphi : (\mathfrak{X}, \mathfrak{B}) \to (\mathbb{R}^1, \mathbb{B}^1)$ be $P_{\boldsymbol{\zeta}}$-integrable for all $\boldsymbol{\zeta} \in \Theta$. Then, the mapping

$$\beta_\varphi : \quad \Theta \to \mathbb{R} : \quad \boldsymbol{\zeta} \mapsto E_{\boldsymbol{\zeta}}[\varphi]$$

is infinitely often differentiable and fulfils

$$\nabla \beta_\varphi(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[\varphi \boldsymbol{T}] - E_{\boldsymbol{\zeta}}[\varphi] E_{\boldsymbol{\zeta}}[\boldsymbol{T}] = \boldsymbol{Cov}_{\boldsymbol{\zeta}}(\varphi, \boldsymbol{T}).$$

Proof. In Witting (1985), Thm. 1.164, pp. 152/153. ■

In the context of situation (2.1.11), we additionally introduce the mapping

$$\pi: \quad \Theta \to \pi(\Theta): \quad \boldsymbol{\zeta} \mapsto E_{\boldsymbol{\zeta}}[\boldsymbol{T}], \tag{2.1.16}$$

which will be helpful in the following section, in particular, with regard to maximum likelihood estimation. Obviously, $\pi(\boldsymbol{\zeta}) = \nabla\kappa(\boldsymbol{\zeta})$ and, thus, π is continuously differentiable with Jacobian matrix $\mathbf{D}_\pi(\boldsymbol{\zeta}) = \mathbf{H}_\kappa(\boldsymbol{\zeta}) = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}) > 0$, $\boldsymbol{\zeta} \in \Theta$ (cf. Thm. 2.1.15 (i) and Thm. 2.1.9). Moreover, π is bijective (cf. Witting (1985), Thm. 1.170, p. 157), and, hence, π possess a continuously differentiable inverse function on $\pi(\Theta)$ which will be denoted by π^{-1}.

Finally, for a better understanding, we illustrate the findings of this subsection by means of an example.

Example 2.1.16
The class of binomial distributions $\mathfrak{P} = \{bin(1,p): p \in (0,1)\}$ forms a one-parameter exponential family, where a density of $bin(1,p)$, $p \in (0,1)$, with respect to the counting measure ϵ is given by

$$f_p(x) = p^x(1-p)^{1-x}\mathbb{1}_{\{0,1\}}(x) = \exp\left\{\ln\left(\frac{p}{1-p}\right)x - (-\ln(1-p))\right\}\mathbb{1}_{\{0,1\}}(x).$$

The reparametrization $\zeta = \ln\left(\frac{p}{1-p}\right)$, $p \in (0,1)$, implies $p = \frac{1}{1+e^{-\zeta}}$ and, thus,

$$-\ln(1-p) = -\ln\left(1 - \frac{1}{1+e^{-\zeta}}\right) = \ln(1+e^{-\zeta}) + \zeta.$$

Setting $\kappa^*(\zeta) = \ln(1+e^{-\zeta}) + \zeta$, $\zeta \in \mathbb{R}$, the natural representation of $\mathfrak{P}$ is given by the ϵ-densities

$$f^*_\zeta(x) = \exp\{\zeta x - \kappa^*(\zeta)\}\mathbb{1}_{\{0,1\}}(x), \quad \zeta \in \mathbb{R}.$$

Here, the natural parameter space Θ^* equals $\mathbb{R}$, and the natural extension of $\mathfrak{P}$ is $\mathfrak{P}^* = \{f^*_\zeta\epsilon : \zeta \in \mathbb{R}\}$ and coincides with $\mathfrak{P}$. As a consequence, $\pi(\zeta) = \frac{d}{d\zeta}\kappa^*(\zeta) = \frac{1}{1+e^{-\zeta}}$ and $\pi(\mathbb{R}) = (0,1)$.

2.1.4 Moment Generating Function

In this subsection, we briefly draw attention to a helpful result related to the moment generating function in the context of exponential families.

Definition 2.1.17 (Moment generating function)
Let $\boldsymbol{X} = (X_1, ..., X_k)'$ be a random vector on the probability space $(\Omega, \mathfrak{A}, P)$. The moment generating function of $\boldsymbol{X}$, respectively $P^{\boldsymbol{X}}$, is defined as

$$m_{\boldsymbol{X}}: \quad \mathbb{R}^k \to \overline{\mathbb{R}}: \quad \boldsymbol{t} = (t_1, ..., t_k)' \mapsto E[\exp\{\boldsymbol{t}'\boldsymbol{X}\}].$$

Now, we obtain the following lemma.

Lemma 2.1.18
Let $\mathfrak{P}$ be a k-parametrical exponential family according to (2.1.11), not necessarily strictly k-parametrical, and let $\boldsymbol{\zeta} \in \Theta$. Moreover, let $V \subseteq \mathbb{R}^k$ be an open neighbourhood of $\mathbf{0} \in \mathbb{R}^k$ with the property that $\boldsymbol{\zeta} + V = \{\boldsymbol{\zeta} + \boldsymbol{v} : \boldsymbol{v} \in V\} \subseteq \Theta$. Then,

$$m_{\boldsymbol{T}}(\boldsymbol{t}) = \frac{C(\boldsymbol{\zeta})}{C(\boldsymbol{\zeta} + \boldsymbol{t})}, \quad \boldsymbol{t} \in V, \tag{2.1.17}$$

where integration is with respect to $P_{\boldsymbol{\zeta}}$.

Proof. In the Appendix. ■

Obviously, remembering the one-to-one-correspondence of distribution and moment generating function, this result might be of use in order to derive the distribution of the vector $\boldsymbol{T}$ of statistics which plays a prominent role in statistical inference, as the following subsection makes clear.

2.1.5 Sufficiency, Minimal Sufficiency and Completeness

For the considered family of distributions, the concepts of *(minimal) sufficiency* and *completeness* of statistics are fundamental for inferential issues. It is well-known that, in case of an underlying exponential family structure, respective statistics can readily be found.

Theorem 2.1.19
Let $\mathfrak{P}$ be a k-parametrical exponential family in $\zeta_1, \dots, \zeta_k$ and $T_1, \dots, T_k$ according to Def. 2.1.1. Then, the following statements are true:

(i) The statistic $\boldsymbol{T}$ is sufficient for $\mathfrak{P}$.

(ii) If the interior of $\boldsymbol{\zeta}(\Theta)$ is not empty, then $\boldsymbol{T}$ is complete for $\mathfrak{P}$.

Proof. Statement (i) can be directly obtained from the factorization criterion (cf., e.g., Lehmann & Romano (2005), Cor. 2.6.1, p. 46, or Witting (1985), Thm. 3.19, pp. 344/345). From Thm. 3.39, p. 356 in Witting (1985), we obtain that $\boldsymbol{T}$ is complete for $\mathfrak{P}$ if the interior of $\boldsymbol{\zeta}(\Theta)$ is not empty. ■

We state our findings explicitly for situation (2.1.11).

Lemma 2.1.20
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11). Then $\boldsymbol{T}$ is sufficient and complete for $\mathfrak{P}$ and, moreover, minimal sufficient for $\mathfrak{P}$, i.e., for every sufficient statistic $\tilde{\boldsymbol{T}}$: $(\mathfrak{X}, \mathfrak{B}) \to (\mathfrak{Z}, \mathfrak{C})$ for $\mathfrak{P}$, there exists a measurable function $h : (\mathfrak{Z}, \mathfrak{C}) \to (\mathbb{R}^k, \mathbb{B}^k)$ with $\boldsymbol{T} = h \circ \tilde{\boldsymbol{T}}$ $[\mathfrak{P}]$.

Proof. By assumption, $\mathfrak{P}$ is of full rank and, hence, the results are obvious from Thm. 2.1.19, where minimal sufficiency is obtained from Lehmann & Casella (1998), Cor. 6.16, p. 39. ■

2.1.6 Score Statistic and Fisher Information Matrix

In mathematical statistics, the terminologies score statistic and Fisher information matrix are closely linked to the quality or, more precisely, to the *efficiency* and *asymptotic efficiency* of estimators, as we will see in Section 2.2. For the moment, we focus on the introduction of the Fisher information matrix and its representation related to the statistic $\boldsymbol{T}$ if an underlying exponential family structure is assumed. In general, the definition is as follows.

Definition 2.1.21 (Score function; Fisher information matrix)
Let μ be a σ-finite measure and $\mathfrak{P} = \{P_{\boldsymbol{\zeta}} = f_{\boldsymbol{\zeta}}\mu : \boldsymbol{\zeta} \in \Theta \subseteq \mathbb{R}^k\}$ be a family of probability measures on a measurable space $(\mathfrak{X}, \mathfrak{B})$. Moreover, let $\boldsymbol{\zeta} \in int(\Theta)$. If all appearing derivatives and integrals with respect to $\boldsymbol{\zeta}$, respectively $P_{\boldsymbol{\zeta}}$, exist, then, the function $\boldsymbol{U}_{\boldsymbol{\zeta}} = \nabla_{\boldsymbol{\zeta}}[\ln(f_{\boldsymbol{\zeta}})]$ on $(\mathfrak{X}, \mathfrak{B})$, is called the score function and

$$\mathbf{I}_f(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[\boldsymbol{U}_{\boldsymbol{\zeta}}\boldsymbol{U}'_{\boldsymbol{\zeta}}] \tag{2.1.18}$$

the Fisher information matrix of $\mathfrak{P}$ at $\boldsymbol{\zeta}$.

If X is a random element (random variable or vector) on some probability space with values in $(\mathfrak{X}, \mathfrak{B})$ and distribution $P_{\boldsymbol{\zeta}}$, $\boldsymbol{U}_{\boldsymbol{\zeta}}(X) = \nabla_{\boldsymbol{\zeta}}[\ln(f_{\boldsymbol{\zeta}}(X))]$ is also termed the *score statistic* of X at $\boldsymbol{\zeta}$.
The exponential family structure allows easy representations of $\boldsymbol{U}_{\boldsymbol{\zeta}}$ and $\mathbf{I}_f(\boldsymbol{\zeta})$ in terms of moments of the statistics $T_1, \ldots, T_k$ in the following way.

Theorem 2.1.22
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11) and let $\boldsymbol{\zeta} \in \Theta$. Then, the following equations are true:

$$\begin{aligned} \boldsymbol{U}_{\boldsymbol{\zeta}} &= \boldsymbol{T} - E_{\boldsymbol{\zeta}}[\boldsymbol{T}], \\ \mathbf{I}_f(\boldsymbol{\zeta}) &= \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}). \end{aligned}$$

Proof. Plugging in the density $f_{\boldsymbol{\zeta}}$ given by (2.1.11), Thm. 2.1.15 (i) leads to $\boldsymbol{U}_{\boldsymbol{\zeta}} = \boldsymbol{T} - \nabla\kappa(\boldsymbol{\zeta}) = \boldsymbol{T} - E_{\boldsymbol{\zeta}}[\boldsymbol{T}]$. Evidently, $E_{\boldsymbol{\zeta}}[\boldsymbol{U}_{\boldsymbol{\zeta}}] = \mathbf{0}$ and, thus, $\mathbf{I}_f(\boldsymbol{\zeta}) = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{U}_{\boldsymbol{\zeta}}) = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T})$. ■

Thm. 2.1.22 in combination with Thm. 2.1.9 implies that, in situation (2.1.11), the Fisher information matrix $\mathbf{I}_f(\boldsymbol{\zeta})$ is positive definite for all $\boldsymbol{\zeta} \in \Theta$. Moreover, we obtain from Thm. 2.1.15 (i), that all entries of $\mathbf{I}_f(\boldsymbol{\zeta})$ are finite, i.e., we have for all $\boldsymbol{\zeta} \in \Theta$

$$\mathbf{I}_f(\boldsymbol{\zeta}) > 0, \quad [\mathbf{I}_f(\boldsymbol{\zeta})]_{i,j} < \infty, \quad 1 \leq i, j \leq k. \tag{2.1.19}$$

The Fisher information matrix clearly depends on the parametrization of $\mathfrak{P}$. We point out how to compute the new Fisher information matrix based on the actual one if the parameters are regularly transformed. Subsequently, we give an example for such a regular transformation.

Lemma 2.1.23
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11) and let $h : \Theta \to \Gamma \subseteq \mathbb{R}^k$ be a continuously differentiable and bijective function with $|\mathbf{D}_h(\boldsymbol{\zeta})| \neq 0$, i.e. the determinant of the Jacobian matrix of h is not zero, for all $\boldsymbol{\zeta} \in \Theta$. We define another representation $\tilde{\mathfrak{P}} = \{\tilde{P}_{\boldsymbol{\gamma}} =$

$P_{h^{-1}(\gamma)} : \gamma \in \Gamma\}$ of $\mathfrak{P}$ and let $\tilde{\mathbf{I}}_f(\boldsymbol{\gamma})$ denote the corresponding Fisher information matrix at $\boldsymbol{\gamma} \in \Gamma$. Then, $h^{-1} : \Gamma \to \Theta$ is continuously differentiable with $\mathbf{D}_{h^{-1}}(\boldsymbol{\gamma}) = \mathbf{D}_h(h^{-1}(\boldsymbol{\gamma}))^{-1}$, $\boldsymbol{\gamma} \in \Gamma$, and

$$\tilde{\mathbf{I}}_f(\boldsymbol{\gamma}) = \mathbf{D}_{h^{-1}}(\boldsymbol{\gamma})'\mathbf{I}_f(h^{-1}(\boldsymbol{\gamma}))\mathbf{D}_{h^{-1}}(\boldsymbol{\gamma}), \quad \gamma \in \Gamma.$$

Proof. In Witting (1985), Thm. 1.167, p. 156. ∎

Notice, that under the assumptions of La. 2.1.23, (2.1.19) yields that $\mathbf{I}_f(\boldsymbol{\zeta})$, $\boldsymbol{\zeta} \in \Theta$, and, for this reason, $\tilde{\mathbf{I}}_f(\boldsymbol{\gamma})$, $\boldsymbol{\gamma} \in \Gamma$, are invertible with

$$\begin{aligned}\tilde{\mathbf{I}}_f(\boldsymbol{\gamma})^{-1} &= \mathbf{D}_{h^{-1}}(\boldsymbol{\gamma})^{-1}\mathbf{I}_f(h^{-1}(\boldsymbol{\gamma}))^{-1}(\mathbf{D}_{h^{-1}}(\boldsymbol{\gamma})^{-1})' \\ &= \mathbf{D}_h(h^{-1}(\boldsymbol{\gamma}))\mathbf{I}_f(h^{-1}(\boldsymbol{\gamma}))^{-1}(\mathbf{D}_h(h^{-1}(\boldsymbol{\gamma})))'. \end{aligned} \tag{2.1.20}$$

Example 2.1.24 (Mean value parametrization)
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11) and remember that $\pi : \Theta \to \pi(\Theta) : \boldsymbol{\zeta} \mapsto E_{\boldsymbol{\zeta}}[\boldsymbol{T}]$ is bijective and continuously differentiable on Θ with $\mathbf{D}_\pi(\boldsymbol{\zeta}) = \mathbf{H}_\kappa(\boldsymbol{\zeta}) = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}) > 0$, $\boldsymbol{\zeta} \in \Theta$ (cf. (2.1.15) and (2.1.16)). Thus, π fulfils the conditions of La. 2.1.23, and we obtain for $\gamma \in \Gamma = \pi(\Theta)$

$$D_{\pi^{-1}}(\boldsymbol{\gamma}) = \mathbf{D}_\pi(\pi^{-1}(\boldsymbol{\gamma}))^{-1} = \mathbf{Cov}_{\pi^{-1}(\boldsymbol{\gamma})}(\boldsymbol{T})^{-1},$$

and, hence, by application of Thm. 2.1.22,

$$\tilde{\mathbf{I}}_f(\gamma) = \mathbf{Cov}_{\pi^{-1}(\boldsymbol{\gamma})}(\boldsymbol{T})^{-1}\mathbf{Cov}_{\pi^{-1}(\boldsymbol{\gamma})}(\boldsymbol{T})\mathbf{Cov}_{\pi^{-1}(\boldsymbol{\gamma})}(\boldsymbol{T})^{-1} = \mathbf{Cov}_{\pi^{-1}(\boldsymbol{\gamma})}(\boldsymbol{T})^{-1}.$$

2.1.7 Product Measures

In this subsection, aiming at statistical inference in Sections 2.2 and 2.3, we continue by considering product probability spaces and measures respectively.

Lemma 2.1.25

(i) Let $\Theta_i \neq \emptyset$, $1 \leq i \leq s$, be parameter sets, and let $\mathfrak{P}_i = \{P_{\vartheta_i;i} = f_{\vartheta_i;i}\,\mu_i : \vartheta_i \in \Theta_i\}$ be an exponential family on $(\mathfrak{X}_i, \mathfrak{B}_i)$ according to Def. 2.1.1, $1 \leq i \leq s$. Then, the family of product probability measures $\{\otimes_{i=1}^s P_{\vartheta_i;i} = \prod_{i=1}^s f_{\vartheta_i;i} \otimes_{i=1}^s \mu_i : \vartheta_i \in \Theta_i,\ 1 \leq i \leq s\}$ forms an exponential family on the product space $(\times_{i=1}^s \mathfrak{X}_i, \otimes_{i=1}^s \mathfrak{B}_i)$.

(ii) Let $\mathfrak{P} = \{P_\vartheta = f_\vartheta \mu : \vartheta \in \Theta\}$ be a k-parametrical exponential family in $\zeta_1, \ldots, \zeta_k$ and $T_1, \ldots, T_k$ according to Def. 2.1.1. Then, a $\mu^{(s)} = (\otimes_{i=1}^s \mu)$-density of $P_\vartheta^{(s)} = \otimes_{i=1}^s P_\vartheta$ is given by

$$\begin{aligned} f_\vartheta^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= C(\vartheta)^s \exp\left\{\sum_{j=1}^k \zeta_j(\vartheta) T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right\}\left(\prod_{i=1}^s h(x^{(i)})\right) \\ &= C(\vartheta)^s \exp\left\{\boldsymbol{\zeta}(\vartheta)'\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right\}\left(\prod_{i=1}^s h(x^{(i)})\right),\ \tilde{\boldsymbol{x}}^{(s)} = (x^{(1)}, \ldots, x^{(s)}) \in \times_{i=1}^s \mathfrak{X}, \end{aligned}$$

where $\boldsymbol{T}^{(s)} = (T_1^{(s)}, ..., T_k^{(s)})'$ *and*

$$T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \sum_{i=1}^{s} T_j(x^{(i)}), \quad \tilde{\boldsymbol{x}}^{(s)} = (x^{(1)}, \ldots, x^{(s)}) \in \times_{i=1}^{s} \mathfrak{X}, \quad 1 \leq j \leq k. \tag{2.1.21}$$

Hence, $\mathfrak{P}^{(s)} = \{P_\vartheta^{(s)} = f_\vartheta^{(s)} \mu^{(s)} : \vartheta \in \Theta\}$ *forms a* k*-parametrical exponential family in* $\zeta_1, \ldots, \zeta_k$ *and* $T_1^{(s)}, \ldots, T_k^{(s)}$ *on the product space* $(\mathfrak{X}^{1\times s} = \times_{i=1}^{s} \mathfrak{X}, \mathfrak{B}^s = \otimes_{i=1}^{s} \mathfrak{B})$*.*
Moreover, if $\mathfrak{P}$ *is strictly* k*-parametrical, then* $\mathfrak{P}^{(s)}$ *is strictly* k*-parametrical, too.*

Proof. In Witting (1985), Thm. 1.157, p. 148. ■

Obviously, the exponential family structure is preserved by considering the respective product probability measures. As an immediate consequence, using the same denotations as in La. 2.1.25, we obtain the following important statements.

Theorem 2.1.26
Let $\mathfrak{P} = \{P_\vartheta = f_\vartheta \mu : \vartheta \in \Theta\}$ *be a* k*-parametrical exponential family in* $\zeta_1, \ldots, \zeta_k$ *and* $T_1, \ldots, T_k$ *according to Def. 2.1.1. Then, the following assertions are true:*

(i) $\boldsymbol{T}^{(s)}$ *is sufficient for* $\mathfrak{P}^{(s)}$*.*

(ii) If $int(\boldsymbol{\zeta}(\Theta)) \neq \emptyset$*, then* $\boldsymbol{T}^{(s)}$ *is sufficient and complete for* $\mathfrak{P}^{(s)}$*.*

Proof. The assertions follow directly by application of La. 2.1.25 (ii) and Thm. 2.1.19. ■

Once again, we separately propose the properties for situation (2.1.11).

Lemma 2.1.27
Let $\mathfrak{P}$ *be a strictly* k*-parametrical exponential family according to (2.1.11). Then,* $\boldsymbol{T}^{(s)}$ *is minimal sufficient and complete for* $\mathfrak{P}^{(s)}$*.*

Proof. By assumption, $\mathfrak{P}$ is of type (2.1.11), and so is $\mathfrak{P}^{(s)}$ by application of La. 2.1.25 (ii). The assertion then follows from La. 2.1.20. ■

We end this subsection by stating the score statistic and the Fisher information matrix of the family $\mathfrak{P}^{(s)}$ of product probability measures, when situation (2.1.11) holds.

Lemma 2.1.28
Let $\mathfrak{P}$ *be a strictly* k*-parametrical exponential family according to (2.1.11). Then, the score function* $\boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)} = \nabla_{\boldsymbol{\zeta}}[\ln(f_{\boldsymbol{\zeta}}^{(s)})]$ *on* $(\mathfrak{X}^{1\times s}, \mathfrak{B}^s)$ *and the Fisher information matrix* $\mathbf{I}_f^{(s)}(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[U_{\boldsymbol{\zeta}}^{(s)}(U_{\boldsymbol{\zeta}}^{(s)})']$ *of* $\mathfrak{P}^{(s)}$ *at* $\boldsymbol{\zeta}$ *are given by the equations*

$$\begin{aligned} \boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)} &= \boldsymbol{T}^{(s)} - E_{\boldsymbol{\zeta}}[\boldsymbol{T}^{(s)}], \\ \mathbf{I}_f^{(s)}(\boldsymbol{\zeta}) &= \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}^{(s)}). \end{aligned}$$

Proof. Again, by assumption and La. 2.1.25, $\mathfrak{P}^{(s)}$ is of type (2.1.11) and, hence, Thm. 2.1.22 yields the assertion. ■

If $X^{(1)}, \ldots, X^{(s)}$ are iid random elements (on the same probability space) with values in $(\mathfrak{X}, \mathfrak{B})$, where $X^{(i)}$ has distribution $P_{\boldsymbol{\zeta}}$, $i = 1, \ldots, s$, $\boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = \nabla_{\boldsymbol{\zeta}}[\ln(f_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}))]$ is also termed the score statistic of $\tilde{\boldsymbol{X}}^{(s)} = (X^{(1)}, \ldots, X^{(s)})$ at $\boldsymbol{\zeta}$.
From the definition of the Fisher information matrix and Thm. 2.1.22, it is easily seen that

$$\mathbf{I}_f^{(s)}(\boldsymbol{\zeta}) = s\,\mathbf{I}_f(\boldsymbol{\zeta}) = s\,\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}), \quad \boldsymbol{\zeta} \in \Theta. \tag{2.1.22}$$

2.1.8 Kullback-Leibler Distance

Preliminary for Section 2.3, we introduce a useful measure of the 'distance' of two distributions. As it is often done in parametrical statistics, we formally define the distance on $\Theta \times \Theta$, where a parameter is identified with the corresponding density and distribution (in virtue of a bijective mapping).

Definition 2.1.29 (Kullback-Leibler distance)
Let μ be a σ-finite measure and $\mathfrak{P} = \{f_{\boldsymbol{\vartheta}}\mu : \boldsymbol{\vartheta} \in \Theta\}$, $\Theta \subseteq \mathbb{R}^k$, be a parametric family of probability measures on a measurable space $(\mathfrak{X}, \mathfrak{B})$, where the densities $f_{\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \in \Theta$, are assumed to have the same support. Then, for $\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)} \in \Theta$,

$$d_{KL}(\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)}) = E_{\boldsymbol{\vartheta}^{(1)}}\left[\ln\left(\frac{f_{\boldsymbol{\vartheta}^{(1)}}}{f_{\boldsymbol{\vartheta}^{(2)}}}\right)\right] = \int \ln\left(\frac{f_{\boldsymbol{\vartheta}^{(1)}}(x)}{f_{\boldsymbol{\vartheta}^{(2)}}(x)}\right) f_{\boldsymbol{\vartheta}^{(1)}}(x)d\mu(x)$$

is called the Kullback-Leibler distance of $\boldsymbol{\vartheta}^{(1)}$ and $\boldsymbol{\vartheta}^{(2)}$. Here, by convention, $\frac{0}{0} = 1$.

Obviously, from Jensen's inequality (e.g., in Billingsley (1995), p. 276), we obtain that $d_{KL}(\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)}) \geq 0$, $\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)} \in \Theta$. However, it should be noted that d_{KL} is not symmetric in general, as it is usually the case for distance measures.
Once more, the exponential family structure simplifies calculations and leads to an easy and manageable representation.

Lemma 2.1.30
Let $\mathfrak{P}$ be a strictly k-parametrical exponential family according to (2.1.11). Then, for $\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)} \in \Theta$,

$$d_{KL}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}) = \kappa(\boldsymbol{\zeta}^{(2)}) - \kappa(\boldsymbol{\zeta}^{(1)}) + (\boldsymbol{\zeta}^{(1)} - \boldsymbol{\zeta}^{(2)})'\pi(\boldsymbol{\zeta}^{(1)}).$$

Proof. In the Appendix. ■

Finally, we remark on some regularity properties of d_{KL}.

Remark 2.1.31
Related to situation (2.1.11), it follows from La. 2.1.30 that, for fixed $\boldsymbol{\zeta}^{(1)} \in \Theta$, the Hessian matrix of the mapping $d_{KL}(\boldsymbol{\zeta}^{(1)}, \bullet)$ coincides with the Hessian matrix of κ and, thus, $d_{KL}(\boldsymbol{\zeta}^{(1)}, \bullet)$ is a strictly convex function on Θ (cf. Kallenberg (1978), La. 2.2.2, p. 15).

2.2 Point Estimation

In this section, we are concerned with the derivation of point estimators and their properties, when the underlying class of distributions forms an exponential family. Throughout Sections 2.2 and 2.3, if not otherwise specified, we consider the following sample situation.
Let $\Theta \subseteq \mathbb{R}^k$ be a parameter space, $(\Omega, \mathfrak{A})$ a measurable space and $\mathfrak{P} = \{P_{\boldsymbol{\zeta}} : \boldsymbol{\zeta} \in \Theta\}$ be a family of probability measures on $(\Omega, \mathfrak{A})$. Let X denote a random element on $(\Omega, \mathfrak{A})$ with values in a measurable space $(\mathfrak{X}, \mathfrak{B})$. The set of corresponding distributions $\mathfrak{P}^X = \{P_{\boldsymbol{\zeta}}^X : \boldsymbol{\zeta} \in \Theta\}$ of X is assumed to form a strictly k-parametrical exponential family in the natural parameters $\zeta_1, \dots, \zeta_k$ and statistics $T_1, \dots, T_k$ on $(\mathfrak{X}, \mathfrak{B})$ according to Def. 2.1.1, where a μ-density $f_{\boldsymbol{\zeta}}^X$ of $P_{\boldsymbol{\zeta}}^X$, $\boldsymbol{\zeta} \in \Theta$, is given by the right-hand side of (2.1.11). For simplicity, let Θ coincide with the natural parameter space of $\mathfrak{P}^X$, which is, moreover, assumed to be open.
Motivated by asymptotic theory, we assume to have an infinite number of iid replicates $X^{(1)}, X^{(2)}, \dots$ of X (on $(\Omega, \mathfrak{A})$) with corresponding observations $x^{(1)}, x^{(2)}, \dots$. For $s \in \mathbb{N}$, we define $\tilde{\boldsymbol{X}}^{(s)} = (X^{(1)}, \dots, X^{(s)})$ and $\tilde{\boldsymbol{x}}^{(s)} = (x^{(1)}, \dots, x^{(s)})$, respectively. Then, for every $s \in \mathbb{N}$, the class of probability measures $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}} = \{P_{\boldsymbol{\zeta}}^{\tilde{\boldsymbol{X}}^{(s)}} = \otimes_{i=1}^s P_{\boldsymbol{\zeta}}^{X^{(i)}} : \boldsymbol{\zeta} \in \Theta\}$ on the product space $(\mathfrak{X}^{1\times s}, \mathfrak{B}^s)$ forms a strictly k-parametrical exponential family in $\zeta_1, \dots, \zeta_k$ and statistics $T_1^{(s)}, \dots, T_k^{(s)}$, as defined in (2.1.21) (cf. La. 2.1.25 (ii)), and a $\mu^{(s)}$-density of $P_{\boldsymbol{\zeta}}^{\tilde{\boldsymbol{X}}^{(s)}}$, $\boldsymbol{\zeta} \in \Theta$, is given by

$$f_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \exp\left\{\sum_{j=1}^k \zeta_j T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - s\kappa(\boldsymbol{\zeta})\right\} \prod_{i=1}^s h(x^{(i)}), \quad \tilde{\boldsymbol{x}}^{(s)} \in \mathfrak{X}^{1\times s}. \tag{2.2.1}$$

Now, suppose that the true parameter $\boldsymbol{\zeta} \in \Theta$ and, thus, the true distribution $P_{\boldsymbol{\zeta}}^X$ of X, is unknown. Then, using well-known results concerning exponential families, convenient estimators of the parameter vector can readily be found along with many useful properties, which will be the subject of the following subsections.

2.2.1 Maximum Likelihood Estimation

The well-known procedure of maximum likelihood estimation is much simplified in the context of exponential families.

Lemma 2.2.1
If $\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \in \pi(\Theta)$ *with* π *as in (2.1.16),*

$$\pi^{-1}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right)$$

is the unique solution of the likelihood equation $\nabla_{\boldsymbol{\zeta}}(\ln(f_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))) = 0$ *and maximizes the likelihood function based on the observations* $x^{(1)}, \dots, x^{(s)}$, $s \in \mathbb{N}$.

Proof. In the Appendix. ■

Inspecting the proof of La. 2.2.1, a solution of the likelihood equation does not exist if the condition $\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \in \pi(\Theta)$ is not fulfilled. In that case, since Θ is assumed to be open, the likelihood function has no maximum in Θ and, hence, the maximum likelihood estimate of $\boldsymbol{\zeta}$ in Θ based on the observations $x^{(1)}, \ldots, x^{(s)}$ does not exist. We give an example.

Example 2.2.2
Suppose that in the situation of Ex. 2.1.16 ζ is unknown, and let $X^{(1)}, ..., X^{(s)} \overset{iid}{\sim} f_{\zeta}^{*}$ with corresponding observations $x^{(1)}, \ldots, x^{(s)}$. If $x^{(1)} = \cdots = x^{(s)} = 0$ or $x^{(1)} = \cdots = x^{(s)} = 1$, $\frac{1}{s}\sum_{i=1}^{s} x^{(i)}$ equals 0 or 1, and both values are not in the range of π. Hence, in these cases, a maximum likelihood estimate of $\zeta \in \mathbb{R}$ based on the observations $x^{(1)}, \ldots, x^{(s)}$ does not exist.

Hence, even if an exponential family structure with densities given by the natural representation is assumed, maximum likelihood estimates based on a given sample of s observations may not exist. However, roughly speaking, the larger the sample size s is, the more likely a solution of the likelihood equation exists, as the following theorem shows.

Theorem 2.2.3
If $P_{\boldsymbol{\zeta}}(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \in \pi(\Theta)) = 1$ for all $\boldsymbol{\zeta} \in \Theta$, then

$$\boldsymbol{\zeta}^{*(s)} = \pi^{-1}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\right)$$

is the unique maximum likelihood estimator (MLE) of $\boldsymbol{\zeta} \in \Theta$ based on s independent observations of X. Moreover, for all $\boldsymbol{\zeta} \in \Theta$,

$$\lim_{s\to\infty} P_{\boldsymbol{\zeta}}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \in \pi(\Theta)\right) = 1,$$

and, thus, the MLE $\boldsymbol{\zeta}^{*(s)}$ of $\boldsymbol{\zeta}$ based on s independent observations exists in Θ with $P_{\boldsymbol{\zeta}}$-probability tending to 1 if the sample size s tends to infinity.

Proof. In the Appendix. ■

In terms of Ex. 2.2.2, the probability of non-existence of the MLE of ζ in $\Theta^* = \mathbb{R}$ based on s independent observations equals $\left(\frac{1}{1+e^{-\zeta}}\right)^s + \left(\frac{e^{-\zeta}}{1+e^{-\zeta}}\right)^s$ if $\zeta \in \mathbb{R}$ is the true parameter, which clearly converges to 0 if s tends to infinity.

Theorem 2.2.4
Let $g : \Theta \to \Gamma \subseteq \mathbb{R}^k$ be a bijective mapping, $\tilde{P}_{\boldsymbol{\gamma}}^{X} = P_{g^{-1}(\boldsymbol{\gamma})}^{X}$, $\boldsymbol{\gamma} \in \Gamma$, and $\tilde{\mathfrak{P}}^{X} = \{\tilde{P}_{\boldsymbol{\gamma}}^{X} : \boldsymbol{\gamma} \in \Gamma\}$ be another parametrization of $\mathfrak{P}^X$. Then, based on s independent replicates $X^{(1)}, \ldots, X^{(s)}$ of X, $\boldsymbol{\zeta}^{*(s)}$ is the MLE of $\boldsymbol{\zeta}$ in Θ if and only if $\boldsymbol{\gamma}^{*(s)} = g(\boldsymbol{\zeta}^{*(s)})$ is the MLE of $\boldsymbol{\gamma} = g(\boldsymbol{\zeta})$ in Γ.

Proof. In Witting (1985), Thm. 1.31, pp. 32/33. ■

2.2.2 Estimation under Simple Order Restriction

In this subsection, the focus is on estimation of parameter vectors under the simple order restriction on its components. For this, let us assume the following sample situation differing from the one introduced at the beginning of Section 2.2.
Let $\tilde{\mathfrak{P}} = \{\tilde{f}_\vartheta \mu : \vartheta \in \Theta\}$ be a one-parameter exponential family on $(\mathbb{R}^1, \mathbb{B}^1)$, where Θ is an open interval of the real line (not necessarily finite), and the μ-densities $\tilde{f}_\vartheta$ are given by

$$\tilde{f}_\vartheta(x) = e^{\zeta(\vartheta)T(x)-\kappa(\vartheta)}h(x), \quad x \in \mathbb{R}.$$

We assume to have k independent samples from the exponential family, where the j^{th} sample has size n_j, $1 \le j \le k$, and the random variables in every sample are iid. More precisely, let $X_{i;j}$, $1 \le i \le n_j$, $1 \le j \le k$, be independent, where, for fixed $j \in \{1, \dots, k\}$, $X_{i;j} \sim \tilde{f}_{\vartheta_j}$, $1 \le i \le n_j$, for some (unknown) $\vartheta_j \in \Theta$. All random variables are formally defined on the same measurable space $(\Omega, \mathfrak{A})$ with corresponding family $\mathfrak{P} = \{P_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k)' \in \Theta^k\}$ of probability measures on $(\Omega, \mathfrak{A})$. Now, the aim is to estimate the parameters $\vartheta_1, \dots, \vartheta_k$ where, e.g., based on some prior experiment, the simple ordering $\vartheta_1 \le \dots \le \vartheta_k$ of the parameters is assumed. Related results can be found in the books of Barlow et al. (1972) and Robertson et al. (1988), and are cited in what follows.

Definition 2.2.5 (Isotonic function)
A real-valued function $g : I \to \mathbb{R}$ on a finite set I is called isotonic on I if $i \le j$ implies $g(i) \le g(j)$ for $i, j \in I$.

In the above sample situation, let $I = \{1, \dots, k\}$. If the MLE $\hat{\boldsymbol{\vartheta}} = (\hat{\vartheta}_1, \dots \hat{\vartheta}_k)'$ of $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k)'$ exists, it can be identified with the (random) mapping r on I defined by $r(j) = \hat{\vartheta}_j$, $1 \le j \le k$.

Definition 2.2.6 (Isotonic regression)
Let $g : I \to \mathbb{R}$ be a real-valued and $w : I \to \mathbb{R}_+$ be a positive function on a finite set I. An isotonic function g^* on I is called an isotonic regression of g with weights w if and only if

$$\sum_{i \in I} (g(i) - g^*(i))^2 w(i) \le \sum_{i \in I} (g(i) - h(i))^2 w(i)$$

for every isotonic function h on I.

According to that definition, we denote the isotonic regression of the MLE $\hat{\boldsymbol{\vartheta}} \equiv (r(1), \dots, r(k))'$ of $\boldsymbol{\vartheta}$ by $r^* = (r^*(1), \dots, r^*(k))'$. The following lemma gives a formula to calculate the isotonic regression of a function.

Lemma 2.2.7
In the situation of Def. 2.2.6 the isotonic regression g^* of g with repect to the weights w is given by

$$g^*(i) = \max_{\{U \in \mathfrak{U}:\, i \in U\}} \min_{\{L \in \mathfrak{L}:\, i \in L\}} Av(L \cap U),$$

where for $A \subseteq I$

$$Av(A) = \frac{\sum_{i \in A} w(i) g(i)}{\sum_{i \in A} w(i)},$$

and $\mathfrak{U}$ and $\mathfrak{L}$ are the upper sets and lower sets of I. At this, $U \subseteq I$ is called an upper set if $i \in U$ and $i \le j$ implies $j \in U$, and $L \subseteq I$ is a lower set of I if $i \in L$ and $j \le i$ implies $j \in L$, respectively.

Proof. E.g., in Robertson et al. (1988), Thm. 1.4.4, p. 23 or Barlow et al. (1972), Thm. 2.8, p. 80. ■

By choosing $I = \{1, \ldots, k\}$, the upper and lower sets of I are given by $\{m, \ldots, k\}$, $m \in I$, and $\{1, \ldots, m\}$, $m \in I$, respectively. The intersection of an upper and a lower set is then either empty or equals $\{j, \ldots, m\}$ for some $1 \leq j \leq m \leq k$.
Now, in the context of the initial sample situation of this subsection, the following theorem holds true.

Theorem 2.2.8
If the unrestricted MLE $\hat{\boldsymbol{\vartheta}} = (\hat{\vartheta}_1, ..., \hat{\vartheta}_k)'$ of $\boldsymbol{\vartheta} = (\vartheta_1, ..., \vartheta_k)'$ exists and if ζ and κ have continuous second derivatives on Θ with the properties that $\zeta'(\vartheta) > 0$ and $\kappa'(\vartheta) = \vartheta\zeta'(\vartheta)$, $\vartheta \in \Theta$, then the restricted MLE of $\boldsymbol{\vartheta}$, subject to the constraint that the estimator be isotonic on $I = \{1, \ldots, k\}$, is given by the isotonic regression of $\hat{\boldsymbol{\vartheta}}$ with weights $w(j) = n_j$, $1 \leq j \leq k$.

Proof. In Robertson et al. (1988), Thm. 1.5.2, p. 34. ■

Hence, the MLE $\tilde{\vartheta}_j$ of ϑ_j, $1 \leq j \leq k$, under the simple order restriction $\vartheta_1 \leq \cdots \leq \vartheta_k$ is given by

$$\tilde{\vartheta}_j = \max_{1 \leq \mu \leq j} \min_{j \leq \nu \leq k} \frac{\sum_{l=\mu}^{\nu} n_l \hat{\vartheta}_l}{\sum_{l=\mu}^{\nu} n_l}.$$

In particular, in case of equal sample sizes $n_1 = \cdots = n_k$, we obtain

$$\tilde{\vartheta}_j = \max_{1 \leq \mu \leq j} \min_{j \leq \nu \leq k} \frac{1}{\nu - \mu + 1} \sum_{l=\mu}^{\nu} \hat{\vartheta}_l.$$

2.2.3 Efficiency of Estimators

The concept of *efficiency* of an estimator is connected to the multivariate Rao-Cramér inequality (see, e.g., Shao (2003), p. 169 or Witting (1985), Thm. 2.133, p. 317), which is applicable here by noticing that in exponential families the order of integration and differentiation can be interchanged (cf. La. 2.1.14). According to the sample situation introduced at the beginning of Section 2.2, the inequality is as follows.
Let $g: \Theta \to \mathbb{R}^l$ be an l-dimensional differentiable function of the parameter $\boldsymbol{\zeta}$ with Jacobian matrix $\mathbf{D}_g(\boldsymbol{\zeta}) \in \mathbb{R}^{l \times k}$ evaluated at $\boldsymbol{\zeta} \in \Theta$. Then, for every unbiased estimator $\hat{g}^{(s)}$ of g based on s independent observations, we obtain

$$\begin{aligned} \mathbf{Cov}_{\boldsymbol{\zeta}}(\hat{g}^{(s)}) &\geq \mathbf{D}_g(\boldsymbol{\zeta}) \mathbf{I}_f^{(s)}(\boldsymbol{\zeta})^{-1} \mathbf{D}_g(\boldsymbol{\zeta})' \\ &= \mathbf{D}_g(\boldsymbol{\zeta})\, \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}^{(s)})^{-1} \mathbf{D}_g(\boldsymbol{\zeta})', \end{aligned} \tag{2.2.2}$$

for $\boldsymbol{\zeta} \in \Theta$ in the sense of the Löwner ordering, i.e. $\mathbf{A} \geq \mathbf{B}$ if and only if $\mathbf{A} - \mathbf{B} \geq 0$ ($\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix).
In particular, by choosing g as identity on Θ, the covariance matrix of every unbiased estimator $\hat{\boldsymbol{\zeta}}^{(s)}$

of $\boldsymbol{\zeta}$ based on s independent observations can be bounded from below:

$$\begin{aligned}\mathbf{Cov}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}}^{(s)}) &\geq \mathbf{I}_f^{(s)}(\boldsymbol{\zeta})^{-1} = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}^{(s)})^{-1} \\ &= \frac{1}{s}\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T})^{-1}, \quad \boldsymbol{\zeta} \in \Theta.\end{aligned}$$

Thus, for every unbiased estimator $\hat{\zeta}_j^{(s)}$ of ζ_j based on s independent observations, its variance is bounded by

$$Var_{\boldsymbol{\zeta}}(\hat{\zeta}_j^{(s)}) \geq \frac{[\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T})^{-1}]_{j,j}}{s}, \quad 1 \leq j \leq k, \quad \zeta \in \Theta.$$

The multivariate Rao-Cramér inequality provides a lower bound for the covariance matrix of an unbiased estimator. Hence, estimators whose covariance matrices attain that lower bound are of particular interest.

Definition 2.2.9 (Efficiency)
In the above situation, $\hat{g}^{(s)}$ is called an efficient estimator of $g(\boldsymbol{\zeta})$ based on s independent observations if $\hat{g}^{(s)}$ is unbiased for $g(\boldsymbol{\zeta})$ and its covariance matrix attains the lower bound of the Rao-Cramér inequality (2.2.2). In particular, $\hat{\boldsymbol{\zeta}}^{(s)}$ is called an efficient estimator of $\boldsymbol{\zeta}$ based on s indenpendent observations if $\hat{\boldsymbol{\zeta}}^{(s)}$ is unbiased for $\boldsymbol{\zeta}$ and $\mathbf{Cov}_{\boldsymbol{\zeta}}(\hat{\zeta}^{(s)}) = \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}^{(s)})^{-1}$, $\boldsymbol{\zeta} \in \Theta$.

As an example for an efficient estimator, notice the following lemma.

Lemma 2.2.10
$\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ is an efficient estimator of $\pi(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[\boldsymbol{T}]$ based on s independent observations.

Proof. In the Appendix. ■

In the actual context of a finite sample size, the MLE of $\boldsymbol{\zeta}$ may turn out to be non-efficient or even biased. However, from an asymptotical and, thus, theoretical point of view, the efficiency of the estimator can be shown as we will point out in Subsection 2.2.5.

2.2.4 (Strong) Consistency of Estimators

In the following subsections, we continue by deriving *asymptotic* properties of sequences of estimators. At first, *(strong) consistency* of a sequence of estimators is the property that an estimator of an unknown parameter gets 'closer' to the parameter when the sample size increases.

Definition 2.2.11 ((Strong) consistency)
A sequence $\hat{\boldsymbol{\zeta}} = \{\hat{\boldsymbol{\zeta}}^{(s)}\}_{s\in\mathbb{N}}$ of estimators of $\boldsymbol{\zeta} \in \Theta$ is called consistent if $\hat{\boldsymbol{\zeta}}^{(s)} \xrightarrow{P_{\boldsymbol{\zeta}}} \boldsymbol{\zeta}$, i.e., if $\hat{\boldsymbol{\zeta}}^{(s)}$ converges in $P_{\boldsymbol{\zeta}}$-probability towards $\boldsymbol{\zeta}$ as s tends to infinity, for every $\boldsymbol{\zeta} \in \Theta$. If the convergence is even $P_{\boldsymbol{\zeta}}$-a.s., $\hat{\boldsymbol{\zeta}}$ is called strongly consistent.

Evidently, strong consistency implies consistency. Assuming an underlying exponential family structure, strong consistency of the sequence of MLEs is easily seen, provided that the sequence exists.

Lemma 2.2.12
If the sequence $\boldsymbol{\zeta}^* = \{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ of the MLEs of $\boldsymbol{\zeta}$ exists, it is strongly consistent. Moreover, in that case, using the denotations of Thm. 2.2.4, the sequence of MLEs $\{g(\boldsymbol{\zeta}^{*(s)})\}_{s\in\mathbb{N}}$ of $g(\boldsymbol{\zeta})$ is strongly consistent for estimating $g(\boldsymbol{\zeta})$, provided g is continuous on Θ.

Proof. Let $\boldsymbol{\zeta} \in \Theta$ be the true parameter vector. From an insight into the proof of Thm. 2.2.3, $s^{-1}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \to \pi(\boldsymbol{\zeta})$, $P_{\boldsymbol{\zeta}} - a.s.$. π^{-1} is a continuous mapping on $\pi(\Theta)$ and, thus, $\boldsymbol{\zeta}^{*(s)} = \pi^{-1}(s^{-1}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})) \to \pi^{-1}(\pi(\boldsymbol{\zeta})) = \boldsymbol{\zeta}$, $P_{\boldsymbol{\zeta}}$-a.s., since a.s. convergence is preserved under continuous mappings (e.g., in Shao (2003), Thm. 1.10 (i), p. 59). Hence, strong consistency of $\boldsymbol{\zeta}^*$ is proven. The statement for the sequence $\{g(\boldsymbol{\zeta}^{*(s)})\}_{s\in\mathbb{N}}$ is then obvious by application of Thm. 1.10 (i) in Shao (2003), p. 59, once again. ■

2.2.5 Asymptotic Efficiency of Estimators

Similarly to the concept of efficiency of estimators in case of finite sample sizes, a respective terminology can be introduced for the asymptotic approach as well. For a better understanding of the definition and results of this subsection, we give a short heuristics (cf. Lehmann & Casella (1998), pp. 437 ff.).
Let $\hat{\delta} = \{\hat{\delta}^{(s)}\}_{s\in\mathbb{N}}$ be a sequence of unbiased estimators of a real-valued parameter $\delta \in D$, where $\hat{\delta}^{(s)}$ is based on s independent replicates $X^{(1)}, \ldots, X^{(s)}$ of a random element X with values in $(\mathfrak{X}, \mathfrak{B})$ having μ-density f_δ^X, where μ is a σ-finite measure on $(\mathfrak{X}, \mathfrak{B})$. We will assume that $\sqrt{s}(\hat{\delta}^{(s)} - \delta)$ has an asymptotic normal distribution with mean zero and variance $v(\delta)$, i.e.

$$\sqrt{s}(\hat{\delta}^{(s)} - \delta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, v(\delta)), \tag{2.2.3}$$

where $v : D \to \mathbb{R}_+$ maps the parameter δ onto the variance $v(\delta)$ of the corresponding asymptotic distribution. If the respective regularity conditions are fulfilled, the Rao-Cramér inequality yields

$$Var_\delta(\hat{\delta}^{(s)}) \geq I_f^{(s)}(\delta)^{-1} = \frac{I_f(\delta)^{-1}}{s}, \quad \forall s \in \mathbb{N}, \tag{2.2.4}$$

where $I_f(\delta)$ denotes the Fisher information of $\mathfrak{P}^X = \{f_\delta^X \mu : \delta \in D\}$, and $I_f^{(s)}(\delta)$ the Fisher information of the corresponding product probability measures at $\delta \in D$.
Obviously, (2.2.4) is equivalent to

$$Var_\delta(\sqrt{s}(\hat{\delta}^{(s)} - \delta)) \geq I_f(\delta)^{-1}, \quad \forall s \in \mathbb{N}.$$

Let us additionally assume that, for every $\delta \in D$, the actual variances $Var_\delta(\sqrt{s}(\hat{\delta}^{(s)} - \delta))$ converge to the asymptotic variance $v(\delta)$ if the sample size s tends to infinity. Then,

$$v(\delta) = \lim_{s\to\infty} Var_\delta(\sqrt{s}(\hat{\delta}^{(s)} - \delta)) \geq I_f(\delta)^{-1}, \tag{2.2.5}$$

and, hence, the asymptotic variance is bounded from below by the inverse of the Fisher information of $\mathfrak{P}^X$. Clearly, a sequence of estimators of δ is the more attractive, the smaller the values $v(\delta)$, $\delta \in D$,

are, and, thus, if equality holds in (2.2.5), the sequence of estimators can be considered as 'efficient' from an asymptotical point of view.
In view of the assumptions made in the above heuristics, the following points should be noted:

- Estimators of interest, e.g. MLEs, usually fulfill condition (2.2.3), and, thus, the assumption of asymptotic normality with mean zero of competing sequences of estimators is not that restrictive (for another approach see Wolfowitz (1965)).

- In general, the limit of the actual variances and the asymptotic variance do not have to coincide. However, the motivation of asymptotic distribution theory is to replace the actual distribution by the possibly more manageable asymptotic one if the number of observations is sufficiently large. Hence, it seems to be reasonable to assume that the variances of the (normalized) estimators converge to that of the asymptotic distribution. Additionally, one might also assume this convergence to be uniform with respect to $\delta \in D$, since the true value of the parameter is unknown.

- Finally, we remark on the fact that the Rao-Cramér inequality holds for unbiased estimators only. Asymptotic normality with mean zero according to (2.2.3) implies consistency of the sequence of estimators but not necessarily unbiasedness of any estimator $\hat{\delta}^{(s)}$, $s \in \mathbb{N}$.

If certain regularity conditions are fulfilled, in particular, if an underlying exponential family structure is assumed, Bahadur (1964) has shown that for every sequence $\hat{\delta} = \{\hat{\delta}^{(s)}\}_{s \in \mathbb{N}}$ of not necessarily unbiased estimators of $\delta \in D$ satisfying (2.2.3) the inequality

$$v(\delta) \geq I_f(\delta)^{-1}, \tag{2.2.6}$$

holds λ^1-a.e. on D. The values of δ which violate inequality (2.2.6) are called *points of superefficiency*. A point of superefficiency might bring along some unpleasant properties related to the risk functions of the estimators evaluated at sequences of parameters converging to that point (Lehmann & Casella (1998), pp. 441/442; see also the Hodges example, e.g., in Lehmann & Casella (1998), Ex.s 2.5 and 2.7, pp. 440 ff.) and, thus, it seems to be advisable to restrict the class of admissible sequences of estimators to those fulfilling (2.2.6) for all $\delta \in D$.
If the set of probability measures forms a one-parameter exponential family according to (2.1.11) where $k = 1$ and ζ_1 is replaced by δ, $\delta \mapsto I_f(\delta)^{-1}$ is a continuous function on Θ and, thus, e.g., the restriction of the set of sequences of estimators satisfying (2.2.3) to those satisfying (2.2.3), where v is a continuous function of δ, ensures that (2.2.6) holds everywhere on Θ.
The preceding result can be extended to the multiparameter case (see Bahadur (1964)). In the situation discussed above, let $\Theta \subseteq \mathbb{R}^k$, $k \geq 1$. If similar regularity conditions are assumed, we obtain that for every sequence of estimators $\hat{\boldsymbol{\delta}} = \{\hat{\boldsymbol{\delta}}^{(s)}\}_{s \in \mathbb{N}}$ of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)' \in \Theta$, satisfying

$$\sqrt{s}(\hat{\boldsymbol{\delta}}^{(s)} - \boldsymbol{\delta}) \xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\delta})), \tag{2.2.7}$$

the inequality

$$\boldsymbol{\Sigma}(\boldsymbol{\delta}) \geq \mathbf{I}_f(\boldsymbol{\delta})^{-1}$$

holds λ^k-a.e. on Θ, where $\mathbf{I}_f$ denotes the Fisher information matrix of $\mathfrak{P}^X = \{f_{\boldsymbol{\delta}}^X \mu : \boldsymbol{\delta} \in \Theta\}$ and $\geq$ is the Löwner ordering. Here and in the following, $\mathcal{N}_k(\boldsymbol{a}, \mathbf{B})$ denotes the k-dimensional normal distribution with mean $\boldsymbol{a}$ and covariance matrix $\mathbf{B}$.
Having elucidated the theoretical background, we define asymptotic efficiency of a sequence of estimators as follows.

Definition 2.2.13 (Asymptotic efficiency)
A sequence $\hat{\boldsymbol{\zeta}} = \{\hat{\boldsymbol{\zeta}}^{(s)}\}_{s\in\mathbb{N}}$ of estimators of $\boldsymbol{\zeta} \in \Theta$ is said to be asymptotically efficient if, for every $\boldsymbol{\zeta} \in \Theta$, $\sqrt{s}(\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta}) \xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \mathbf{I}_f(\boldsymbol{\zeta})^{-1})$, i.e. if $\sqrt{s}(\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta})$ has a k-dimensional asymptotic normal distribution with mean zero and covariance matrix $\mathbf{I}_f(\boldsymbol{\zeta})^{-1}$, where $\mathbf{I}_f(\boldsymbol{\zeta})$ denotes the Fisher information matrix of $\mathfrak{P}^X$ at $\boldsymbol{\zeta} \in \Theta$.

Remark 2.2.14
As mentioned before, asymptotic normality implies consistency. For this, let $\{\hat{\boldsymbol{\zeta}}^{(s)}\}_{s\in\mathbb{N}}$ fulfill condition (2.2.7) with $\boldsymbol{\delta}$ replaced by $\boldsymbol{\zeta}$. Then

$$\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta} = \underbrace{\frac{1}{\sqrt{s}}}_{\substack{s\to\infty \\ \longrightarrow 0}} \cdot \underbrace{\sqrt{s}(\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta})}_{\xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \Sigma(\boldsymbol{\zeta}))},$$

and the multivariate version of Slutsky's theorem (cf., e.g., Sen & Singer (1993), Thm. 3.4.3, p. 130) yields that $\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta} \xrightarrow{\mathcal{D}} \mathbf{0}$. Applying Thm. 1.8 (vii) in Shao (2003), p. 51, we obtain that $\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta} \xrightarrow{P_{\boldsymbol{\zeta}}} \mathbf{0}$ and, hence, the assertion is shown.

Now, according to the sample situation introduced at the beginning of Section 2.2, the following lemma is true.

Lemma 2.2.15
If the sequence $\boldsymbol{\zeta}^* = \{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ of the MLEs of $\boldsymbol{\zeta}$ exists, it is asymptotically efficient. Moreover, in that case, if in Thm. 2.2.4 g is continuously differentiable with $|\mathbf{D}_g(\boldsymbol{\zeta})| \neq 0$ $\forall \boldsymbol{\zeta} \in \Theta$, then the sequence $\{\boldsymbol{\gamma}^{*(s)}\}_{s\in\mathbb{N}}$, $\boldsymbol{\gamma}^{*(s)} = g(\boldsymbol{\zeta}^{*(s)})$, $s \in \mathbb{N}$, is asymptotically efficient for estimating $\boldsymbol{\gamma} = g(\boldsymbol{\zeta})$, i.e.

$$\sqrt{s}(\boldsymbol{\gamma}^{*(s)} - \boldsymbol{\gamma}) \xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \tilde{\mathbf{I}}_f(\boldsymbol{\gamma})^{-1}),$$

where $\tilde{\mathbf{I}}_f(\boldsymbol{\gamma})$ denotes the Fisher information matrix of $\tilde{\mathfrak{P}}^X$ at $\boldsymbol{\gamma} \in \Gamma$.

Proof. In the Appendix. ■

Finally, we mention that, within the class of all asymptotically efficient sequences of estimators of $\boldsymbol{\zeta} \in \Theta$, a certain subset, i.e. the set of sequences $\{\hat{\boldsymbol{\zeta}}^{(s)}\}_{s\in\mathbb{N}}$ that fulfill

$$\sqrt{s}(\hat{\boldsymbol{\zeta}}^{(s)} - \boldsymbol{\zeta}) - \frac{1}{\sqrt{s}} \mathbf{I}_f(\boldsymbol{\zeta})^{-1} \boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \xrightarrow{P_{\boldsymbol{\zeta}}} \mathbf{0}, \quad \forall \boldsymbol{\zeta} \in \Theta,$$

are termed best asymptotic normal estimators (BAN estimators). If the sequence $\boldsymbol{\zeta}^*$ of MLEs of $\boldsymbol{\zeta} \in \Theta$ exists, it is a BAN estimator of $\boldsymbol{\zeta}$, which can easily be seen by application of the multivariate mean value theorem (in the Appendix: Proof Δ_1). Several authors (see below) have shown that BAN estimators can be considered as asymptotically optimal estimators from another point of view, when comparisons of competing estimators are made based on *Pitman's measure of closeness*. At this, an estimator $\hat{\delta}$ of a real-valued parameter $\delta \in D$ is said to be a *Pitman closer estimator* than $\tilde{\delta}$ if

$$P_\delta(|\hat{\delta} - \delta| < |\tilde{\delta} - \delta|) \geq 0.5 \quad \forall \delta \in D, \tag{2.2.8}$$

with a strict inequality sign for at least one $\delta \in D$. $\hat{\delta}$ is termed a *Pitman closest estimator within the class* $\mathfrak{D}$ if $\hat{\delta}$ is Pitman closer than every other estimator in $\mathfrak{D}$. These concepts can both be extended to the multivariate case, where the absolute value $|\bullet|$ is replaced by some norm on $\mathbb{R}^k$, e.g., the Mahalanobis norm involving the inverse of the Fisher information matrix (see, e.g., Rencher (1998), pp. 22/23), and, moreover, respective definitions of Pitman's measure of closeness related to asymptotic theory and sequences of estimators have been made. For a comprehensive treatment of Pitman's measure of closeness, we refer to Keating et al. (1993), where asymptotic optimality in Pitmans's sense is considered in Thm. 6.1.8, pp. 176/177, for the univariate case. For a respective result in a multivariate framework, see Sen (1986).

2.3 Statistical Tests

Having derived point estimators and their (asymptotic) properties in Section 2.2, we are now concerned with statistical tests on the model parameters $\zeta_1, \dots, \zeta_k$. Throughout this section, if not otherwise specified, we assume the same sample situation and denotations as in the beginning of Section 2.2. Based on the underlying exponential family structure, optimal univariate and multivariate tests are obtained, and, in particular, tests under order restrictions are established.

2.3.1 Uniformly Most Powerful Unbiased Tests

It is well-known that, based on an underlying multivariate exponential family structure, uniformly most powerful unbiased (UMPU) one- and two-sided tests on a single parameter ζ_j, $1 \le j \le k$, can be established. We consider the five following test problems involving ζ_1. Regarding the structure of $f_{\boldsymbol{\zeta}}^{(s)}$, respective modifications of the statements of this subsection are also valid for every other choice of ζ_j, $2 \le j \le k$.

(I) $H_0: \ \zeta_1 \le \zeta_0 \quad \leftrightarrow \quad H_1: \ \zeta_1 > \zeta_0,$

(II) $H_0: \ \zeta_1 \ge \zeta_0 \quad \leftrightarrow \quad H_1: \ \zeta_1 < \zeta_0,$

(III) $H_0: \ \zeta_1 = \zeta_0 \quad \leftrightarrow \quad H_1: \ \zeta_1 \neq \zeta_0,$

(IV) $H_0: \ \zeta_0^{(1)} \le \zeta_1 \le \zeta_0^{(2)} \quad \leftrightarrow \quad H_1: \ \zeta_1 < \zeta_0^{(1)}$ or $\zeta_1 > \zeta_0^{(2)},$

(V) $H_0: \zeta_1 \le \zeta_0^{(1)}$ or $\zeta_1 \ge \zeta_0^{(2)} \quad \leftrightarrow \quad H_1: \ \zeta_0^{(1)} < \zeta_1 < \zeta_0^{(2)},$

where $\zeta_0, \zeta_0^{(1)}, \zeta_0^{(2)} \in \tilde{\Theta} = \{\zeta_1 \in \mathbb{R}: \ (\zeta_1, \dots, \zeta_k)' \in \Theta\}$ with $\zeta_0^{(1)} < \zeta_0^{(2)}$.
(I)-(III) are the typical one- and two-sided test problems, whereas (IV) and (V), corresponding to the question whether the true parameter lies in a given interval or not, do not as frequently appear in literature. The latter tests are known as *tests of equivalence* and a survey on that topic can be found in Wellek (2003).
As a general result concerning multivariate exponential families, in all of the cases (I)-(V), UMPU level-α tests can be established, where a level-α test φ is called *unbiased* if its power is bounded from below by α, i.e., $E_{\boldsymbol{\zeta}}[\varphi] \ge \alpha$, if the alternative is true.
For the sake of brevity, all theorems of this subsection are given for a single observation ($s = 1$). From La. 2.1.25, it is obvious that respective assertions are true if we consider a sample of size $s > 1$ and replace T_1 by $T_1^{(s)}$, $\boldsymbol{T}$ by $\boldsymbol{T}^{(s)}$ and so on. Moreover, throughout the subsection, for a better reading, $P_{\boldsymbol{\zeta}}^X$ is replaced by $P_{\boldsymbol{\zeta}}$.
Introducing, additionally, the statistic $\tilde{\boldsymbol{T}} = (T_2, \dots, T_k)'$ with corresponding observation $\tilde{\boldsymbol{t}} = (t_2, \dots, t_k)'$, the theorems are as follows.

Theorem 2.3.1
For the one-sided test problem (I), $\varphi^* = \Psi^* \circ (T_1, \tilde{\boldsymbol{T}})$ *defined by*

$$\Psi^*(t_1, \tilde{\boldsymbol{t}}) = \mathbb{1}_{(c(\tilde{\boldsymbol{t}}), \infty)}(t_1) + \gamma(\tilde{\boldsymbol{t}}) \mathbb{1}_{\{c(\tilde{\boldsymbol{t}})\}}(t_1),$$

where $c, \gamma : (\mathbb{R}^{k-1}, \mathbb{B}^{k-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$ *with* $0 \leq \gamma \leq 1$ *solve the equations*

$$P_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}((c(\tilde{\boldsymbol{t}}), \infty)) + \gamma(\tilde{\boldsymbol{t}}) P_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}(\{c(\tilde{\boldsymbol{t}})\}) = \alpha, \quad \tilde{\boldsymbol{t}} \in \mathbb{R}^{k-1},$$

is a UMPU level-α *test.*

Proof. In Witting (1985), Thm. 3.60, p. 376, or Shao (2003), Thm. 6.4 (i), pp. 406/407. ■

At this, since the conditional distribution of T_1 given $\tilde{\boldsymbol{T}} = \tilde{\boldsymbol{t}}$ only depends on ζ_0 if $\boldsymbol{\zeta} = (\zeta_0, \zeta_2, \ldots, \zeta_k)'$ is true (cf., e.g., Lehmann & Romano (2005), La. 2.7.2, p. 48), we omit $\zeta_2, \ldots, \zeta_k$ and use, for short, the expression $P_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}$.

Theorem 2.3.2
For the one-sided test problem (II), $\varphi^* = \Psi^* \circ (T_1, \tilde{\boldsymbol{T}})$ *defined by*

$$\Psi^*(t_1, \tilde{\boldsymbol{t}}) = \mathbb{1}_{(-\infty, c(\tilde{\boldsymbol{t}}))}(t_1) + \gamma(\tilde{\boldsymbol{t}}) \mathbb{1}_{\{c(\tilde{\boldsymbol{t}})\}}(t_1), \tag{2.3.1}$$

where $c, \gamma : (\mathbb{R}^{k-1}, \mathbb{B}^{k-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$ *with* $0 \leq \gamma \leq 1$ *solve the equations*

$$P_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}(-\infty, c(\tilde{\boldsymbol{t}})) + \gamma(\tilde{\boldsymbol{t}}) P_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}(\{c(\tilde{\boldsymbol{t}})\}) = \alpha, \quad \tilde{\boldsymbol{t}} \in \mathbb{R}^{k-1}, \tag{2.3.2}$$

is a UMPU level-α *test.*

Proof. In the Appendix. ■

We proceed by considering the two-sided test problems (III)-(V).

Theorem 2.3.3
For the two-sided test problem (III), $\varphi^* = \Psi^* \circ (T_1, \tilde{\boldsymbol{T}})$ *defined by*

$$\Psi^*(t_1, \tilde{\boldsymbol{t}}) = \mathbb{1}_{(-\infty, c_1(\tilde{\boldsymbol{t}}))}(t_1) + \mathbb{1}_{(c_2(\tilde{\boldsymbol{t}}), \infty)}(t_1) + \sum_{i=1}^{2} \gamma_i(\tilde{\boldsymbol{t}}) \mathbb{1}_{\{c_i(\tilde{\boldsymbol{t}})\}}(t_1), \tag{2.3.3}$$

where $c_1, c_2, \gamma_1, \gamma_2 : (\mathbb{R}^{k-1}, \mathbb{B}^{k-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$ *with* $0 \leq \gamma_1, \gamma_2 \leq 1$ *solve the equations*

$$\begin{aligned} &\int \Psi^*(t_1, \tilde{\boldsymbol{t}}) dP_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}(t_1) = \alpha \\ \textit{and} \quad &\int t_1 \, \Psi^*(t_1, \tilde{\boldsymbol{t}}) dP_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}(t_1) = \alpha \int t_1 \, dP_{\zeta_0,\bullet}^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}(t_1), \quad \tilde{\boldsymbol{t}} \in \mathbb{R}^{k-1}, \end{aligned}$$

is a UMPU level-α *test.*

Proof. In Witting (1985), Thm. 3.62, pp. 377/378, or Shao (2003), Thm. 6.4 (iv), pp. 406/407. ■

Theorem 2.3.4
For the two-sided test problem (IV), $\varphi^* = \Psi^* \circ (T_1, \tilde{\boldsymbol{T}})$ with Ψ^* as defined in (2.3.3), where $c_1, c_2, \gamma_1, \gamma_2 : (\mathbb{R}^{k-1}, \mathbb{B}^{k-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$ with $0 \leq \gamma_1, \gamma_2 \leq 1$ solve the equations

$$\int \Psi^*(t_1, \tilde{\boldsymbol{t}}) dP^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}_{\zeta_0^{(1)},\bullet}(t_1) = \int \Psi^*(t_1, \tilde{\boldsymbol{t}}) dP^{T_1|\tilde{\boldsymbol{T}}=\tilde{\boldsymbol{t}}}_{\zeta_0^{(2)},\bullet}(t_1) = \alpha, \quad \tilde{\boldsymbol{t}} \in \mathbb{R}^{k-1}, \tag{2.3.4}$$

is a UMPU level-α test.

Proof. In Shao (2003), Thm. 6.4, pp. 406/407. ■

Theorem 2.3.5
For the two-sided test problem (V), $\varphi^* = \Psi^* \circ (T_1, \tilde{\boldsymbol{T}})$ defined by

$$\Psi^*(t_1, \tilde{\boldsymbol{t}}) = \mathbb{1}_{(c_1(\tilde{\boldsymbol{t}}), c_2(\tilde{\boldsymbol{t}}))}(t_1) + \sum_{i=1}^{2} \gamma_i(\tilde{\boldsymbol{t}}) \mathbb{1}_{\{c_i(\tilde{\boldsymbol{t}})\}}(t_1),$$

where $c_1, c_2, \gamma_1, \gamma_2 : (\mathbb{R}^{k-1}, \mathbb{B}^{k-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$ with $0 \leq \gamma_1, \gamma_2 \leq 1$ solve the equations (2.3.4), is a UMPU level-α test.

Proof. In Shao (2003), Thm. 6.4, pp. 406/407. ■

Finally, we remark on an important additional result.

Remark 2.3.6
If $k = 1$, i.e. if $\mathfrak{P}$ is a one-parameter exponential family, the (unconditional) UMPU tests of Thm.s 2.3.1, 2.3.2 and 2.3.5 are even uniformly most powerful (UMP) tests (see, e.g., in Shao (2003), Thm. 6.2, p. 399, and Thm. 6.3, pp. 401/402).

2.3.2 Likelihood Ratio Test

In the preceding subsection, based on the fact that the conditional distribution of T_1 given $(T_2, \ldots, T_k)' = (t_2, \ldots, t_k)'$ forms a one-parameter exponential family in ζ_1 and the identity (see, e.g., Lehmann & Romano (2005), La. 2.7.2 (ii), p. 48), and, thus, has *monotone likelihood ratio*, UMPU tests on the single parameter ζ_1 have been established. However, in case of testing hypotheses involving more than one of the ζ's, mathematical difficulties arise due to the fact that the likelihood ratio then depends on at least two parameters. Usually, in that case, UMPU tests for the considered test problems do not exist. Hence, there is need for other procedures that lead to reasonable tests and test statistics, respectively.
Meanwhile, many multivariate test principles have been developed, and properties of the related tests and test statistics, in particular concerning asymptotic distribution theory, have been examined as well. Among them, the probably best-known is the likelihood ratio test that was introduced in Neyman & Pearson (1928).

Definition 2.3.7 (Likelihood ratio test)
Let $\Theta_0 \subset \Theta$ and $\Theta_1 = \Theta \setminus \Theta_0$. We consider the test problem $H_0 : \boldsymbol{\zeta} \in \Theta_0$ against $H_1 : \boldsymbol{\zeta} \in \Theta_1$. If the MLE of $\boldsymbol{\zeta}$ in Θ and the restricted MLE of $\boldsymbol{\zeta}$ in Θ_0 based on $\tilde{\boldsymbol{X}}^{(s)} = (X^{(1)}, \ldots, X^{(s)})$ both exist, the level-α likelihood ratio test (LR test) is defined as

$$\varphi_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(d_s(\alpha),\infty)}(T_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})) + \delta_s(\alpha)\mathbb{1}_{\{d_s(\alpha)\}}(T_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})), \quad \tilde{\boldsymbol{x}}^{(s)} \in \mathfrak{X}^{1\times s},$$

where

$$T_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = -s^{-1}\ln\left(\frac{\sup_{\boldsymbol{\zeta}^{(0)}\in\Theta_0}\prod_{i=1}^{s} f_{\boldsymbol{\zeta}^{(0)}}(x^{(i)})}{\sup_{\boldsymbol{\zeta}\in\Theta}\prod_{i=1}^{s} f_{\boldsymbol{\zeta}}(x^{(i)})}\right), \quad \tilde{\boldsymbol{x}}^{(s)} \in \mathfrak{X}^{1\times s}, \tag{2.3.5}$$

and the constants $d_s(\alpha)$ and $\delta_s(\alpha)$ are such that

$$\sup_{\boldsymbol{\zeta}^{(0)}\in\Theta_0} E_{\boldsymbol{\zeta}^{(0)}}[\varphi_{LR}^{(s)}] = \alpha.$$

At first glance, the LR test seems to be a 'reasonable' test. The smaller the 'probability' $\prod_{i=1}^{s} f_{\boldsymbol{\zeta}}(x^{(i)})$ of observing $\tilde{\boldsymbol{x}}^{(s)}$ is if $\boldsymbol{\zeta} \in \Theta_0$ is assumed, the larger is the value of the test statistic $T_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})$ and, thus, the more likely H_0 will be rejected. Indeed, this intuition is right, as we will see below, where several useful properties of the LR test are stated.
Without further mentioning, we will assume throughout this section that both, the sequence of MLEs and the sequence of restricted MLEs, exist. Applying Thm. 2.2.3, the test statistic $T_{LR}^{(s)}$ of the LR test can be rewritten in the following way.

Lemma 2.3.8
$T_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ is the (random) Kullback-Leibler distance of the MLE $\boldsymbol{\zeta}^{*(s)}$ of $\boldsymbol{\zeta}$ based on $X^{(1)}, \ldots, X^{(s)}$ and Θ_0, i.e.,

$$T_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = d_{KL}(\boldsymbol{\zeta}^{*(s)}, \Theta_0),$$

where $d_{KL}(\boldsymbol{\zeta}, \Theta_0) = \inf_{\boldsymbol{\zeta}^{(0)}\in\Theta_0} d_{KL}(\boldsymbol{\zeta}, \boldsymbol{\zeta}^{(0)})$ for $\boldsymbol{\zeta} \in \Theta$.

Proof. In the Appendix. ∎

Hence, for $\tilde{\boldsymbol{X}}^{(s)} = (X^{(1)}, \ldots, X^{(s)})$,

$$\varphi_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = \mathbb{1}_{(d_s(\alpha),\infty)}(d_{KL}(\boldsymbol{\zeta}^{*(s)}, \Theta_0)) + \delta_s(\alpha)\mathbb{1}_{\{d_s(\alpha)\}}(d_{KL}(\boldsymbol{\zeta}^{*(s)}, \Theta_0)). \tag{2.3.6}$$

In many applications, the case of a simple null hypothesis, i.e. $\Theta_0 = \{\boldsymbol{\zeta}^{(0)}\}$ for some $\boldsymbol{\zeta}^{(0)} \in \Theta$, is of particular interest. The corresponding test problem is

$$H_0 : \boldsymbol{\zeta} = \boldsymbol{\zeta}^{(0)} \quad \leftrightarrow \quad H_1 : \boldsymbol{\zeta} \in \Theta \setminus \{\boldsymbol{\zeta}^{(0)}\} \tag{2.3.7}$$

and (2.3.6) simplifies to

$$\varphi_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = \mathbb{1}_{(d_s(\alpha),\infty)}(d_{KL}(\boldsymbol{\zeta}^{*(s)}, \boldsymbol{\zeta}^{(0)})) + \delta_s(\alpha)\mathbb{1}_{\{d_s(\alpha)\}}(d_{KL}(\boldsymbol{\zeta}^{*(s)}, \boldsymbol{\zeta}^{(0)})). \tag{2.3.8}$$

In a finite sample set-up, properties of the LR test are difficult to derive since the distribution of $T_{LR}^{(s)}$ can usually not be computed analytically. As a consequence, critical values of the LR test are frequently obtained numerically or are replaced by the critical values of the asymptotic distribution if the sample size is sufficiently large. For the latter case, notice the following well-known theorem.

Theorem 2.3.9
Let $\boldsymbol{\zeta}^{(0)} \in \Theta$ be fixed and consider the test problem (2.3.7) with a simple null hypothesis. Then, the following assertions hold true:

(i) $2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *is asymptotically* $\chi^2(k)$*-distributed if* H_0 *is true, where* $\chi^2(k)$ *denotes the chi-square distribution with* k *degrees of freedom.*

(ii) The sequence $\{\tilde{\varphi}_{LR}^{(s)}\}_{s\in\mathbb{N}}$ *of tests defined by* $\tilde{\varphi}_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(\chi^2_{1-\alpha}(k),\infty)}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$, $s \in \mathbb{N}$, *has asymptotic level* α *and is consistent, i.e., if the alternative is true,* $E_{\boldsymbol{\zeta}}[\tilde{\varphi}_{LR}^{(s)}] \to 1$ *as* s *tends to infinity. At this,* $\chi^2_{1-\alpha}(k)$ *denotes the* $(1-\alpha)$*-quantile of the* $\chi^2(k)$ *distribution.*

Proof. E.g., in Wilks (1962), pp. 408-419. A proof of the first result can also be found in Serfling (1980), pp. 155/156, Sen & Singer (1993), Thm. 5.6.1, p. 236, and the Appendix. ■

Thm. 2.3.9 can be extended to the case where under the null hypothesis only m, $1 \leq m \leq k-1$, of the ζ's are assumed to equal pre-fixed values and no restrictions are imposed on the remaining parameters. Then, the theorem remains true replacing the number k of degrees of freedom by m (e.g., in Wilks (1962), pp. 419-422).
Another interesting question in practise is whether all ζ's are equal, i.e. $\zeta_1 = \cdots = \zeta_k$. The case, where the null hypothesis is given by $\zeta_1 = \cdots = \zeta_r = \zeta^{(0)}$ for some fixed real number $\zeta^{(0)}$ is covered in Thm. 2.3.9 by setting $\boldsymbol{\zeta}^{(0)} = (\zeta^{(0)}, \ldots, \zeta^{(0)})'$. However, if there is no pre-fixed value the parameters are compared with, i.e. if $H_0: \zeta_1 = \cdots = \zeta_r$, some more investigations are necessary (see, e.g., Sen & Singer (1993), pp. 239 ff.).
Let $q \leq k$ and $h: \Theta \subseteq \mathbb{R}^k \to \mathbb{R}^q$ a vector-valued function with existing Jacobian matrix $\mathbf{D}_h(\boldsymbol{\zeta}) \in \mathbb{R}^{q\times k}$ of rank q at every $\boldsymbol{\zeta} \in \Theta$. We consider the test problem

$$H_0: h(\boldsymbol{\zeta}) = \mathbf{0} \quad \leftrightarrow \quad H_1: h(\boldsymbol{\zeta}) \neq \mathbf{0}, \tag{2.3.9}$$

and define $\Theta_0 = \{\boldsymbol{\zeta} \in \Theta : h(\boldsymbol{\zeta}) = \mathbf{0}\}$. Furthermore, let $g: \Gamma \subseteq \mathbb{R}^{k-q} \to \Theta_0$ be bijective with existing Jacobian matrix $\mathbf{D}_g(\gamma) \in \mathbb{R}^{k\times k-q}$ of rank $k-q$ at $\gamma \in \Gamma$. Then, (2.3.9) is equivalent to

$$H_0: \boldsymbol{\zeta} \in g(\Gamma) \quad \leftrightarrow \quad H_1: \boldsymbol{\zeta} \notin g(\Gamma). \tag{2.3.10}$$

We illustrate the relationship of h and g by means of an example.

Example 2.3.10
Let $\Theta = \mathbb{R}^k_+$, $q = k-1$ *and* $h: \mathbb{R}^k_+ \to \mathbb{R}^{k-1}$ *defined by* $h(\boldsymbol{\zeta}) = (\zeta_2 - \zeta_1, \ldots, \zeta_k - \zeta_{k-1})'$, $\boldsymbol{\zeta} \in \mathbb{R}^k_+$. *Then,* $H_0: \zeta_1 = \cdots = \zeta_k$, $\Theta_0 = \{\boldsymbol{\zeta} \in \mathbb{R}^k_+ : \zeta_1 = \cdots = \zeta_k\}$ *and a corresponding function* $g: \Gamma = \mathbb{R}_+ \to \Theta_0$ *is given by* $g(\zeta) = (\zeta, \ldots, \zeta)'$, $\zeta > 0$.

Now, the following theorem holds true.

Theorem 2.3.11
Let test problem (2.3.9) or (2.3.10), respectively, be given. Then, $2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *is asymptotically* $\chi^2(q)$*-distributed if* H_0 *is true.*

Proof. In Serfling (1980), pp. 156-160 (cf. Sen & Singer (1993), Thm. 5.6.3, p. 240). ∎

2.3.3 Wald's Test and Rao's Score Test

Related to test problems concerning several parameters, the LR test is one of the best-known tests in multivariate statistics, and it is frequently used in applications. In fact, as we will see in Subsections 2.3.6 and 2.3.7, if the sample size is sufficiently large, the usage of the LR test is encouraged by certain asymptotic optimality properties provided that the set of underlying distributions forms an exponential family. However, we shall introduce at least two other well-known test statistics as alternatives to the LR test statistic, i.e. Wald's statistic introduced in Wald (1943) and Rao's score statistic first mentioned in Rao (1948). In literature, there are many articles dealing with properties of the LR test, Wald's test and Rao's score test, and comparisons have been drawn between these three tests as well. For a survey containing many references on that topic, we refer to Rao (2005).
For test problem (2.3.7) with a simple null hypothesis, we define Wald's statistic and Rao's score statistic as follows (cf. Sen & Singer (1993), pp. 235/236).

- Wald's statistic $T_W = \{T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\}_{s\in\mathbb{N}}$:

$$T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = (\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)})'\mathbf{I}_f^{(s)}(\boldsymbol{\zeta}^{*(s)})(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)}) \tag{2.3.11}$$
$$= s(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)})'\mathbf{I}_f(\boldsymbol{\zeta}^{*(s)})(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)}), \tag{2.3.12}$$

  provided that the MLE $\boldsymbol{\zeta}^{*(s)}$ of $\boldsymbol{\zeta}$ based on a sample of size s exists, and

- Rao's score statistic $T_R = \{T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\}_{s\in\mathbb{N}}$:

$$T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = (\boldsymbol{U}_{\boldsymbol{\zeta}^{(0)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}))'\mathbf{I}_f^{(s)}(\boldsymbol{\zeta}^{(0)})^{-1}\boldsymbol{U}_{\boldsymbol{\zeta}^{(0)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \tag{2.3.13}$$
$$= s^{-1}(\boldsymbol{U}_{\boldsymbol{\zeta}^{(0)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}))'\mathbf{I}_f(\boldsymbol{\zeta}^{(0)})^{-1}\boldsymbol{U}_{\boldsymbol{\zeta}^{(0)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}), \tag{2.3.14}$$

  where $\boldsymbol{U}_{\boldsymbol{\zeta}^{(0)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ denotes the score statistic of $\tilde{\boldsymbol{X}}^{(s)}$ at $\boldsymbol{\zeta}^{(0)}$.

If in the above definition (2.3.11) of Wald's statistic, $\mathbf{I}_f(\boldsymbol{\zeta}^{*(s)})$ is replaced by $\mathbf{I}_f(\boldsymbol{\zeta}^{(0)})$, we will refer to Wald's *modified* statistic $T_{\tilde{W}} = \{T_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\}_{s\in\mathbb{N}}$ (cf. Sen & Singer (1993), p. 235).
In general, in the context of a finite sample size s, neither the LR test, nor Wald's test or Rao's score test dominates the other ones uniformly in terms of their power functions. Moreover, as it is the case for the LR test, the analytical derivation of the distributions of the test statistics $T_W^{(s)}$ and $T_R^{(s)}$ under H_0 is difficult or almost impossible. Again, criticical values of the tests are empirically obtained from simulations or from the respective quantiles of the asymptotic distributions if the sample size is sufficiently large.

Theorem 2.3.12
In case of test problem (2.3.7), $T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *and* $T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *are asymptotically* $\chi^2(k)$*-distributed if* H_0 *is true.*

Proof. E.g., in Serfling (1980), pp. 155/156, or in Sen & Singer (1993), Thm. 5.6.1, p. 236. ■

Additionally, we define Wald's statistic and Rao's score statistic in case of test problem (2.3.9) and (2.3.10), respectively, where the null hypothesis is composite (cf. Sen & Singer (1993), pp. 239/240).

- Wald's statistic $T_W = \{T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\}_{s\in\mathbb{N}}$:

$$T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = sh(\boldsymbol{\zeta}^{*(s)})'\tilde{\mathbf{I}}(\boldsymbol{\zeta}^{*(s)})^{-1}h(\boldsymbol{\zeta}^{*(s)}), \tag{2.3.15}$$

  provided that the MLE $\boldsymbol{\zeta}^{*(s)}$ of $\boldsymbol{\zeta}$ based on a sample of size s exists, where

$$\tilde{\mathbf{I}}(\boldsymbol{\zeta}) = \mathbf{D}_h(\boldsymbol{\zeta})\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\mathbf{D}_h(\boldsymbol{\zeta})', \tag{2.3.16}$$

  and $\mathbf{D}_h(\boldsymbol{\zeta})$ denotes the Jacobian matrix of h at $\boldsymbol{\zeta} \in \Theta$,

- Rao's score statistic $T_R = \{T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\}_{s\in\mathbb{N}}$:

$$T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = s^{-1}(\boldsymbol{U}_{\tilde{\boldsymbol{\zeta}}^{(s)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}))'\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{-1}\boldsymbol{U}_{\tilde{\boldsymbol{\zeta}}^{(s)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}), \tag{2.3.17}$$

  where $\boldsymbol{U}_{\tilde{\boldsymbol{\zeta}}^{(s)}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ denotes the score statistic of $\tilde{\boldsymbol{X}}^{(s)}$ at $\tilde{\boldsymbol{\zeta}}^{(s)}$ and $\tilde{\boldsymbol{\zeta}}^{(s)}$ is the MLE of $\boldsymbol{\zeta}$ based on a sample of size s under the restriction $h(\boldsymbol{\zeta}) = 0$ (under Θ_0).

As in the case of a simple null hypothesis, the asymptotic distribution of the LR test statistic, Wald's statistic and Rao's score statistic is the same if H_0 is true.

Theorem 2.3.13
In case of test problem (2.3.9) and (2.3.10), respectively, $T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *and* $T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *are asymptotically* $\chi^2(q)$*-distributed if* H_0 *is true.*

Proof. E.g., in Serfling (1980), pp. 156-160, or in Sen & Singer (1993), Thm. 5.6.3., p. 240. ■

2.3.4 Tests under Simple Order Restriction

For a simple null hypothesis and different composite null hypotheses, we have already proposed multivariate test statistics and discussed their asymptotic behaviour. However, in applications, questions may arise concerning the simple ordering of the unknown parameters. The corresponding test problems are not considered in this section so far and are subject matter of what follows.
We consider the same sample situation as in Subsection 2.2.2, where n_j is replaced by $n_j^{(s)}$, $1 \le j \le k$, and $s = \sum_{j=1}^{k} n_j^{(s)}$ denotes the total number of observations. Then, for $s \in \mathbb{N}$, k independent samples from the exponential family $\tilde{\mathfrak{P}}$ are assumed, where the j^{th} sample has size $n_j^{(s)}$, $1 \le j \le k$,

and the random variables $X_{i;j} \overset{iid}{\sim} \tilde{f}_{\vartheta_j}$, $1 \leq i \leq n_j^{(s)}$, in each sample are iid. Additionally, let $\tilde{\boldsymbol{X}}^{(s)} = (X_{1;1}, \ldots, X_{n_1^{(s)};1}, \ldots, X_{1;k}, \ldots, X_{n_k^{(s)};k})$, $s \in \mathbb{N}$.
Introducing the parameter sets

$$\Theta_{=} = \{\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_k)' \in \Theta^k : \vartheta_1 = \cdots = \vartheta_k\}$$

and

$$\Theta_{\leq} = \{\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_k)' \in \Theta^k : \vartheta_1 \leq \cdots \leq \vartheta_k\},$$

we consider the test problems

$$H_0 : \boldsymbol{\vartheta} \in \Theta_{=} \quad \leftrightarrow \quad H_1 : \boldsymbol{\vartheta} \in \Theta_{\leq} \setminus \Theta_{=} \tag{2.3.18}$$

and

$$H_0 : \boldsymbol{\vartheta} \in \Theta_{\leq} \quad \leftrightarrow \quad H_1 : \boldsymbol{\vartheta} \in \Theta \setminus \Theta_{\leq}. \tag{2.3.19}$$

Then, Robertson et al. (1988) derived the following asymptotic properties of the sequence of LR tests.

Theorem 2.3.14
Let $\boldsymbol{\zeta}$ and κ have continuous second derivatives on Θ with the properties that $\boldsymbol{\zeta}'(\vartheta) > 0$ and $\kappa'(\vartheta) = \vartheta\boldsymbol{\zeta}'(\vartheta)$, $\vartheta \in \Theta$, and let $\frac{n_j^{(s)}}{s} \to a_j \in (0,1)$, as s tends to infinity, $1 \leq j \leq k$.

(i) Related to test problem (2.3.18), for every $\boldsymbol{\vartheta} \in \Theta_{=}$ and $c \in \mathbb{R}$, we have

$$\lim_{s\to\infty} P_{\boldsymbol{\vartheta}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) = \sum_{j=1}^{k} \frac{|S_k^j|}{k!} P(\chi^2(j-1) \geq c).$$

(ii) Related to test problem (2.3.19), for every $\boldsymbol{\vartheta} \in \Theta_{\leq}$, $\boldsymbol{\eta} \in \Theta_{=}$ and $c \in \mathbb{R}$, we have

$$\begin{aligned} \lim_{s\to\infty} P_{\boldsymbol{\vartheta}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) &\leq \lim_{s\to\infty} P_{\boldsymbol{\eta}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) \\ &= \sum_{j=1}^{k} \frac{|S_k^j|}{k!} P(\chi^2(k-j) \geq c). \end{aligned}$$

At this, S_k^j are the Stirling numbers of the first kind (cf. Abramowitz & Stegun (1965), p. 824), and $\chi^2(0) = 0$.

Proof. In Robertson et al. (1988), Thm. 4.1.1, p. 164, in combination with Cor. B, p. 82. ■

2.3.5 (Asymptotic) Relative Efficiency of Tests

Different approaches have been proposed in order to compare tests and test statistics in the multivariate case, i.e. in situations where more than a single parameter is on test. Most of them are connected

to the concept of *relative efficiency* of test sequences. We will explain this terminology in some more detail (see Nikitin (1995), pp. 1 ff.).
We assume the sample situation in the beginning of Section 2.2 with $\mathfrak{X} \in \mathbb{B}^k$ and $\mathfrak{B} = \mathfrak{X} \cap \mathbb{B}^k$, and consider an infinite number of iid random vectors $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots$ with values in $(\mathfrak{X}, \mathfrak{B})$ having distribution $P_{\boldsymbol{\zeta}}^{\boldsymbol{X}}$ for some $\boldsymbol{\zeta} \in \Theta$. Additionally, w.l.o.g., let $\boldsymbol{T} = id_{\mathfrak{X}}$, (otherwise switch to the exponential family $\mathfrak{P}^{\boldsymbol{T}}$), and, for simplicity, let $\mathfrak{X}$ be an open convex set in $\mathbb{R}^k$ and all appearing random vectors and statistics be continuously distributed. We are interested in testing $H_0 : \boldsymbol{\zeta} \in \Theta_0$ against $H_1 : \boldsymbol{\zeta} \in \Theta_1 = \Theta \setminus \Theta_0$. Let $\varphi = \{\varphi^{(s)}\}_{s\in\mathbb{N}}$ be a sequence of tests with corresponding sequence of test statistics $V = \{V^{(s)}\}_{s\in\mathbb{N}}$, where $V^{(s)} = V^{(s)}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)})$, $s \in \mathbb{N}$, only depends on $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$. For large values of $V^{(s)}$, say, larger than a real number c, H_0 will be rejected. Then, the level of the test $\varphi^{(s)}$ is given by

$$\sup_{\boldsymbol{\zeta}_0 \in \Theta_0} P_{\boldsymbol{\zeta}_0}(V^{(s)} > c)$$

and its power function is given by

$$\beta(\boldsymbol{\zeta}^{(1)}) = P_{\boldsymbol{\zeta}^{(1)}}(V^{(s)} > c), \quad \boldsymbol{\zeta}^{(1)} \in \Theta_1.$$

Now, for every $s \in \mathbb{N}$, $\beta \in (0,1)$ and $\boldsymbol{\zeta} \in \Theta_1$ let $c_s = c_s(\beta, \boldsymbol{\zeta}^{(1)})$, $s \in \mathbb{N}$, be a real number with the property that

$$P_{\boldsymbol{\zeta}^{(1)}}(V^{(s)} > c_s) = \beta.$$

c_s can be chosen as the $(1-\beta)$-quantile of the distribution $P_{\boldsymbol{\zeta}^{(1)}}^{V^{(s)}}$. By doing this, $\varphi^{(s)}$ with level $\alpha_s(\beta, \boldsymbol{\zeta}^{(1)}) = \sup_{\boldsymbol{\zeta}_0 \in \Theta_0} P_{\boldsymbol{\zeta}_0}(V^{(s)} > c_s)$ has power β at $\boldsymbol{\zeta}^{(1)}$. Now, define for every level $\alpha \in (0,1)$

$$N_\varphi(\alpha, \beta, \boldsymbol{\zeta}^{(1)}) = \inf\{s \in \mathbb{N} : \alpha_m(\beta, \boldsymbol{\zeta}^{(1)}) \leq \alpha \ \ \forall m \geq s\},$$

which is the minimal number of observations such that $\varphi^{(m)}$, $m \geq N_\varphi(\alpha, \beta, \boldsymbol{\zeta}^{(1)})$, has level α (and power β at $\boldsymbol{\zeta}^{(1)}$). Clearly, a test is the more attractive, the smaller this number is.

Definition 2.3.15 (Relative efficiency)
With the assumptions and denotations above, let φ and $\tilde{\varphi}$ be two competing test sequences. The relative efficiency of $\tilde{\varphi}$ with respect to φ is defined as

$$e_{\tilde{\varphi},\varphi}(\alpha, \beta, \boldsymbol{\zeta}^{(1)}) = \frac{N_\varphi(\alpha, \beta, \boldsymbol{\zeta}^{(1)})}{N_{\tilde{\varphi}}(\alpha, \beta, \boldsymbol{\zeta}^{(1)})}. \tag{2.3.20}$$

Depending on whether this ratio is larger or smaller than 1, test $\tilde{\varphi}$ or φ, respectively, might be preferred, because less observations are needed to perform as well as the other test in the above sense.
However, given two sequences of tests, the calculation of the corresponding relative efficiency, which depends on three arguments, is difficult, and, for this reason, one continues by considering limiting values of $e_{\tilde{\varphi},\varphi}(\alpha, \beta, \boldsymbol{\zeta}^{(1)})$, hoping that the resulting calculations might be easier (Nikitin (1995), p. 2).
In order to do this, three possibilities are near at hand:

(a) Decrease the level, i.e. send α to zero (*vanishing levels*).

(b) Increase the power, i.e. send β to one.

(c) Move the alternative to the null hypothesis, i.e. send $\boldsymbol{\zeta}^{(1)}$ to $\boldsymbol{\zeta}_0 \in \partial\Theta_0$ in some topology (*contiguous alternatives*).

Most of the approaches in literature that have been made in order to define a concept of *asymptotic relative efficiency (ARE)* of test sequences can be matched to one or a combination of two of the rudiments (a)-(c). At this, the best-known concepts named after the researchers who have developed them are *Bahadur ARE* (a), *Hodges-Lehmann ARE* (b) and *Pitman ARE* (c). Two essential approaches implying more than one of points (a)-(c) are *Chernoff ARE* and *Kallenberg ARE*, where the latter is also termed *Intermediate ARE*.
An exhaustive treatment of all of the contributions to the different types of ARE would exceed the scope of this work. However, in case of an underlying exponential family structure, it turns out that the LR test is *asymptotically optimal* in the sense of Bahadur and Kallenberg, and, for this reason, we will discuss these two concepts in some more detail. A survey on the topic of ARE in the parametric and nonparametric case, as well as many references concerning the different types of ARE can be found in Nikitin (1995) (see also Serfling (1980), Ch. 10).

2.3.6 Bahadur Asymptotic Relative Efficiency

The approach of Bahadur to define asymptotic efficiency of test sequences is related to the concept of the *exact slope* which built the fundament of the stochastic comparison of test sequences in the original definition of Bahadur ARE (e.g., in Bahadur (1971), p. 26). With the assumptions and denotations of Subsection 2.3.5, we define for $s \in \mathbb{N}$ the random variable

$$L_s(\omega) = \sup_{\boldsymbol{\zeta}_0 \in \Theta_0} P_{\boldsymbol{\zeta}_0}(V^{(s)} > V^{(s)}(\omega)), \quad \omega \in \Omega,$$

which is called the *(random) p-value* of the test $\varphi^{(s)}$. For fixed $\omega \in \Omega$, $L_s(\omega)$ is the level, i.e. the error probability of the first kind, of test $\varphi^{(s)}$ with critical value $V^{(s)}(\omega)$. Hence, it is desirable that $L_s(\omega)$, $\omega \in \Omega$, tends to zero as fast as possible if the sample size s becomes large. Typically, if H_0 is true, L_s is asymptotically uniformly distributed over $(0,1)$, and if $\boldsymbol{\zeta}^{(1)} \in \Theta_1$ obtains, $L_s \to 0$ $P_{\boldsymbol{\zeta}^{(1)}}$-a.s. with an exponential rate depending on $\boldsymbol{\zeta}^{(1)}$, i.e.

$$\lim_{s\to\infty} \frac{\ln L_s}{s} = -\frac{1}{2} c_\varphi(\boldsymbol{\zeta}^{(1)}), \quad P_{\boldsymbol{\zeta}^{(1)}}\text{-a.s.}, \tag{2.3.21}$$

where $c_\varphi(\boldsymbol{\zeta}^{(1)}) > 0$ is a positive constant depending on $\boldsymbol{\zeta}^{(1)}$ (Bahadur (1971), p. 26). In that case, i.e. if the limit exists, $c_\varphi(\boldsymbol{\zeta}^{(1)})$ is called the *(strong) Bahadur exact slope of φ at $\boldsymbol{\zeta}^{(1)}$*, and, since (2.3.21) implies that for fixed $\beta \in (0,1)$ and for every $\boldsymbol{\zeta}^{(1)} \in \Theta_1$

$$N_\varphi(\alpha, \beta, \boldsymbol{\zeta}^{(1)}) \sim \frac{-2\ln(\alpha)}{c_\varphi(\boldsymbol{\zeta}^{(1)})}, \quad \text{as } \alpha \searrow 0,$$

(e.g., in Nikitin (1995), Thm. 1.2.1, pp. 5/6), where $f \sim g$ means asymptotic equivalence of f and g, we define and rewrite the *Bahadur ARE of $\tilde{\varphi}$ with respect to φ* in the following way:

$$e^B_{\tilde{\varphi},\varphi}(\beta, \boldsymbol{\zeta}^{(1)}) = \lim_{\alpha \searrow 0} \frac{N_\varphi(\alpha, \beta, \boldsymbol{\zeta}^{(1)})}{N_{\tilde{\varphi}}(\alpha, \beta, \boldsymbol{\zeta}^{(1)})} = \frac{c_{\tilde{\varphi}}(\boldsymbol{\zeta}^{(1)})}{c_\varphi(\boldsymbol{\zeta}^{(1)})}, \quad \boldsymbol{\zeta}^{(1)} \in \Theta_1.$$

Usually, the ratio of the slopes does not depend on the choice of β (Nikitin (1995), p. 3). Now, if $e^B_{\tilde{\varphi},\varphi}(\beta, \boldsymbol{\zeta}^{(1)}) > 1$, $\tilde{\varphi}$ performs better than φ in the 'Bahadur sense', and vice versa. With the aim of defining an optimal test in the 'Bahadur sense', we give a result of Raghavachari (1970), i.e., for every sequence $\varphi = \{\varphi^{(s)}\}_{s \in \mathbb{N}}$ of tests holds

$$\liminf_{s \to \infty} \frac{\ln L_s}{s} \geq -d_{KL}(\boldsymbol{\zeta}^{(1)}, \Theta_0), \quad P_{\boldsymbol{\zeta}^{(1)}}\text{-a.s.},$$

for $\boldsymbol{\zeta}^{(1)} \in \Theta_1$, where $d_{KL}(\boldsymbol{\zeta}^{(1)}, \Theta_0)$ denotes the Kullback-Leibler distance of $\boldsymbol{\zeta}^{(1)}$ and Θ_0 (cf. Subsection 2.1.8). This inequality implies that if the exact Bahadur slope of a test $\varphi = \{\varphi^{(s)}\}_{s \in \mathbb{N}}$ at $\boldsymbol{\zeta}^{(1)} \in \Theta_1$ exists, it is bounded from above:

$$c_\varphi(\boldsymbol{\zeta}^{(1)}) \leq 2 d_{KL}(\boldsymbol{\zeta}^{(1)}, \Theta_0). \tag{2.3.22}$$

Hence, if in (2.3.22) equality holds true for some test sequence $\tilde{\varphi} = \{\tilde{\varphi}^{(s)}\}_{s \in \mathbb{N}}$, it follows that $c_\varphi(\boldsymbol{\zeta}^{(1)}) \leq c_{\tilde{\varphi}}(\boldsymbol{\zeta}^{(1)})$ and, thus, $e^B_{\tilde{\varphi},\varphi}(\beta, \boldsymbol{\zeta}^{(1)}) \geq 1$ for every sequence $\varphi = \{\varphi^{(s)}\}_{s \in \mathbb{N}}$ of tests with existing Bahadur exact slope $c_\varphi(\boldsymbol{\zeta}^{(1)})$ at $\boldsymbol{\zeta}^{(1)}$.

Definition 2.3.16 (Bahadur asymptotic optimality)
Let $\tilde{\varphi} = \{\tilde{\varphi}^{(s)}\}_{s \in \mathbb{N}}$ be a sequence of tests satisfying (2.3.21) with $c_{\tilde{\varphi}}(\boldsymbol{\zeta}^{(1)}) = 2 d_{KL}(\boldsymbol{\zeta}^{(1)}, \Theta_0)$ at some $\boldsymbol{\zeta}^{(1)} \in \Theta_1$. Then, $\tilde{\varphi}$ is said to be asymptotically optimal in the sense of Bahadur at $\boldsymbol{\zeta}^{(1)}$.

The class of sequences of tests with equality in (2.3.22), for every $\boldsymbol{\zeta}^{(1)} \in \Theta_1$, is small (Nikitin (1995), p. 9). If the underlying class of probability measures forms an exponential family according to the assumptions of Subsection 2.3.5, Kim (1997) proved the following formula related to the case of a simple null hypothesis.

Theorem 2.3.17
For the test problem $H_0 : \boldsymbol{\zeta} = \boldsymbol{\zeta}^{(0)}$ against $H_1 : \boldsymbol{\zeta} \in \Theta_1 = \Theta \backslash \{\boldsymbol{\zeta}^{(0)}\}$, let $\varphi = \{\varphi^{(s)}\}_{s \in \mathbb{N}}$ be a sequence of tests with $\varphi^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(c_s, \infty)}(V^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$, $s \in \mathbb{N}$. For $s \in \mathbb{N}$, we assume that the test statistic $V^{(s)}$ depends upon $\tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)})$ only through $\overline{\tilde{\boldsymbol{x}}^{(s)}} = \frac{1}{s} \sum_{i=1}^{s} \boldsymbol{x}^{(i)}$, i.e. $V^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = W(\overline{\tilde{\boldsymbol{x}}^{(s)}})$ for some test statistic W on $(\mathfrak{X}, \mathfrak{B})$. Moreover, let the following conditions be fulfilled:

(i) For every $s \in \mathbb{N}$, the set $\{\boldsymbol{x} \in \mathfrak{X} : W(\boldsymbol{x}) \leq c_s\}$ is convex in $\mathfrak{X}$ (or, equivalently, the acceptance region $\{\tilde{\boldsymbol{x}}^{(s)} \in \mathfrak{X}^{1 \times s} : V^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \leq c_s\}$ of $\varphi^{(s)}$ is convex in $\mathfrak{X}^{1 \times s}$).

(ii) W is minimized at $\pi(\boldsymbol{\zeta}^{(0)})$.

(iii) W is strictly increasing on rays from $\pi(\boldsymbol{\zeta}^{(0)})$, i.e. for every $\boldsymbol{x} \in \mathfrak{X}$, the mapping $\rho \mapsto W(\pi(\boldsymbol{\zeta}^{(0)}) + \rho(\boldsymbol{x} - \pi(\boldsymbol{\zeta}^{(0)})))$ is strictly increasing in $\rho \in (0, 1)$.

(iv) W *is continuously differentiable on* $\mathfrak{X}$.

Then, the exact slope $c_\varphi(\boldsymbol{\zeta})$ *of* φ *at* $\boldsymbol{\zeta} \in \Theta_1$ *exists and equals*

$$2 \inf\{d_{KL}(\tilde{\boldsymbol{\zeta}}, \boldsymbol{\zeta}^{(0)}) : \, W(\pi(\tilde{\boldsymbol{\zeta}})) = W(\pi(\boldsymbol{\zeta})), \; \tilde{\boldsymbol{\zeta}} \in \Theta_1\}$$

Proof. In Kim (1997), Thm. 3.1. ∎

Notice that, from a theorem of Mathes & Truax (1967), the class of all sequences of tests with property (i) of Thm. 2.3.17 is *essentially complete*, i.e., for every arbitrary test Ψ, there exists a test $\tilde{\Psi}$ that satisfies (i) with $E_{\boldsymbol{\zeta}^{(0)}}(\tilde{\Psi}) \leq E_{\boldsymbol{\zeta}^{(0)}}(\Psi)$ and $E_{\boldsymbol{\zeta}}(\tilde{\Psi}) \geq E_{\boldsymbol{\zeta}}(\Psi)$ for all $\boldsymbol{\zeta} \in \Theta_1$ ($\tilde{\Psi}$ has uniformly not greater error probabilities than Ψ).
In the above context of a simple null hypothesis, from representation (2.3.8) in Subsection 2.3.2, the LR test statistic is given by $d_{KL}(\pi^{-1}(\overline{\tilde{\boldsymbol{x}}^{(s)}}), \boldsymbol{\zeta}^{(0)})$. Hence, by setting $W(\bullet) = d_{KL}(\pi^{-1}(\bullet), \boldsymbol{\zeta}^{(0)})$, (ii) is obvious, and, by application of La. 2.1.30 and some simple analysis, it can be shown that (iii) and (iv) hold for the sequence of LR tests, too. Finally, Birnbaum (1955) has shown that condition (i) is true for the sequence of LR tests. Hence, the sequence of LR test fulfils the conditions of Thm. 2.3.17, and it follows that $c_{\varphi_{LR}}(\boldsymbol{\zeta}^{(1)}) = 2 d_{KL}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(0)})$, $\boldsymbol{\zeta}^{(1)} \in \Theta_1$, and, thus, the sequence of LR test is asymptotically optimal in the sense of Bahadur. Moreover, for every other sequence of tests satisfying the conditions of Thm. 2.3.17 for some statistic $\tilde{W}$ on $(\mathfrak{X}, \mathfrak{B})$, its Bahadur exact slope at $\boldsymbol{\zeta}^{(1)}$ is given by

$$\inf\{c_{\varphi_{LR}}(\tilde{\boldsymbol{\zeta}}) : \, \tilde{W}(\pi(\tilde{\boldsymbol{\zeta}})) = \tilde{W}(\pi(\boldsymbol{\zeta}^{(1)})), \; \tilde{\boldsymbol{\zeta}} \in \Theta_1\},$$

and thus equals $c_{\varphi_{LR}}(\boldsymbol{\zeta}^{(1)})$ if and only if $c_{\varphi_{LR}}(\tilde{\boldsymbol{\zeta}})$ is constant on $\{\tilde{\boldsymbol{\zeta}} \in \Theta_1 : \tilde{W}(\pi(\tilde{\boldsymbol{\zeta}})) = \tilde{W}(\pi(\boldsymbol{\zeta}^{(1)}))\}$ (cf. Kim (1997)).

When a sequence of tests turns out to be asymptotically optimal in the sense of Bahadur, the related concept of *Bahadur deficiency* provides more information about the performance of the test (Kallenberg (1978), p. 3).

Definition 2.3.18
Let the test sequence $\tilde{\varphi} = \{\tilde{\varphi}^{(s)}\}_{s \in \mathbb{N}}$ *be asymptotically optimal in the sense of Bahadur at* $\boldsymbol{\zeta}^{(1)} \in \Theta_1$, *and let*

$$N_+(\alpha, \beta, \boldsymbol{\zeta}^{(1)}) = \inf_{\varphi} N_\varphi(\alpha, \beta, \boldsymbol{\zeta}^{(1)}), \quad 0 < \alpha < 1, \quad 0 < \beta < 1,$$

where the infimum is taken over all tests φ *satisfying the assumptions in the beginning of Subsection 2.3.5. If for all* $0 < \beta < 1$

$$\lim_{\alpha \searrow 0} \frac{N_{\tilde{\varphi}}(\alpha, \beta, \boldsymbol{\zeta}^{(1)}) - N_+(\alpha, \beta, \boldsymbol{\zeta}^{(1)})}{g(N_+(\alpha, \beta, \boldsymbol{\zeta}^{(1)}))} \leq a(\beta, \boldsymbol{\zeta}^{(1)}) \tag{2.3.23}$$

for some constant $a(\beta, \boldsymbol{\zeta}^{(1)}) > 0$, *we say that* $\tilde{\varphi}$ *is deficient in the sense of Bahadur at* $\boldsymbol{\zeta}^{(1)}$ *of order* $O(g(N_+(\alpha, \beta, \boldsymbol{\zeta}^{(1)})))$ *as* $\alpha \searrow 0$, *where* $g : \mathbb{R}_+ \to \mathbb{R}$ *is an increasing function.*

Obviously, the smaller the order of deficiency is, the faster the sample size $N_{\tilde{\varphi}}(\alpha, \beta, \boldsymbol{\zeta}^{(1)})$ needed to ensure that $\tilde{\varphi}$ has power β at $\boldsymbol{\zeta}^{(1)}$ tends to the purely theoretically optimal (minimal) sample size.
Assuming the sample situation of Subsection 2.3.5, Kallenberg (1978) proved the following theorem in the context of an underlying exponential family structure.

Theorem 2.3.19
Let for $A \subseteq \Theta$ and $\epsilon \geq 0$

$$U_\epsilon(A) = \{\boldsymbol{\zeta} \in \Theta : \ d_{KL}(\boldsymbol{\zeta}, A) \leq \epsilon\}$$

and

$$\iota(A) = \sup\{\epsilon \geq 0 : \ \exists K \subseteq \Theta \text{ compact with } U_\epsilon(A) \subseteq K\}.$$

Then, the following assertions hold true:

(i) Let $\Theta_0 \subseteq K \subseteq \Theta$, K compact. Then, for every $\boldsymbol{\zeta}^{(1)} \in int(\Theta_1)$ with $d_{KL}(\boldsymbol{\zeta}^{(1)}, \Theta_0) < \iota(\Theta_0)$, the sequence of LR tests is deficient in the sense of Bahadur at $\boldsymbol{\zeta}^{(1)}$ of order $O(\ln(N^+(\alpha, \beta, \boldsymbol{\zeta}^{(1)})))$ as $\alpha \searrow 0$.

(ii) Let $\Theta_1 \subseteq K \subseteq \Theta$, K compact. Then, for every $\boldsymbol{\zeta}^{(1)} \in \Theta_1$ with $d_{KL}(\boldsymbol{\zeta}^{(1)}, \Theta_0) > 0$, the sequence of LR tests is deficient in the sense of Bahadur at $\boldsymbol{\zeta}^{(1)}$ of order $O(\ln(N^+(\alpha, \beta, \boldsymbol{\zeta}^{(1)})))$ as $\alpha \searrow 0$.

Proof. In Kallenberg (1978), Thm.s 5.3.2, p. 111, and 5.3.3, p. 121. ■

2.3.7 Intermediate Asymptotic Relative Efficiency

Just pointing out the basic idea and a result of Kallenberg (1983), we will briefly introduce into another approach to asymptotic efficiency of tests, i.e. the concept of Intermediate or Kallenberg ARE, that combines the rudiments *vanishing levels* and *moving alternatives*, and, hence, can be considered a mixture of the approaches of Bahadur and Pitman.
We assume the same sample situation as in Subsection 2.3.5 and consider the test problem $H_0 : \boldsymbol{\zeta} \in \Theta_0$ against $H_1 : \boldsymbol{\zeta} \in \Theta_1 = \Theta \setminus \Theta_0$. Let $\{\varphi_\alpha^{(s)}\}_{s \in \mathbb{N}, 0<\alpha<1}$ be a family of tests, where $\varphi_\alpha^{(s)}$ has level-α and depends only on $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$. Let $\{\alpha_s\}_{s \in \mathbb{N}}$ be a sequence of levels that vanishes, but not too exponentially fast, i.e.

$$\lim_{s \to \infty} \alpha_s = 0 = \lim_{s \to \infty} \frac{\ln(\alpha_s)}{s},$$

and let $\{\tilde{\boldsymbol{\zeta}}^{(s)}\}_{s \in \mathbb{N}} \subseteq \Theta_1$ be a sequence of alternatives that converges (not too fast) to Θ_0 in the sense that

$$\lim_{s \to \infty} H(\tilde{\boldsymbol{\zeta}}^{(s)}, \Theta_0) = 0, \quad \lim_{s \to \infty} s\, H(\tilde{\boldsymbol{\zeta}}^{(s)}, \Theta_0)^2 = \infty,$$

where, for $\boldsymbol{\zeta}, \tilde{\boldsymbol{\zeta}} \in \Theta$, $H(\boldsymbol{\zeta}, \tilde{\boldsymbol{\zeta}}) = \left(\frac{1}{2}\int(\sqrt{f_{\boldsymbol{\zeta}}^{\boldsymbol{X}}} - \sqrt{f_{\tilde{\boldsymbol{\zeta}}}^{\boldsymbol{X}}})^2 d\mu\right)^{\frac{1}{2}}$ denotes the *Hellinger distance of $\boldsymbol{\zeta}$ and* $\tilde{\boldsymbol{\zeta}}$, and $H(\boldsymbol{\zeta}, \Theta_0) = \inf_{\boldsymbol{\zeta}_0 \in \Theta_0} H(\boldsymbol{\zeta}, \boldsymbol{\zeta}_0)$. Moreover, we assume that

$$0 < \liminf_{s \to \infty} E_{\tilde{\boldsymbol{\zeta}}^{(s)}}[\varphi_{\alpha_s}^{(s)}] \leq \limsup_{s \to \infty} E_{\tilde{\boldsymbol{\zeta}}^{(s)}}[\varphi_{\alpha_s}^{(s)}] < 1.$$

In particular, if the limit of the power of $\varphi_{\alpha_s}^{(s)}$ at $\tilde{\boldsymbol{\zeta}}^{(s)}$ exists as s tends to infinity, it lies in the open interval $(0, 1)$.
Let $\{\Psi_\alpha^{(s)}\}_{s\in\mathbb{N},0<\alpha<1}$ be another family of tests with the above properties. We define the integer

$$m_{\varphi,\Psi}(s) = \inf\{m \in \mathbb{N} : \ E_{\tilde{\boldsymbol{\zeta}}^{(s)}}[\Psi_{\alpha_s}^{(\tilde{m})}] \geq E_{\tilde{\boldsymbol{\zeta}}^{(s)}}[\varphi_{\alpha_s}^{(s)}] \quad \forall \tilde{m} \geq m\},$$

which is the minimal sample size $m_{\varphi,\Psi}(s)$ such that, for fixed $s \in \mathbb{N}$, $\Psi_{\alpha_s}^{(\tilde{m})}$ has power not less than $\varphi_{\alpha_s}^{(s)}$ at $\tilde{\boldsymbol{\zeta}}^{(s)}$ for all $\tilde{m} \geq m_{\varphi,\Psi}(s)$. Then, the *strong asymptotic i-efficiency* (or *Intermediate ARE*, or *Kallenberg ARE*) of φ with respect to Ψ is defined as

$$e_{\varphi,\Psi}^K = \lim_{s\to\infty} \frac{m_{\varphi,\Psi}(s)}{s},$$

provided that the limit exists and does not depend on the choice of the sequences $\{\alpha_s\}_{s\in\mathbb{N}}$ and $\{\tilde{\boldsymbol{\zeta}}^{(s)}\}_{s\in\mathbb{N}}$ (see Kallenberg (1983)). Similar to the approach of Bahadur to ARE, the Intermediate ARE is defined as the limit of the ratios of minimal sample sizes needed to attain a certain power, when the sequence $\{\varphi_{\alpha_s}^{(s)}\}_{s\in\mathbb{N}}$, respectively $\{\Psi_{\alpha_s}^{(s)}\}_{s\in\mathbb{N}}$, is used. The main difference between both approaches lies in the fact that the alternative, where powers are evaluated at, is fixed in the Bahadur case.

Definition 2.3.20 (Strong i-efficiency)
A family of tests $\varphi = \{\varphi_\alpha^{(s)}\}_{s\in\mathbb{N},0<\alpha<1}$ is called strongly i-efficient if $e_{\varphi,\Psi}^K \geq 1$ for every family $\Psi = \{\Psi_\alpha^{(s)}\}_{s\in\mathbb{N},0<\alpha<1}$ of tests, for that this limit exists.

Related to the case that the underlying set of distributions forms a multivariate exponential family, we cite a result of Kallenberg (1983).

Theorem 2.3.21
Let $\Theta_0 \subseteq K \subseteq \Theta$ for some compact subset K of Θ. Then, the sequence of LR tests is strongly i-efficient.

Proof. In Kallenberg (1983), Thm. 3.1. ■

Chapter 3

SOSs with Known Baseline Distribution

Throughout this chapter, we consider *SOSs with conditional proportional hazard rates* which result from the general sequential model (cf. Def.s 1.1.1 and 1.1.2) by setting

$$F_j = 1 - (1-F)^{\alpha_j}, \quad 1 \leq j \leq n, \tag{3.0.1}$$

where F is an absolutely continuous baseline distribution function with corresponding density function f and $\alpha_1, ..., \alpha_n$ are positive parameters. The hazard rate of F_j is then proportional to the hazard rate of F and given by $\alpha_j f/(1-F)$. When the model is used to describe a sequential $(n-r+1)$-out-of-n system, the interpretation is as follows. All components of the system start operating at hazard rate $\alpha_1 f/(1-F)$. Then, upon occurrence of the first failure of a component, the hazard rate is supposed to change from $\alpha_1 f/(1-F)$ to $\alpha_2 f/(1-F)$, and the system continues to work with $n-1$ remaining operative components. Upon failure of the second component, the failure rate is supposed to change from $\alpha_2 f/(1-F)$ to $\alpha_3 f/(1-F)$, and so on. Finally, the r^{th} SOS is the lifetime of the system.
In the above situation, the joint density of the first r SOSs $X_*^{(1)}, \ldots, X_*^{(r)}$ is given by

$$\begin{aligned} f_{\boldsymbol{\alpha}}^{X_*^{(1)},\ldots,X_*^{(r)}}(x_1,\ldots,x_r) &= \tfrac{n!}{(n-r)!} \left(\prod_{j=1}^{r} \alpha_j\right) \left(\prod_{j=1}^{r-1} (1-F(x_j))^{m_j} f(x_j)\right) \\ &\quad \times (1-F(x_r))^{\alpha_r(n-r+1)-1} f(x_r) \end{aligned} \tag{3.0.2}$$

on the cone $F^{-1}(0) < x_1 < \cdots < x_r < F^{-1}(1)$, with $1 \leq r \leq n$, and $m_j = (n-j+1)\,\alpha_j - (n-j)\,\alpha_{j+1} - 1, \quad 1 \leq j \leq r-1$ (cf. Kamps (1995a,b) and Cramer & Kamps (2001*b*)). The index $\boldsymbol{\alpha}$ denotes the vector $(\alpha_1, \ldots, \alpha_r)'$ of model parameters. Notice that, in the distribution theoretical sense, common OSs based on F are included in the model of SOSs with conditional proportional hazard rates by setting $\alpha_1 = \cdots = \alpha_r = 1$.
In this chapter, we show that the joint distribution of SOSs with conditional proportional hazard rates forms a multivariate exponential family in the model parameters, when the baseline distribution function F is assumed to be known. As a consequence, the application of the results of Chapter 2 leads to much simplified proofs of former results in literature, and also enables us to state new useful

properties, in particular concerning statistical inference on the model parameters. To name only two, MLEs of the model parameters can easily be obtained, where asymptotic efficiency of the estimators can be shown, and UMPU one- and two-sided tests on the model parameters $\alpha_1, ..., \alpha_r$ are derived as well as interval hypotheses are examined. These particular findings have already been published (Bedbur et al. (2010), Bedbur (2010)).

3.1 SOSs as Exponential Family in Model Parameters

Upon introducing the statistics

$$\begin{aligned} T_1(x_1, \ldots, x_r) &= n \ln(1 - F(x_1)), \\ T_j(x_1, \ldots, x_r) &= (n - j + 1) \ln\left(\frac{1 - F(x_j)}{1 - F(x_{j-1})}\right), \quad 2 \leq j \leq r, \end{aligned} \tag{3.1.1}$$

where $F^{-1}(0) < x_1 < \cdots < x_r < F^{-1}(1)$, the joint density of the first r SOSs $X_*^{(1)}, \ldots, X_*^{(r)}$ with conditional proportional hazard rates (see (3.0.2)) can be rewritten as

$$\begin{aligned} f_{\boldsymbol{\alpha}}^{X_*^{(1)},\ldots,X_*^{(r)}}(x_1, \ldots, x_r) &= \frac{n!}{(n-r)!} \left(\prod_{j=1}^{r} \alpha_j\right) \left(\prod_{j=1}^{r-1} (1 - F(x_j))^{m_j} f(x_j)\right) \\ &\quad \times (1 - F(x_r))^{\alpha_r(n-r+1)-1} f(x_r) \\ &= \frac{n!}{(n-r)!} \left(\prod_{j=1}^{r} \alpha_j\right) \left(\prod_{j=1}^{r-1} (1 - F(x_j))^{(n-j+1)\alpha_j - (n-j)\alpha_{j+1} - 1} f(x_j)\right) \\ &\quad \times (1 - F(x_r))^{\alpha_r(n-r+1)-1} f(x_r) \\ &= \left(\prod_{j=1}^{r} \alpha_j\right) \left(\prod_{j=1}^{r-1} (1 - F(x_j))^{(n-j+1)\alpha_j - (n-j)\alpha_{j+1}}\right) \\ &\quad \times (1 - F(x_r))^{\alpha_r(n-r+1)} \left(\frac{n!}{(n-r)!} \prod_{j=1}^{r} \frac{f(x_j)}{1 - F(x_j)}\right) \\ &= \left(\prod_{j=1}^{r} \alpha_j\right) \exp\left\{ \left(\sum_{j=1}^{r-1} ((n-j+1)\alpha_j - (n-j)\alpha_{j+1}) \ln(1 - F(x_j))\right)\right. \\ &\quad \left. + \alpha_r(n-r+1) \ln(1 - F(x_r)) \right\} \left(\frac{n!}{(n-r)!} \prod_{j=1}^{r} \frac{f(x_j)}{1 - F(x_j)}\right) \end{aligned}$$

$$
\begin{aligned}
&= \left(\prod_{j=1}^{r}\alpha_j\right)\exp\Bigg\{\left(\sum_{j=1}^{r-1}(n-j+1)\alpha_j\ln(1-F(x_j))\right)\\
&\quad -\left(\sum_{j=2}^{r}(n-j+1)\alpha_j\ln(1-F(x_{j-1}))\right)\\
&\quad +\alpha_r(n-r+1)\ln(1-F(x_r))\Bigg\}\left(\frac{n!}{(n-r)!}\prod_{j=1}^{r}\frac{f(x_j)}{1-F(x_j)}\right)\\
&= \left(\prod_{j=1}^{r}\alpha_j\right)\exp\left\{n\alpha_1\ln(1-F(x_1))+\sum_{j=2}^{r}(n-j+1)\alpha_j\ln\left(\frac{1-F(x_j)}{1-F(x_{j-1})}\right)\right\}\\
&\quad \times\left(\frac{n!}{(n-r)!}\prod_{j=1}^{r}\frac{f(x_j)}{1-F(x_j)}\right)\\
&= \left(\prod_{j=1}^{r}\alpha_j\right)\exp\left\{\sum_{j=1}^{r}\alpha_j\,T_j\,(x_1,\ldots,x_r)\right\}\left(\frac{n!}{(n-r)!}\prod_{j=1}^{r}\frac{f(x_j)}{1-F(x_j)}\right).
\end{aligned}
$$

3.1.1 First Results

For fixed $r \in \{1,...,n\}$, let $\boldsymbol{X} = (X_*^{(1)},\ldots,X_*^{(r)})'$ denote the column vector of the first r SOSs with conditional proportional hazard rates. $\boldsymbol{X}$ takes on values in the measurable space $(\mathbb{R}^r_<, \mathbb{R}^r_< \cap \mathbb{B}^r)$, where $\mathbb{R}^r_< = \{\boldsymbol{x} = (x_1,...,x_r)' \in \mathbb{R}^r : \ F^{-1}(0) < x_1 < \cdots < x_r < F^{-1}(1)\}$ denotes the (truncated) cone of increasing real numbers and $\mathbb{R}^r_< \cap \mathbb{B}^r$ the Borel sets of $\mathbb{R}^r_<$. Let $\mathfrak{P}^{\boldsymbol{X}} = \{P^{\boldsymbol{X}}_{\boldsymbol{\alpha}} = f^{\boldsymbol{X}}_{\boldsymbol{\alpha}}\lambda^r|_{\mathbb{R}^r_<} : \boldsymbol{\alpha} = (\alpha_1,...,\alpha_r)' \in \mathbb{R}^r_+\}$, where λ^r denotes Lebesgue measure on $(\mathbb{R}^r,\mathbb{B}^r)$ and $\cdot|_B$ the restriction of a measure to a measurable subset $B \in \mathbb{B}^r$. The densities $f^{\boldsymbol{X}}_{\boldsymbol{\alpha}}$ are given by

$$
f^{\boldsymbol{X}}_{\boldsymbol{\alpha}}(\boldsymbol{x}) \;=\; C(\boldsymbol{\alpha})\exp\left\{\sum_{j=1}^{r}\alpha_jT_j(\boldsymbol{x})\right\}h(\boldsymbol{x}),\quad \boldsymbol{x}=(x_1,...,x_r)'\in\mathbb{R}^r_<,\quad \lambda^r|_{\mathbb{R}^r_<}\text{-a.e.}, \tag{3.1.2}
$$

with $C(\boldsymbol{\alpha}) = \prod_{j=1}^{r}\alpha_j$, statistics T_j, $1 \le j \le r$, as in (3.1.1), and

$$
h(\boldsymbol{x}) = \frac{n!}{(n-r)!}\prod_{j=1}^{r}\frac{f(x_j)}{1-F(x_j)},\ \boldsymbol{x}=(x_1,...,x_r)'\in\mathbb{R}^r_< \,. \tag{3.1.3}
$$

Then $\mathfrak{P}^{\boldsymbol{X}}$ forms a r-parametrical exponential family in the model parameters $\alpha_1,...,\alpha_r$ and statistics $T_1,...,T_r$. Here, the natural parameter space Θ^* of the exponential family is given by $\mathbb{R}^r_+$, which can be seen as follows. Clearly, $\mathbb{R}^r_+ \subseteq \Theta^*$, and $\boldsymbol{\alpha} \notin \Theta^*$ if the components of $\boldsymbol{\alpha}$ are nonnegative and (at least) one component equals zero. Suppose, there exists $\boldsymbol{\alpha} \in \Theta^*$ having some negative component.

Then, since Θ^* is a convex subset of $\mathbb{R}^r$ (cf. Thm. 2.1.13), there exists $\tilde{\boldsymbol{\alpha}} \in \Theta^*$ on the line between $\boldsymbol{\alpha}$ and $(1, \ldots, 1)'$ with nonnegative components and at least one component equal to zero, which forms a contradiction. Thus, $\Theta^* = \mathbb{R}^r_+$ is shown. Moreover, the density of $P^{\boldsymbol{X}}_{\boldsymbol{\alpha}}$ is given in the *canonical form* (cf. Def. 2.1.11).

Defining the measure $\nu = h\lambda^r|_{\mathbb{R}^r_<}$, we obtain another representation of $\mathfrak{P}^{\boldsymbol{X}}$, i.e. $P^{\boldsymbol{X}}_{\boldsymbol{\alpha}}$ has a ν-density

$$C(\boldsymbol{\alpha}) \exp\left\{\sum_{j=1}^r \alpha_j T_j(\boldsymbol{x})\right\}, \quad \boldsymbol{x} = (x_1, ..., x_r)' \in \mathbb{R}^r_<,$$

and, moreover, the distribution family $\mathfrak{P}^{\boldsymbol{T}} = \{P^{\boldsymbol{T}}_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathbb{R}^r_+\}$ on $(\mathbb{R}^r, \mathbb{B}^r)$ forms a r-parametrical exponential family in $\alpha_1, \ldots, \alpha_r$ and the projections $p_1, \ldots, p_r$, where $P^{\boldsymbol{T}}_{\boldsymbol{\alpha}}$ has a $\nu^{\boldsymbol{T}}$-density

$$g^{\boldsymbol{T}}_{\boldsymbol{\alpha}}(\boldsymbol{t}) = C(\boldsymbol{\alpha}) e^{\boldsymbol{\alpha}'\boldsymbol{t}}, \quad \boldsymbol{t} = (t_1, \ldots, t_r)' \in \mathbb{R}^r_-. \tag{3.1.4}$$

It is well-known that the statistics $-T_1(\boldsymbol{X}), \ldots, -T_r(\boldsymbol{X})$ are jointly independent random variables, and $-T_j(\boldsymbol{X}) \sim Exp\,(\alpha_j^{-1})$, $\quad 1 \le j \le r$, has an exponential distribution with scale parameter α_j^{-1}, i.e. a λ^1-density of $-T_j(\boldsymbol{X})$ is given by $f^{-T_j(\boldsymbol{X})}(x) = \alpha_j \exp\{-\alpha_j x\} \mathbb{1}_{(0,\infty)}(x)$ (cf. Kamps (1995*b*), p. 81, and Cramer & Kamps (1996)). Once having observed the exponential family structure of $\mathfrak{P}^{\boldsymbol{X}}$, this result can also be obtained by deriving the moment generating function of $-\boldsymbol{T} = (-T_1, ..., -T_r)'$, where the expectation is computed with respect to $P^{\boldsymbol{X}}_{\boldsymbol{\alpha}}$. For this, let $\boldsymbol{\alpha} \in \mathbb{R}^r_+$ be fixed and let $U = (-\delta, \delta)^r$, where $\delta = \min\{\alpha_1, \ldots, \alpha_r\}/2 > 0$. Then, for $\boldsymbol{t} = (t_1, ..., t_r)' \in U$, application of La. 2.1.18 yields

$$m_{-\boldsymbol{T}}(\boldsymbol{t}) = m_{\boldsymbol{T}}(-\boldsymbol{t}) = \frac{C(\boldsymbol{\alpha})}{C(\boldsymbol{\alpha} - \boldsymbol{t})} = \prod_{j=1}^r \frac{\alpha_j}{\alpha_j - t_j}, \tag{3.1.5}$$

and, hence, the assertion is established. Moreover, it can be shown that $\nu^{\boldsymbol{T}} = (h\lambda^r|_{\mathbb{R}^r_<})^{\boldsymbol{T}} = \mathbb{1}_{\mathbb{R}^r_-} \lambda^r$.

For brevity, we state our findings in what follows only for the family $\mathfrak{P}^{\boldsymbol{X}}$ of distributions. Clearly, respective results are also true by considering the exponential family $\mathfrak{P}^{\boldsymbol{T}}$.

Lemma 3.1.1
In the above situation, we find:

(i) $\mathfrak{P}^{\boldsymbol{X}}$ *is strictly* r*-parametrical and of full rank.*

(ii) $\boldsymbol{T} = (T_1, \ldots, T_r)'$ *is a minimal sufficient and complete statistic for* $\mathfrak{P}^{\boldsymbol{X}}$*.*

Proof. (i). By application of Thm. 2.1.9 (i), $\mathfrak{P}^{\boldsymbol{X}}$ is strictly r-parametrical if and only if (the projections) $p_j(\boldsymbol{\alpha}) = \alpha_j$, $j \in \{1, \ldots, r\}$, are affinely independent and $T_1, ..., T_r$ are $\mathfrak{P}^{\boldsymbol{X}}$-affinely independent. The first condition is obvious. For the latter, it is sufficient to show that $T_1, ..., T_r$ are $P^{\boldsymbol{X}}_{\boldsymbol{\alpha}}$-affinely independent for an arbitrary fixed $\boldsymbol{\alpha} \in \mathbb{R}^r_+$ (cf. Rem 2.1.4). Setting $\boldsymbol{T} = (T_1, ..., T_r)'$, this assertion immediately follows from Thm. 2.1.9 (ii) since

$$\mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}) = \mathbf{diag}\left(\frac{1}{\alpha_1^2}, ..., \frac{1}{\alpha_r^2}\right) > 0.$$

Here and in the following, $\mathbf{diag}(d_1, \ldots, d_r)$ denotes a diagonal matrix with diagonal elements $d_1, \ldots, d_r$.
(ii). The assertion follows immediately by application of (i) and La. 2.1.20. ■

Using the denotations of Chapter 2, we obtain

$$\kappa : \quad \mathbb{R}_+^r \to \mathbb{R} : \quad \boldsymbol{\alpha} \mapsto -\ln[C(\boldsymbol{\alpha})] = -\sum_{j=1}^{r} \ln(\alpha_j) \tag{3.1.6}$$

(cf. (2.1.3)), and

$$\frac{dP_{\boldsymbol{\alpha}}^{\boldsymbol{X}}}{d\nu}(\boldsymbol{x}) = e^{\boldsymbol{\alpha}'\boldsymbol{T}(\boldsymbol{x})-\kappa(\boldsymbol{\alpha})}, \quad \boldsymbol{x} \in \mathbb{R}_<^r, \quad \nu\text{-a.e.},$$

where $\nu = h\lambda^r|_{\mathbb{R}_+^r}$. Moreover, the mapping

$$\pi : \quad \mathbb{R}_+^r \to \pi(\mathbb{R}_+^r) : \quad \boldsymbol{\alpha} \mapsto E_{\boldsymbol{\alpha}}[\boldsymbol{T}] = -\left(\frac{1}{\alpha_1}, \ldots, \frac{1}{\alpha_r}\right)', \tag{3.1.7}$$

where $\pi(\mathbb{R}_+^r) = \mathbb{R}_-^r$, is bijective with inverse function

$$\pi^{-1} : \quad \mathbb{R}_-^r \to \mathbb{R}_+^r : \quad \boldsymbol{t} \mapsto -\left(\frac{1}{t_1}, \ldots, \frac{1}{t_r}\right)'. \tag{3.1.8}$$

Both functions, κ and π, are infinitely often differentiable with respect to $\boldsymbol{\alpha}$ (cf. Thm. 2.1.15), and, in particular, are connected by the relation $\pi(\boldsymbol{\alpha}) = \nabla\kappa(\boldsymbol{\alpha})$ (see (2.1.14) and (2.1.16)). In virtue of (2.1.15), the Hessian matrix of κ is

$$\mathbf{H}_\kappa(\boldsymbol{\alpha}) = \mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}) = \mathbf{diag}\left(\frac{1}{\alpha_1^2}, \ldots, \frac{1}{\alpha_r^2}\right).$$

Moreover, by application of Thm. 2.1.22, the score function of $\mathfrak{P}^{\boldsymbol{X}}$ is given by

$$\boldsymbol{U}_{\boldsymbol{\alpha}} = \left(T_1 + \frac{1}{\alpha_1}, \ldots, T_r + \frac{1}{\alpha_r}\right)', \tag{3.1.9}$$

and the Fisher information matrix of $\mathfrak{P}^{\boldsymbol{X}}$ equals

$$\mathbf{I}_f(\boldsymbol{\alpha}) = \mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}) = \mathbf{diag}\left(\frac{1}{\alpha_1^2}, \ldots, \frac{1}{\alpha_r^2}\right) \tag{3.1.10}$$

at $\boldsymbol{\alpha} \in \mathbb{R}_+^r$.

3.1.2 Product Measures

Aiming at statistical inference with SOSs given by (3.1.2), we continue by considering the family of corresponding product probability measures.
We will assume to have s iid vectors $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$ of SOSs and corresponding vectors $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ of observations with $\boldsymbol{X}^{(i)} \sim f_{\boldsymbol{\alpha}}^{\boldsymbol{X}}$, $1 \leq i \leq s$. Let $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}} = \{P_{\boldsymbol{\alpha}}^{\tilde{\boldsymbol{X}}^{(s)}} = \otimes_{i=1}^{s} P_{\boldsymbol{\alpha}}^{\boldsymbol{X}^{(i)}} : \boldsymbol{\alpha} \in \mathbb{R}_{+}^{r}\}$ denote the family of the respective product probability measures. Defining the product measure $\nu^{(s)} = \otimes_{i=1}^{s} \nu$, a joint $\nu^{(s)}$-density of $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$ is given by

$$\begin{aligned} f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= C(\boldsymbol{\alpha})^{s} e^{\boldsymbol{\alpha}' \boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})} \\ &= e^{\boldsymbol{\alpha}' \boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - s\kappa(\boldsymbol{\alpha})}, \quad \tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}) \in (\mathbb{R}_{<}^{r})^{1 \times s}, \end{aligned} \tag{3.1.11}$$

where $\boldsymbol{T}^{(s)} = (T_1^{(s)}, \ldots, T_r^{(s)})'$ and

$$T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \sum_{i=1}^{s} T_j(\boldsymbol{x}^{(i)}), \quad \tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}) \in (\mathbb{R}_{<}^{r})^{1 \times s}, \quad 1 \leq j \leq r.$$

Hence, $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}$ forms a r-parametrical exponential family in the model parameters $\alpha_1, \ldots, \alpha_r$ and statistics $T_1^{(s)}, \ldots, T_r^{(s)}$ (cf. La. 2.1.25). As above, for fixed $\boldsymbol{\alpha} \in \mathbb{R}_{+}^{r}$, $\delta = \min\{\alpha_1, \ldots, \alpha_r\}/2 > 0$, and $U = (-\delta, \delta)^r$, the moment generating function of $-\boldsymbol{T}^{(s)}$ at $\boldsymbol{t} \in U$, where the expectation is computed with respect to $P_{\boldsymbol{\alpha}}^{\tilde{\boldsymbol{X}}^{(s)}}$, can be derived according to

$$m_{-\boldsymbol{T}^{(s)}}(\boldsymbol{t}) = m_{\boldsymbol{T}^{(s)}}(-\boldsymbol{t}) = \left(\frac{C(\boldsymbol{\alpha})}{C(\boldsymbol{\alpha} - \boldsymbol{t})} \right)^{s} = \prod_{j=1}^{r} \left(\frac{\alpha_j}{\alpha_j - t_j} \right)^{s},$$

and, hence, $T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)}), \ldots, T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ are jointly independent random variables, where $-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ has a gamma distribution with scale parameter α_j^{-1} and shape parameter s, i.e.,

$$-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = \sum_{i=1}^{s} -T_j(\boldsymbol{X}^{(i)}) \sim \Gamma(s, \alpha_j^{-1}), \ 1 \leq j \leq r,$$

and a λ^1-density of $-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ is given by

$$f^{-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}(x) = \frac{\alpha_j^{s}}{(s-1)!} x^{s-1} e^{-\alpha_j x} \mathbb{1}_{(0,\infty)}(x).$$

In La. 3.1.2, we summarize our findings for the class $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}$ of distributions.

Lemma 3.1.2
In the above situation, we obtain:

(i) $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}$ *is strictly* r*-parametrical and of full rank.*

(ii) The statistic $\boldsymbol{T}^{(s)} = (T_1^{(s)}, ..., T_r^{(s)})'$ *is minimal sufficient and complete for* $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}$.

Proof. Since

$$\mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}^{(s)}) = \mathbf{diag}\left(\frac{s}{\alpha_1^2}, ..., \frac{s}{\alpha_r^2}\right) > 0,$$

the assertions are obvious by the same arguments as in the proof of La. 3.1.1 (see also La. 2.1.25). ∎

The score function of $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}$ and the score statistic of $\tilde{\boldsymbol{X}}^{(s)}$ are given by

$$\begin{aligned} \boldsymbol{U}_{\boldsymbol{\alpha}}^{(s)} &= \left(T_1^{(s)} + \frac{s}{\alpha_1}, \ldots, T_r^{(s)} + \frac{s}{\alpha_r}\right)', \\ \boldsymbol{U}_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= \left(T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + \frac{s}{\alpha_1}, \ldots, T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + \frac{s}{\alpha_r}\right)', \end{aligned} \tag{3.1.12}$$

and the Fisher information matrix of $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}$ equals

$$\mathbf{I}_f^{(s)}(\boldsymbol{\alpha}) = \mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}^{(s)}) = \mathbf{diag}\left(\frac{s}{\alpha_1^2}, \ldots, \frac{s}{\alpha_r^2}\right)$$

at $\boldsymbol{\alpha} \in \mathbb{R}_+^r$.

3.1.3 The Univariate and the General Model

It is worth mentioning some univariate and more general results. Suppose, for some $j \in \{1, \ldots, r\}$, $\alpha_1, \ldots, \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_r$ are considered as fixed nuisance parameters. Then, by setting

$$\tilde{h}(\boldsymbol{x}) = \left(\prod_{\substack{l=1 \\ l\neq j}}^{r} \alpha_l\right) \exp\left\{\sum_{\substack{l=1 \\ l\neq j}}^{r} \alpha_l T_l(\boldsymbol{x})\right\} h(x), \ \boldsymbol{x} = (x_1, ..., x_r)' \in \mathbb{R}_<^r, \tag{3.1.13}$$

with h as in (3.1.3), (3.1.2) can be rewritten as

$$f_{\boldsymbol{\alpha}}^{\boldsymbol{X}}(\boldsymbol{x}) = \alpha_j e^{\alpha_j T_j(\boldsymbol{x})} \tilde{h}(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, ..., x_r)' \in \mathbb{R}_<^r, \quad \lambda^r|_{\mathbb{R}_<^r}\text{-a.e..}$$

Hence, $\mathfrak{P}_j^{\boldsymbol{X}} = \{P_{\boldsymbol{\alpha}}^{\boldsymbol{X}} : \alpha_j > 0\}$ forms a one-parameter exponential family in α_j and T_j, and the findings of Chapter 2 can be applied to the univariate case, too. In particular, by setting $\tilde{\nu} = \tilde{h}\lambda^r|_{\mathbb{R}_<^r}$, $\tilde{\nu}^{(s)} = \otimes_{i=1}^s \tilde{\nu}$ and $\tilde{\kappa}(\alpha) = -\ln(\alpha)$, $\alpha > 0$, a joint $\tilde{\nu}^{(s)}$-density of $\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(s)}$ is given by

$$\begin{aligned} f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= \alpha_j^s e^{\alpha_j T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})} \\ &= e^{\alpha_j T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - s\tilde{\kappa}(\alpha_j)}, \quad \tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}) \in (\mathbb{R}_<^r)^{1\times s}, \end{aligned} \tag{3.1.14}$$

(cf. (3.1.11)). $\mathfrak{P}_j^{\tilde{\boldsymbol{X}}^{(s)}} = \{P_{\boldsymbol{\alpha}}^{\tilde{\boldsymbol{X}}^{(s)}} : \alpha_j > 0\}$ forms a one-parameter exponential family in α_j and $T_j^{(s)}$. For the sake of brevity, we state the following results directly for the family of product measures.

Lemma 3.1.3
In the above situation, $T_j^{(s)}$ is a minimal sufficient and complete statistic for $\mathfrak{P}_j^{\tilde{\boldsymbol{X}}^{(s)}}$.

Proof. Application of La. 2.1.20 yields the assertion ($k = 1$). ■

In the univariate case, independent of j, the functions κ and π (cf. (2.1.3) and (2.1.16)) are given by $\tilde{\kappa}(\alpha) = -\ln(\alpha)$ and $\tilde{\pi}(\alpha) = -\alpha^{-1}$, $\alpha > 0$. It should be noted that $\pi(\boldsymbol{\alpha}) = (\tilde{\pi}(\alpha_1), \ldots, \tilde{\pi}(\alpha_r))'$ with the mapping π from the multivariate case (see (3.1.7)). Moreover, the Fisher information of $\mathfrak{P}_j^{\boldsymbol{X}}$ is given by $\tilde{I}_f(\alpha) = \alpha^{-2}$ at $\alpha > 0$.

Finally, we point out the following important remark, which, once again, demonstrates the favourable and useful structure of the densities considered.

Remark 3.1.4
As we have shown in this subsection, $\mathfrak{P}^{\boldsymbol{X}}$ can be considered as a one-parameter exponential family in α_j and T_j if the remaining parameters are assumed to be fixed nuisance parameters. All arguments and conclusions mentioned above remain valid in the following genzeralized set-up. Let $I \subseteq \{1, \ldots, r\}$ be an index set with the interpretation that α_j is a fixed nuisance parameter if and only if $j \notin I$. Then, $\mathfrak{P}^{\boldsymbol{X}}$ forms a strictly $|I|$-parametrical exponential family in the model parameters α_j and statistics T_j, $j \in I$, and by introducing $\tilde{h}_I(\boldsymbol{x}) = \left(\prod_{j \notin I} \alpha_j\right) \exp\{\sum_{j \notin I} \alpha_j T_j(\boldsymbol{x})\} h(\boldsymbol{x})$, $f_{\boldsymbol{\alpha}}^{\boldsymbol{X}}$ can be written as

$$f_{\boldsymbol{\alpha}}^{\boldsymbol{X}}(\boldsymbol{x}) = \left(\prod_{j \in I} \alpha_j\right) \exp\left\{\sum_{j \in I} \alpha_j T_j(\boldsymbol{x})\right\} \tilde{h}_I(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, ..., x_r)' \in \mathbb{R}^r_<, \quad \lambda^r|_{\mathbb{R}^r_<}\text{-a.e..}$$

The cases $|I| = 1$ and $|I| = r$ corresponding to the situations, where a single parameter or all parameters are of interest, are included in this set-up. For the sake of brevity and a simplified notation, we shall state subsequent results only for these two particular cases. If $1 < |I| < r$, respective statements can similarly be shown.

3.2 Estimation of Model Parameters

In practical applications, one is interested in estimating the model parameters $\alpha_1, \ldots, \alpha_r$ based on a sample of size s of independent SOSs. In Cramer & Kamps (1996), the MLEs $\alpha_1^*, \ldots, \alpha_r^*$ of $\alpha_1, \ldots, \alpha_r$ have been calculated directly and some useful properties have been shown. Based on differently structured samples, MLEs are presented in Cramer & Kamps (2001*b*).
Once having observed the exponential family structure of SOSs with conditional proportional hazard rates, these and further results are immediate consequences, where, in particular, efficiency notions can easily be addressed. This is subject matter of Subsections 3.2.1 and 3.2.2.
If SOSs are used to describe sequential systems, the prior information that the α's are simply ordered can be taken into account. This particular case is discussed in Subsection 3.2.3.

3.2.1 MLEs and UMVUEs

Theorem 3.2.1
Let sample situation (3.1.11) be given and let $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(s)})$. Then, the following statements hold true:

(i) The unique MLE of $\boldsymbol{\alpha}$ based on the independent observations $\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(s)}$ of $\boldsymbol{X}$ is given by

$$\boldsymbol{\alpha}^{*(s)} = \left(-\frac{s}{T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \ldots, -\frac{s}{T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}\right)'.$$

Moreover, if $g : \mathbb{R}_+^r \to \Gamma$ is a bijective function, $g(\boldsymbol{\alpha}^{*(s)})$ is the MLE of $g(\boldsymbol{\alpha})$ based on s independent observations of $\boldsymbol{X}$.

(ii) The unique MLE of α_j based on the independent observations $\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(s)}$ of $\boldsymbol{X}$ is given by

$$\alpha_j^{*(s)} = -\frac{s}{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \quad 1 \leq j \leq r.$$

$\alpha_1^{*(s)}, ..., \alpha_r^{*(s)}$ are jointly independent, and $\alpha_j^{*(s)}$ is inverted gamma distributed with shape parameter s and scale parameter $s\alpha_j$, $1 \leq j \leq r$, i.e. $\alpha_j^{*(s)}$ has a λ^1-density

$$f^{\alpha_j^{*(s)}}(x) = \frac{(s\alpha_j)^s}{(s-1)!}\left(\frac{1}{x}\right)^{s+1} e^{-s\alpha_j x^{-1}} \mathbb{1}_{(0,\infty)}(x). \tag{3.2.1}$$

Moreover, if $g : \mathbb{R}_+ \to \Gamma$ is a bijective function, $g(\alpha_j^{*(s)})$ is the MLE of $g(\boldsymbol{\alpha}_j)$ based on s independent observations of $\boldsymbol{X}$.

Proof. (i). In virtue of (3.1.8) and $P_{\boldsymbol{\alpha}}(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \in \mathbb{R}_-^r) = 1$, it follows from Thm. 2.2.3 that the unique MLE of $\boldsymbol{\alpha}$ based on the independent observations $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$ is given by

$$\boldsymbol{\alpha}^{*(s)} = \pi^{-1}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\right) = \left(-\frac{s}{T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \ldots, -\frac{s}{T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}\right)'.$$

The respective result for g is then obvious from Thm. 2.2.4.
(ii). Similarly to the proof of (i), the MLE of α_j based on the independent observations $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$ can be derived by noticing of the results of Subsection 3.1.3. In the given situation, the vector of the MLEs of the α's coincides with the MLE of $\boldsymbol{\alpha}$. $T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)}), ..., T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ are jointly independent and so are the MLEs $\alpha_1^{*(s)}, ..., \alpha_r^{*(s)}$. With $-s^{-1}T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \sim \Gamma(s, (s\alpha_j)^{-1})$ for $1 \leq j \leq r$, the distribution of $\alpha_j^{*(s)}$ is obvious. The result for g follows again by application of Thm. 2.2.4. ■

Subsequently, we state some moments of the MLEs which can also be found in Cramer & Kamps (1996).

Lemma 3.2.2
In the above situation, we find:

(i) $E_{\boldsymbol{\alpha}}[\alpha_j^{*(s)}] = \frac{s}{s-1}\alpha_j,\ s > 1, \quad E_{\boldsymbol{\alpha}}[(\alpha_j^{*(s)})^k] = \frac{(s-k-1)!}{(s-1)!}(s\alpha_j)^k,\ 2 \leq k \leq s-1, s > 2.$

(ii) $Var_{\boldsymbol{\alpha}}(\alpha_j^{*(s)}) = \frac{s^2}{(s-1)^2(s-2)}\alpha_j^2,\ s > 2.$

(iii) $MSE_{\boldsymbol{\alpha}}(\alpha_j^{*(s)}) = \frac{s+2}{(s-1)(s-2)}\alpha_j^2,\ 1 \leq j \leq r,\ s > 2.$

Proof. By means of (3.2.1), (i) can readily be computed. Then, (ii) and (iii) follow by (i) and some easy algebra (for the mean and the variance of the inverted gamma distribution see also Kotz & Johnson (1983), p. 259). ■

Theorem 3.2.3
Let sample situation (3.1.11) be given and let $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(s)})$*. Then, for* $s > 1$*, the uniformly minimum variance unbiased estimator (UMVUE) of* α_j *based on the independent observations* $\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(s)}$ *of* $\boldsymbol{X}$ *is given by*

$$\alpha_j^{**(s)} = -\frac{s-1}{T_j^{(s)}(\tilde{\boldsymbol{X}^{(s)}})}, \quad 1 \leq j \leq r.$$

Proof. The statement is obvious from La. 3.1.3 in combination with the Lehmann-Scheffé theorem (cf., e.g., Shao (2003), Thm. 3.1, p. 162) and La. 3.2.2 (i). ■

If we compare the mean squared error of the MLE and UMVUE of α_j based on $s > 1$ independent observations, we obtain

$$Var_{\boldsymbol{\alpha}}(\alpha_j^{**(s)}) < MSE_{\boldsymbol{\alpha}}(\alpha_j^{*(s)}) \quad \Leftrightarrow \quad \frac{\alpha_j^2}{s-2} < \frac{s+2}{(s-1)(s-2)}\alpha_j^2 \quad \Leftrightarrow \quad -1 < 2.$$

Hence, from that point of view, the UMVUE performs better than the MLE as an estimator of α_j. We will demonstrate that, by another approach, the MLE turns out to be the more attractive choice.

Lemma 3.2.4
In the above situation, for $1 \leq j \leq r$ *and* $s > 1$, $\alpha_j^{*(s)}$ *is a Pitman closer estimator of* α_j *than* $\alpha_j^{**(s)}$ *in the sense of (2.2.8).*

Proof. For later use, we first prove assertion

$$(*) \quad \text{For every } a \in \mathbb{R} \text{ holds: } |a| < |a+1| \quad \Leftrightarrow \quad a > -\frac{1}{2}.$$

Thereto:

'$\Leftarrow$' Let $a > -\frac{1}{2}$. Then $-a < \frac{1}{2}$ and $a + \frac{1}{2} > 0$. Thus,

$$|a| = \max\{-a, a\} < \max\left\{\frac{1}{2}, a + \frac{1}{2}\right\} < \frac{1}{2} + a + \frac{1}{2} = a + 1 = |a+1|.$$

'$\Rightarrow$' Let $a \leq -\frac{1}{2}$. Then $a < 0$ and $a + \frac{1}{2} \leq 0$. Thus,

$$|a| = -a = -\left(a + \frac{1}{2}\right) + \frac{1}{2} = \left|a + \frac{1}{2}\right| + \frac{1}{2} \geq \left|a + \frac{1}{2} + \frac{1}{2}\right| = |a+1|.$$

Hence, assertion $(*)$ is shown.

Now, let $\boldsymbol{\alpha} \in \mathbb{R}_+^r$ be the true parameter, and let $s > 1$ and $j \in \{1, \ldots, r\}$ be fixed. We define the random variable $Y_s = -\alpha_j T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ and, thus, by application of Thm. 3.2.1 and Thm. 3.2.3, $\alpha_j^{*(s)} = \frac{s\alpha_j}{Y_s}$ and $\alpha_j^{**(s)} = \frac{(s-1)\alpha_j}{Y_s}$, where $Y_s \sim \Gamma(s, 1)$. Then,

$$\begin{aligned} P_{\boldsymbol{\alpha}}(|\alpha_j^{*(s)} - \alpha_j| < |\alpha_j^{**(s)} - \alpha_j|) &= P_{\boldsymbol{\alpha}}\left(\left|\frac{s\alpha_j}{Y_s} - \alpha_j\right| < \left|\frac{(s-1)\alpha_j}{Y_s} - \alpha_j\right|\right) \\ &= P(|Y_s - s| < |Y_s - s + 1|) \\ &= P\left(Y_s > s - \frac{1}{2}\right), \end{aligned}$$

where the last equation results by application of $(*)$ ($a = Y_s(\omega) - s$). The median m_s of the distribution $\Gamma(s, 1)$ fullfils the inequality $m_s > s - \frac{1}{3}$ (cf. Chen & Rubin (1986) and Choi (1994)). Hence,

$$P_{\boldsymbol{\alpha}}(|\alpha_j^{*(s)} - \alpha_j| < |\alpha_j^{**(s)} - \alpha_j|) = P\left(Y_s > s - \frac{1}{2}\right) > P\left(Y_s > m_s\right) = 0.5.$$

■

Interpreting the above result in terms of absolute frequencies, we find in more than a half of all cases where both estimators are used to obtain an estimate of α_j based on s independent observations that the MLE is closer to α_j than the UMVUE in terms of the absolute value. The following table contains the probabilities $P\left(Y_s > s - \frac{1}{2}\right)$ for $s \in \{2, \ldots, 5, 10, 20, 50, 100, 1000\}$.

s	2	3	4	5	10	20	50	100	1000
$P\left(Y_s > s - \frac{1}{2}\right)$	0.558	0.544	0.537	0.532	0.522	0.515	0.509	0.507	0.502

Applying the multivariate Rao-Cramér inequality (cf. Subsection 2.2.3), the covariance matrix of every unbiased estimator $\hat{\boldsymbol{\alpha}}^{(s)}$ of $\boldsymbol{\alpha}$ based on s independent observations fulfils

$$\mathbf{Cov}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}^{(s)}) \geq \mathbf{I}_f^{(s)}(\boldsymbol{\alpha})^{-1} = \mathbf{diag}\left(\frac{\alpha_1^2}{s}, ..., \frac{\alpha_r^2}{s}\right), \quad \boldsymbol{\alpha} \in \mathbb{R}_+^r,$$

in the sense of the Löwner ordering. This inequality implies that for every unbiased estimator $\hat{\alpha}_j{}^{(s)}$ of α_j based on s independent observations, its variance is bounded from below by

$$Var_{\boldsymbol{\alpha}}(\hat{\alpha}_j^{(s)}) \geq \frac{\alpha_j^2}{s}, \quad \boldsymbol{\alpha} \in \mathbb{R}_+^r, \quad 1 \leq j \leq r.$$

Hence, for $s > 2$, since $Var_{\boldsymbol{\alpha}}(\alpha_j^{**(s)}) > \alpha_j^2/s$, $1 \leq j \leq r$, an efficient estimator of α_j and, thus, of $\boldsymbol{\alpha}$ does not exist.

We extend the multivariate result to the more general case, where $g : \mathbb{R}_+^r \to \mathbb{R}^l$ is an l-dimensional differentiable function of $\boldsymbol{\alpha}$ with Jacobian matrix $\mathbf{D}_g(\boldsymbol{\alpha}) \in \mathbb{R}^{l \times r}$ evaluated at $\boldsymbol{\alpha} \in \mathbb{R}_+^r$. Then, for every unbiased estimator $\hat{g}^{(s)}$ of g based on s independent observations, we obtain

$$\begin{aligned} \mathbf{Cov}_{\boldsymbol{\alpha}}(\hat{g}^{(s)}) &\geq \mathbf{D}_g(\boldsymbol{\alpha})\mathbf{I}_f^{(s)}(\boldsymbol{\alpha})^{-1}\mathbf{D}_g(\boldsymbol{\alpha})' \\ &= \mathbf{D}_g(\boldsymbol{\alpha})\,\mathbf{diag}\left(\frac{\alpha_1^2}{s}, \ldots, \frac{\alpha_r^2}{s}\right)\mathbf{D}_g(\boldsymbol{\alpha})', \end{aligned}$$

for $\boldsymbol{\alpha} \in \mathbb{R}_+^r$ in the sense of the Löwner ordering.

We end this subsection by pointing out that the reciprocal values of the α's can be estimated efficiently.

Lemma 3.2.5
In the above situation, we find:

(i) $-s^{-1}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *is an efficient estimator of* $(\alpha_1^{-1}, \ldots, \alpha_r^{-1})'$, *i.e.* $-s^{-1}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ *has uniformly minimal covariance matrix in the sense of the Löwner ordering among all unbiased estimators of* $(\alpha_1^{-1}, \ldots, \alpha_r^{-1})'$ *based on* s *independent observations. The lower bound of the Rao-Cramér inequality is attained at* $\mathbf{diag}(s^{-1}\alpha_1^{-2}, \ldots, s^{-1}\alpha_r^{-2})$.

(ii) $-s^{-1}T_j(\tilde{\boldsymbol{X}}^{(s)})$ *is an efficient estimator and, thus, the UMVUE of* α_j^{-1}, $1 \leq j \leq r$, *i.e. it has minimal variance among all unbiased estimators of* α_j^{-1} *based on* s *independent observations. The lower bound of the Rao-Cramér inequality is attained at* $s^{-1}\alpha_j^{-2}$.

Proof. Both assertions are obvious from La. 2.2.10. ■

3.2.2 Asymptotic Properties

As a useful property, strong consistency of the sequences of MLEs and UMVUEs can readily be seen.

Theorem 3.2.6
In the above situation, we find:

(i) The sequence of MLEs $\{\boldsymbol{\alpha}^{*(s)}\}_{s\in\mathbb{N}}$ *and the sequence* $\{\boldsymbol{\alpha}^{**(s)}\}_{s\in\mathbb{N}}$, $\boldsymbol{\alpha}^{**(s)} = (\alpha_1^{**(s)}, ..., \alpha_r^{**(s)})'$, $s \in \mathbb{N}$, *of UMVUEs are strongly consistent for estimating* $\boldsymbol{\alpha}$*. Moreover, if* g *in Thm. 3.2.1* (i) *is continuous, the sequences* $\{g(\boldsymbol{\alpha}^{*(s)})\}_{s\in\mathbb{N}}$ *and* $\{g(\boldsymbol{\alpha}^{**(s)})\}_{s\in\mathbb{N}}$ *are strongly consistent for estimating* $g(\boldsymbol{\alpha})$*.*

(ii) The sequence of MLEs $\{\alpha_j^{*(s)}\}_{s\in\mathbb{N}}$ *and the sequence of UMVUEs* $\{\alpha_j^{**(s)}\}_{s\in\mathbb{N}}$ *of* α_j *are strongly consistent for estimating* α_j, $1 \leq j \leq r$*. Moreover, if* g *in Thm. 3.2.1* (ii) *is continuous, the sequences* $\{g(\alpha_j^{*(s)})\}_{s\in\mathbb{N}}$ *and* $\{g(\alpha_j^{**(s)})\}_{s\in\mathbb{N}}$ *are strongly consistent for estimating* $g(\alpha_j)$, $1 \leq j \leq r$*.*

Proof. Application of La. 2.2.12 yields strong consistency of $\{\boldsymbol{\alpha}^{*(s)}\}_{s\in\mathbb{N}}$. The respective result for the other sequences, in particular, for the univariate conclusions, are then evident. ■

Moreover, asymptotic efficiency is easily addressed.

Theorem 3.2.7
In the above situation, we find:

(i) The sequence $\{\boldsymbol{\alpha}^{*(s)}\}_{s\in\mathbb{N}}$ *of MLEs and the sequence* $\{\boldsymbol{\alpha}^{**(s)}\}_{s\in\mathbb{N}}$ *of UMVUEs are asymptotically efficient for estimating* $\boldsymbol{\alpha}$*, i.e., we have*

$$\begin{aligned}\sqrt{s}(\boldsymbol{\alpha}^{*(s)} - \boldsymbol{\alpha}) &\xrightarrow{\mathcal{D}} \mathcal{N}_r(\mathbf{0}, \mathbf{diag}(\alpha_1^2, ..., \alpha_r^2)),\\ \sqrt{s}(\boldsymbol{\alpha}^{**(s)} - \boldsymbol{\alpha}) &\xrightarrow{\mathcal{D}} \mathcal{N}_r(\mathbf{0}, \mathbf{diag}(\alpha_1^2, ..., \alpha_r^2))\,.\end{aligned}$$

Moreover, if in Thm. 3.2.1 (i) g *is continuously differentiable with* $|\mathbf{D}_g(\boldsymbol{\alpha})| \neq 0\ \forall \boldsymbol{\alpha} \in \mathbb{R}_+^r$, *then the sequences* $\{g(\boldsymbol{\alpha}^{*(s)})\}_{s\in\mathbb{N}}$ *and* $\{g(\boldsymbol{\alpha}^{**(s)})\}_{s\in\mathbb{N}}$ *are asymptotically efficient for estimating* $g(\boldsymbol{\alpha})$*, i.e., e.g.,*

$$\sqrt{s}(g(\boldsymbol{\alpha}^{*(s)}) - g(\boldsymbol{\alpha})) \xrightarrow{\mathcal{D}} \mathcal{N}_r(\mathbf{0}, \mathbf{D}_g(\boldsymbol{\alpha})\mathbf{diag}(\alpha_1^2, ..., \alpha_r^2)\mathbf{D}_g(\boldsymbol{\alpha})').$$

(ii) The sequences of estimators $\{\alpha_j^{*(s)}\}_{s\in\mathbb{N}}$ *and* $\{\alpha_j^{**(s)}\}_{s\in\mathbb{N}}$ *for estimating* α_j, $1 \leq j \leq r$, *are asymptotically efficient for estimating* α_j*, i.e.,*

$$\begin{aligned}\sqrt{s}(\alpha_j^{*(s)} - \alpha_j) &\xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha_j^2),\\ \sqrt{s}(\alpha_j^{**(s)} - \alpha_j) &\xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha_j^2)\,.\end{aligned}$$

Moreover, if in Thm. 3.2.1 (ii) g is continuously differentiable with $g'(\alpha_j) \neq 0 \; \forall \boldsymbol{\alpha_j} \in \mathbb{R}_+$, then the sequences $\{g(\alpha_j^{*(s)})\}_{s\in\mathbb{N}}$ and $\{g(\alpha_j^{**(s)})\}_{s\in\mathbb{N}}$ are asymptotically efficient for estimating $g(\alpha_j)$, i.e., e.g.,

$$\sqrt{s}(g(\alpha_j^{*(s)}) - g(\alpha_j)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, (g'(\alpha_j))^2\, \alpha_j^2).$$

Proof. Both assertions follow by application of La. 2.2.15 and the multivariate version of Slutsky's theorem (cf., e.g., Sen & Singer (1993), Thm. 3.4.3, p. 130). ∎

In particular, sequences of the estimators presented in La. 3.2.5 are strongly consistent (and asymptotically efficient) for estimating the respective reciprocals of the α's.
When the model of SOSs with conditional proportional hazard rates is used to describe (possibly sequential) systems, aiming at detecting increasing load put on remaining components, several choices of g are near at hand, e.g.,

$$\begin{aligned} g_1(\boldsymbol{\alpha}) &= (\alpha_1, \alpha_2 - \alpha_1, \ldots, \alpha_r - \alpha_{r-1})', \\ g_2(\boldsymbol{\alpha}) &= \left(\alpha_1, \frac{\alpha_2}{\alpha_1}, \ldots, \frac{\alpha_r}{\alpha_{r-1}}\right)', \quad \text{or} \\ g_3(\boldsymbol{\alpha}) &= \left(\alpha_1, \frac{\alpha_2}{\alpha_1}, \ldots, \frac{\alpha_r}{\alpha_1}\right)', \quad \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)' \in \mathbb{R}_+^r. \end{aligned}$$

3.2.3 Estimation under Simple Order Restriction

When the model of SOSs with conditional proportional hazard rates is used to describe and analyse sequential systems, the additional assumption that the α's are arranged in ascending order of magnitude can be made. In that context, the simple order restriction $\alpha_1 \leq \cdots \leq \alpha_r$ on the model parameters is naturally justified in order to model increasing stress on remaining components upon failure of some component of the system. Hence, the prior information of ordered model parameters should be taken into account when these parameters are estimated. In Balakrishnan et al. (2008), MLEs of the model parameters under simple order restriction have been established. Making use of the exponential family structure, this result can also be obtained from Thm. 2.2.8 as we will show.

Theorem 3.2.8
The MLE of $\boldsymbol{\alpha}$ under the simple order restriction $\alpha_1 \leq \cdots \leq \alpha_r$ based on $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)})$ is uniquely determined and given by $\boldsymbol{\alpha}_{\leq}^{*(s)}$ with components

$$(\boldsymbol{\alpha}_{\leq}^{*(s)})_j = \min_{j\leq\mu\leq r} \max_{1\leq\nu\leq j} -\frac{s(\mu-\nu+1)}{\sum_{l=\nu}^{\mu} T_l^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \quad 1 \leq j \leq r.$$

Proof. Let $\tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)})$ be a realization of $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)})$. We have seen that the log likelihood function based on $\tilde{\boldsymbol{x}}^{(s)}$ is strictly concave on $\mathbb{R}_+^r$ (see, e.g., the proof of La. 2.2.1 in the Appendix), and, thus, in particular, on the convex subset $\Theta_{\leq} = \{\boldsymbol{\alpha} \in \mathbb{R}_+^r : \alpha_1 \leq \cdots \leq \alpha_r\}$ of

$\mathbb{R}^r_+$. Hence, if the restricted MLE of $\boldsymbol{\alpha}$ based on $\tilde{\boldsymbol{x}}^{(s)}$ in $\Theta_\leq$ exists, it is uniquely determined. Moreover, we have already shown, that the random variables $T_j(\boldsymbol{X}^{(i)})$, $1 \leq j \leq r$, $1 \leq i \leq s$, are independent, and $-T_j(\boldsymbol{X}^{(1)}), \dots, -T_j(\boldsymbol{X}^{(s)})$ have an exponential distribution with scale parameter $\alpha_j^{-1} > 0$, $1 \leq j \leq r$. Hence, by reparametrization of the model parameters via $\tilde{\alpha}_j = \alpha_{r-j+1}^{-1}$, $1 \leq j \leq r$, for fixed $1 \leq j \leq r$, the transformed observations $t_{i;j} = T_j(\boldsymbol{x}^{(i)})$, $1 \leq i \leq s$, are realizations of s iid random variables having λ^1-density $\tilde{f}_{\tilde{\alpha}_{r-j+1}}$, where

$$\tilde{f}_{\tilde{\alpha}}(t) = e^{\zeta(\tilde{\alpha})(-t)-\kappa(\tilde{\alpha})} \mathbb{1}_{(-\infty,0)}(t)$$

with $\zeta(\tilde{\alpha}) = -\tilde{\alpha}^{-1}$ and $\kappa(\tilde{\alpha}) = \ln(\tilde{\alpha})$, $\tilde{\alpha} > 0$. Obviously, $\zeta'(\tilde{\alpha}) = \tilde{\alpha}^{-2} > 0$ and $\kappa'(\tilde{\alpha}) = \tilde{\alpha}^{-1} = \tilde{\alpha}\,\zeta'(\tilde{\alpha})$. Now, a joint density of all rs random variables is given by

$$\tilde{f}^{(rs)}_{\tilde{\boldsymbol{\alpha}}}(\tilde{\boldsymbol{t}}^{(rs)}) = \prod_{j=1}^{r}\prod_{i=1}^{s} \tilde{f}_{\tilde{\alpha}_{r-j+1}}(t_{i;j}), \quad \tilde{\boldsymbol{t}}^{(rs)} = (t_{1;1}, \dots, t_{s;1}, \dots, t_{1;r}, \dots, t_{s;r}),$$

where $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_r)'$. Notice, that $\tilde{f}^{(rs)}_{\tilde{\boldsymbol{\alpha}}}(\tilde{\boldsymbol{t}}^{(rs)}) \prod_{i=1}^s h(\boldsymbol{x}^{(i)}) = f^{(s)}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{x}}^{(s)})$. It is easily shown that the (unrestricted) MLE of $\tilde{\boldsymbol{\alpha}}$ based on $\tilde{\boldsymbol{t}}^{(rs)}$ exists and is given by $(-s^{-1}t_r^{(s)}, \dots, -s^{-1}t_1^{(s)})'$, where $t_j^{(s)} = \sum_{i=1}^s t_{i;j}$, $1 \leq j \leq r$. Hence, the assumptions from Thm. 2.2.8 are fulfilled, and we obtain that $\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq}$ is the MLE of $\tilde{\boldsymbol{\alpha}}$ based on $\tilde{\boldsymbol{t}}^{(rs)}$ with respect to the constraint $\tilde{\alpha}_1 \leq \dots \leq \tilde{\alpha}_r$, where the components of $\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq}$ are given by

$$(\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq})_j = \max_{1\leq\mu\leq j}\min_{j\leq\nu\leq r} -\frac{1}{s(\nu-\mu+1)}\sum_{l=\mu}^{\nu} t^{(s)}_{r-l+1}, \quad 1 \leq j \leq r.$$

The mapping $g:\ \Theta_\leq \to \Theta_\leq:\ \boldsymbol{\alpha} \mapsto \tilde{\boldsymbol{\alpha}}$ is bijective, and it follows

$$\begin{aligned}
& \tilde{f}^{(rs)}_{\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq}}(\tilde{\boldsymbol{t}}^{(rs)}) \geq \tilde{f}^{(rs)}_{\tilde{\boldsymbol{\alpha}}}(\tilde{\boldsymbol{t}}^{(rs)}) \quad \forall \tilde{\boldsymbol{\alpha}} \in \Theta_\leq \\
\Leftrightarrow \quad & f^{(s)}_{g^{-1}(\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq})}(\tilde{\boldsymbol{x}}^{(s)}) \geq f^{(s)}_{g^{-1}(\tilde{\boldsymbol{\alpha}})}(\tilde{\boldsymbol{x}}^{(s)}) \quad \forall \tilde{\boldsymbol{\alpha}} \in \Theta_\leq \\
\Leftrightarrow \quad & f^{(s)}_{g^{-1}(\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq})}(\tilde{\boldsymbol{x}}^{(s)}) \geq f^{(s)}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{x}}^{(s)}) \quad \forall \boldsymbol{\alpha} \in \Theta_\leq.
\end{aligned}$$

Hence, with regard to the first part of the proof, $\boldsymbol{\alpha}^{*(s)}_{\leq} = g^{-1}(\tilde{\boldsymbol{\alpha}}^{*(rs)}_{\leq})$ is the unique MLE of $\boldsymbol{\alpha}$ under the simple order restriction $\alpha_1 \leq \dots \leq \alpha_r$ based on the observation $\tilde{\boldsymbol{x}}^{(s)}$, where the components of

$\boldsymbol{\alpha}_{\leq}^{*(s)}$ are given by

$$\begin{aligned}(\boldsymbol{\alpha}_{\leq}^{*(s)})_j &= ((\tilde{\alpha}_{\leq}^{*(rs)})_{r-j+1})^{-1} \\ &= \left(\max_{1\leq\mu\leq r-j+1} \min_{r-j+1\leq\nu\leq r} - \frac{1}{s(\nu-\mu+1)} \sum_{l=\mu}^{\nu} T_{r-l+1}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \right)^{-1} \\ &= \left(\max_{j\leq\mu\leq r} \min_{1\leq\nu\leq j} - \frac{1}{s(\mu-\nu+1)} \sum_{l=\nu}^{\mu} T_{l}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \right)^{-1} \\ &= \min_{j\leq\mu\leq r} \max_{1\leq\nu\leq j} - \frac{s(\mu-\nu+1)}{\sum_{l=\nu}^{\mu} T_l^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}, \quad 1\leq j\leq r.\end{aligned}$$

■

By noticing that $-\sum_{l=\nu}^{\mu} s^{-1}T_l^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \to \sum_{l=\nu}^{\mu} \alpha_l^{-1}$ $P_{\boldsymbol{\alpha}}$-a.s. for $1\leq\nu\leq\mu\leq r$, it follows that $\boldsymbol{\alpha}_{\leq}^{*(s)} \to \boldsymbol{\alpha}$ $P_{\boldsymbol{\alpha}}$-a.s. for $\boldsymbol{\alpha}\in\Theta_{\leq}$, i.e. strong consistency of the sequence $\{\boldsymbol{\alpha}_{\leq}^{*(s)}\}_{s\in\mathbb{N}}$ for estimating $\boldsymbol{\alpha}$ provided that $\alpha_1\leq\cdots\leq\alpha_r$.

3.3 Statistical Tests on Model Parameters

In this section, we focus on univariate and multivariate statistical tests on the model parameters $\alpha_1, \ldots, \alpha_r$. Throughout, if not otherwise specified, we consider sampling situation (3.1.11) with $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(s)}$ being realizations of s independent random vectors $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)}$ having density $f_{\boldsymbol{\alpha}}^{\boldsymbol{X}}$.
In Subsections 3.3.1 and 3.3.2, UMPU level-α tests on single parameters are established, where a level-α test φ is called *unbiased* if its power is bounded from below by α, i.e., $E_{\boldsymbol{\alpha}}[\varphi] \geq \alpha$, if the alternative is true. On the one hand, these tests can directly be interpreted in terms of the hazard rate $\alpha_j f/(1-F)$ of the remaining components of the system upon occurrence of the $(j-1)^{\text{th}}$ failure of some component. On the other hand, they can also be used for model checking in the following sense. Consider, e.g., a (possibly sequential) 3-out-of-4 system and the null hypothesis $H_0: \ \alpha_2 = 1$, which is tested against the alternative $H_1: \ \alpha_2 \neq 1$, where, based on some prior information, $\alpha_1 = 1$ is assumed. If the alternative is accepted on the basis of some experiment, the assumption of a common 3-out-of-4 system can no longer be maintained, and the system is supposed to be of the more general sequential type. Hence, in that case, the model of SOSs is more appropriate (than the model of OSs) for describing and analysing the structure of the system.
In Subsections 3.3.3 and 3.3.4, model tests are proposed for the case, where no prior information on the model parameters is available. For simple and composite null hypotheses, the test statistics of the LR test, Wald's test and Rao's score test are presented, and asymptotic properties of the tests are derived. In case of a simple null hypothesis, asymptotic optimality of the sequence of LR tests is obtained (Bahadur sense, Kallenberg sense).
In Subsection 3.3.5, multivariate tests are discussed, where some of the α's are considered as fixed nuisance parameters.
Finally, in Subsection 3.3.6, in the context of test problems with hypotheses on the simple ordering of the model parameters, the asymptotic distribution of the LR test statistic under the null hypothesis is considered.

3.3.1 One-sided Test Problems

We consider the two following one-sided test problems concerning α_1. Notice that all statements of this subsection are also valid for every other choice of α_j, $2 \leq j \leq r$. We will go into that point in Rem. 3.3.10.

(I) $H_0: \ \alpha_1 \leq \alpha_0 \quad \leftrightarrow \quad H_1: \ \alpha_1 > \alpha_0$,

(II) $H_0: \ \alpha_1 \geq \alpha_0 \quad \leftrightarrow \quad H_1: \ \alpha_1 < \alpha_0$,

where α_0 is a positive constant. Throughout this subsection, for a short notation and a better reading, $P_{\boldsymbol{\alpha}}^{\tilde{\boldsymbol{X}}^{(s)}}$ is replaced by $P_{\boldsymbol{\alpha}}$.

Theorem 3.3.1
For $\alpha \in (0,1)$, $\alpha_0 \in \mathbb{R}_+$, *and the test problem* $H_0: \ \alpha_1 \leq \alpha_0 \ \leftrightarrow \ H_1: \ \alpha_1 > \alpha_0$,

$$\varphi^*: \quad (\mathbb{R}^r_<)^{1\times s} \to [0,1]: \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto \mathbb{1}_{\left(-\frac{\chi^2_\alpha(2s)}{2\alpha_0}, \infty\right)}(T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on s independent observations of $\boldsymbol{X}$, where $\chi^2_\alpha(2s)$ denotes the α-quantile of the χ^2-distribution with $2s$ degrees of freedom.

Proof. Let $\boldsymbol{V} = (T_2^{(s)}, \dots, T_r^{(s)})'$. Applying Thm 2.3.1, a UMPU level-α test is given by $\varphi^* = \Psi^* \circ (T_1^{(s)}, \boldsymbol{V})$, $\Psi^*(u, \boldsymbol{v}) = \mathbb{1}_{(c(\boldsymbol{v}),\infty)}(u) + \gamma(\boldsymbol{v})\mathbb{1}_{\{c(\boldsymbol{v})\}}(u)$, where $c, \gamma : (\mathbb{R}^{r-1}, \mathbb{B}^{r-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$, $0 \le \gamma \le 1$, fulfill

$$P_{\alpha_0,\bullet}^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}((c(\boldsymbol{v}), \infty)) + \gamma(\boldsymbol{v}) P_{\alpha_0,\bullet}^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}(\{c(\boldsymbol{v})\}) \stackrel{!}{=} \alpha.$$

Since the conditional distribution of $T_1^{(s)}$ given $(T_2^{(s)}, \dots, T_r^{(s)})' = \boldsymbol{v}$ only depends on α_0 if $\boldsymbol{\alpha} = (\alpha_0, \alpha_2, \dots, \alpha_r)'$ is true (cf., e.g., Lehmann & Romano (2005), La. 2.7.2, p. 48), we omit $\alpha_2, \dots, \alpha_r$ and use the expression $P_{\alpha_0,\bullet}^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})=\boldsymbol{v}}$.
$T_1^{(s)}, \dots, T_r^{(s)}$ are jointly independent and $-T_1^{(s)} \sim \Gamma(s, \alpha_0^{-1})$ is, in particular, continuously distributed if $\alpha_1 = \alpha_0$ is true. Thus, the problem simplifies to finding the constant (function) $c \equiv c(\boldsymbol{v})$, $\boldsymbol{v} \in \mathbb{R}^{r-1}$, with $\alpha \stackrel{!}{=} P_{\alpha_0,\bullet}(T_1^{(s)} > c)$.
Since $-2\alpha_0 T_1^{(s)} \sim \chi^2(2s)$, the proof is completed. ■

Analogously, we obtain for test problem (II) the following theorem.

Theorem 3.3.2
For $\alpha \in (0, 1)$, $\alpha_0 \in \mathbb{R}_+$, and the test problem $H_0 : \alpha_1 \ge \alpha_0 \leftrightarrow H_1 : \alpha_1 < \alpha_0$,

$$\varphi^* : \quad (\mathbb{R}^r_<)^{1\times s} \to [0, 1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto \mathbb{1}_{\left(-\infty, -\frac{\chi^2_{1-\alpha}(2s)}{2\alpha_0}\right)}(T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on s independent observations of $\boldsymbol{X}$, where $\chi^2_{1-\alpha}(2s)$ denotes the $(1-\alpha)$-quantile of the χ^2-distribution with $2s$ degrees of freedom.

Proof. Applying Thm. 2.3.2, the assertion can be shown by the same arguments as in the proof of Thm. 3.3.1. ■

3.3.2 Two-sided Test Problems

We continue by considering the three following two-sided test problems concerning α_1. As in Subsection 3.3.1, all statements of this section are also valid for every other choice of α_j, $2 \le j \le r$, and, again, $P_{\boldsymbol{\alpha}}^{\tilde{\boldsymbol{X}}^{(s)}}$ is replaced by $P_{\boldsymbol{\alpha}}$.

(III) $H_0 : \alpha_1 = \alpha_0 \quad \leftrightarrow \quad H_1 : \alpha_1 \ne \alpha_0$,

(IV) $H_0 : \alpha_0^{(1)} \le \alpha_1 \le \alpha_0^{(2)} \quad \leftrightarrow \quad H_1 : \alpha_1 < \alpha_0^{(1)}$ or $\alpha_1 > \alpha_0^{(2)}$,

(V) $H_0 : \alpha_1 \le \alpha_0^{(1)}$ or $\alpha_1 \ge \alpha_0^{(2)} \quad \leftrightarrow \quad H_1 : \alpha_0^{(1)} < \alpha_1 < \alpha_0^{(2)}$, where $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ are positive constants with $\alpha_0^{(1)} < \alpha_0^{(2)}$.

In the following, $F_{a,n}$ denotes the distribution function of the Erlang distribution $\Gamma(n, a^{-1})$ with shape parameter $n \in \mathbb{N}$ and scale parameter $a^{-1} > 0$, i.e.

$$F_{a,n}(x) = 1 - e^{-ax} \sum_{j=0}^{n-1} \frac{(ax)^j}{j!}, \quad x > 0,$$

with corresponding density function

$$f_{a,n}(x) = \frac{a^n}{(n-1)!} x^{n-1} e^{-ax} \mathbb{1}_{(0,\infty)}(x).$$

For every $a > 0$ and $n \in \mathbb{N}$, $F_{a,n}$ is strictly increasing on $(0, \infty)$. Hence, its quantile function $F_{a,n}^{-1}$ coincides with its continuously differentiable inverse function on $(0,1)$, and $F_{a,n}^{-1}(0) = \lim_{x \searrow 0} F_{a,n}^{-1}(x) = 0$, $F_{a,n}^{-1}(1) = \lim_{x \nearrow 1} F_{a,n}^{-1}(x) = \infty$. Moreover, $F_{a,n}(x)$ is strictly increasing in a for fixed $n \in \mathbb{N}$ and $x > 0$ and strictly decreasing in n for fixed $a > 0$ and $x > 0$.
For $\alpha \in (0,1)$, $\alpha_0 \in \mathbb{R}_+$, and $s \in \mathbb{N}$, we introduce the mapping

$$\begin{aligned} \tau^{(\alpha,\alpha_0,s)} : \ (0,\alpha) &\rightarrow \{(c,d) \in \mathbb{R}_-^{1\times 2} : \ F_{\alpha_0,s}(-c) - F_{\alpha_0,s}(-d) = 1-\alpha\} : \\ \beta &\mapsto (\tau_1^{(\alpha,\alpha_0,s)}(\beta), \tau_2^{(\alpha,\alpha_0,s)}(\beta)) = (-F_{\alpha_0,s}^{-1}(1-\alpha+\beta), -F_{\alpha_0,s}^{-1}(\beta)). \quad (3.3.1) \end{aligned}$$

Corollary 3.3.3
For every $\alpha \in (0,1)$, $\alpha_0 \in \mathbb{R}_+$ and $s \in \mathbb{N}$, $\tau^{(\alpha,\alpha_0,s)}$ is well-defined and bijective.

Proof. Let $\alpha \in (0,1)$, $\alpha_0 \in \mathbb{R}_+^r$ and $s \in \mathbb{N}$ be fixed. For brevity, let $\tau = \tau^{(\alpha,\alpha_0,s)}$ and $F = F_{\alpha_0,s}$. Then, for every $\beta \in (0,\alpha)$, it follows that $\tau_1(\beta), \tau_2(\beta) < 0$ and

$$\begin{aligned} F(-\tau_1(\beta)) - F(-\tau_2(\beta)) &= F(-(-F^{-1}(1-\alpha+\beta))) - F(-(-F^{-1}(\beta))) \\ &= 1 - \alpha + \beta - \beta = 1 - \alpha \end{aligned}$$

and, thus, τ is well-defined.
Let $\tau(\beta_1) = \tau(\beta_2)$ for some $\beta_1, \beta_2 \in (0,\alpha)$. Then,

$$\beta_1 = F(-(-F^{-1}(\beta_1))) = F(-\tau_2(\beta_1)) = F(-\tau_2(\beta_2)) = F(-(-F^{-1}(\beta_2))) = \beta_2$$

yields that τ is injective.
For $(c,d) \in \mathbb{R}_-^{1\times 2}$ with $F(-c) - F(-d) = 1-\alpha$, set $\beta_0 = F(-d)$. Since $d < 0$, it follows that $\beta_0 > 0$ and

$$\beta_0 = F(-c) - (1-\alpha) < 1 - (1-\alpha) = \alpha.$$

We conclude that $\beta_0 \in (0,\alpha)$. Moreover,

$$\begin{aligned} \tau(\beta_0) &= (\tau_1(\beta_0), \tau_2(\beta_0)) = (-F^{-1}(1-\alpha+\beta_0), -F^{-1}(\beta_0)) \\ &= (-F^{-1}(1-\alpha+F(-c)-(1-\alpha)), -F^{-1}(F(-d))) = (c,d). \end{aligned}$$

Hence, τ is surjective and the proof completed. ■

With these denotations, we turn towards the two-sided test problem (III).

Theorem 3.3.4
For $\alpha \in (0,1)$, $\alpha_0 \in \mathbb{R}_+$, and the test problem $H_0:\ \alpha_1 = \alpha_0 \ \leftrightarrow\ H_1:\ \alpha_1 \neq \alpha_0$,

$$\varphi^*: \quad (\mathbb{R}^r_<)^{1\times s} \to [0,1]: \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto 1 - \mathbb{1}_{\left(\tau_1^{(\alpha,\alpha_0,s)}(\beta^*),\tau_2^{(\alpha,\alpha_0,s)}(\beta^*)\right)}(T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on s independent observations of $\boldsymbol{X}$, where β^* is the unique solution of the equation

$$F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(1-\alpha+\beta)) - F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(\beta)) \stackrel{!}{=} 1-\alpha \tag{3.3.2}$$

with respect to $\beta \in (0,\alpha)$.

We state some remarks and, subsequently, prove the theorem.

Remark 3.3.5
It is easily shown that there exists exactly one solution of (3.3.2). For fixed $\alpha \in (0,1)$, $\alpha_0 \in \mathbb{R}_+$ and $s \in \mathbb{N}$, we define the mapping

$$g(\beta) = F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(1-\alpha+\beta)) - F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(\beta)), \quad \beta \in (0,\alpha).$$

Then,

$$\begin{aligned}\lim_{\beta\searrow 0} g(\beta) &= F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(1-\alpha)) < F_{\alpha_0,s}(F^{-1}_{\alpha_0,s}(1-\alpha)) = 1-\alpha,\\ \lim_{\beta\nearrow\alpha} g(\beta) &= 1 - F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(\alpha)) > 1 - F_{\alpha_0,s}(F^{-1}_{\alpha_0,s}(\alpha)) = 1-\alpha.\end{aligned}$$

Obviously, g is differentiable and

$$\begin{aligned} g'(\beta) > 0 \quad &\Leftrightarrow \quad \frac{f_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(1-\alpha+\beta))}{f_{\alpha_0,s}(F^{-1}_{\alpha_0,s}(1-\alpha+\beta))} - \frac{f_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(\beta))}{f_{\alpha_0,s}(F^{-1}_{\alpha_0,s}(\beta))} > 0\\ &\Leftrightarrow \quad \frac{f_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(1-\alpha+\beta))}{f_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(\beta))} > \frac{f_{\alpha_0,s}(F^{-1}_{\alpha_0,s}(1-\alpha+\beta))}{f_{\alpha_0,s}(F^{-1}_{\alpha_0,s}(\beta))}\\ &\Leftrightarrow \quad \frac{F^{-1}_{\alpha_0,s}(1-\alpha+\beta)}{F^{-1}_{\alpha_0,s}(\beta)} > 1 \quad \Leftrightarrow \quad 1-\alpha > 0.\end{aligned}$$

Hence, applying the intermediate value theorem, there exists one and only one solution of $g(\beta) = 1-\alpha$.
Moreover, the expression on the left-hand side of (3.3.2) and, hence, its solution β^* do not depend on the value of α_0. For this, note that for $y > 0$ and $s \in \mathbb{N}$ $F_{\alpha_0,s}(y) = F_{1,s}(\alpha_0 y)$ and, therewith, for $u \in (0,1)$, $F_{1,s}(\alpha_0 F^{-1}_{\alpha_0,s}(u)) = u$ and $\alpha_0 F^{-1}_{\alpha_0,s}(u) = F^{-1}_{1,s}(u)$ hold. Then,

$$F_{\alpha_0,s+1}(F^{-1}_{\alpha_0,s}(u)) = F_{1,s+1}(\alpha_0 F^{-1}_{\alpha_0,s}(u)) = F_{1,s+1}(F^{-1}_{1,s}(u)),$$

regardless of the value of α_0.
Thus, w.l.o.g., $\alpha_0 = 1$ can be assumed. Then, for every choice of $\alpha \in (0,1)$ and $s \in \mathbb{N}$, β^* can be

obtained numerically, e.g, with Newton's procedure. For that, note that the mapping $N^{(\alpha,s)}$ defined by

$$N^{(\alpha,s)}(\beta) = F_{1,s+1}(F_{1,s}^{-1}(1-\alpha+\beta)) - F_{1,s+1}(F_{1,s}^{-1}(\beta)) - (1-\alpha), \quad \beta \in (0,\alpha), \tag{3.3.3}$$

is differentiable with respect to β and its derivative is given by

$$\frac{d}{d\beta}N^{(\alpha,s)}(\beta) = \frac{1}{s}\left(F_{1,s}^{-1}(1-\alpha+\beta) - F_{1,s}^{-1}(\beta)\right), \quad \beta \in (0,\alpha).$$

In the following, we provide a proof of Thm. 3.3.4.

Proof of Thm. 3.3.4 Let $\boldsymbol{V} = (T_2^{(s)}, \dots, T_r^{(s)})'$. Applying Thm. 2.3.3, a UMPU level-α test is given by $\varphi^* = \Psi^* \circ (T_1^{(s)}, \boldsymbol{V})$, $\Psi^*(u,\boldsymbol{v}) = 1 - \mathbb{1}_{[c_1(\boldsymbol{v}),c_2(\boldsymbol{v})]}(u) + \sum_{i=1}^2 \gamma_i(\boldsymbol{v})\mathbb{1}_{\{c_i(\boldsymbol{v})\}}(u)$, where $c_1, c_2, \gamma_1, \gamma_2 : (\mathbb{R}^{r-1}, \mathbb{B}^{r-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$, $0 \le \gamma_1, \gamma_2 \le 1$, are such that

$$\begin{aligned} & \int \Psi^*(u,\boldsymbol{v}) dP_{\alpha_0,\bullet}^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}(u) \stackrel{!}{=} \alpha \\ \text{and} \quad & \int u\,\Psi^*(u,\boldsymbol{v}) dP_{\alpha_0,\bullet}^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}(u) \stackrel{!}{=} \alpha \int u\, dP_{\alpha_0,\bullet}^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}(u). \end{aligned}$$

$T_1^{(s)}, \dots., T_r^{(s)}$ are jointly independent and $-T_1^{(s)} \sim \Gamma(s, \alpha_0^{-1})$ is, in particular, continuously distributed if $\alpha_1 = \alpha_0$ is true. Thus, the problem left is to find the constants, respectively constant functions, $c_1 \equiv c_1(\boldsymbol{v})$ and $c_2 \equiv c_2(\boldsymbol{v})$, $\boldsymbol{v} \in \mathbb{R}^{r-1}$, with

$$\int_{c_1}^{c_2} dP_{\alpha_0,\bullet}^{T_1^{(s)}}(u) \stackrel{!}{=} 1-\alpha \quad \text{and} \quad \int_{c_1}^{c_2} u\, dP_{\alpha_0,\bullet}^{T_1^{(s)}}(u) \stackrel{!}{=} -\frac{s(1-\alpha)}{\alpha_0}.$$

Since

$$\begin{aligned} -\frac{\alpha_0}{s}\int_{c_1}^{c_2} u\, dP_{\alpha_0,\bullet}^{T_1^{(s)}}(u) &= -\frac{\alpha_0}{s}\int_{c_1}^{c_2} \frac{\alpha_0^s}{(s-1)!} u(-u)^{s-1} e^{\alpha_0 u}\mathbb{1}_{(-\infty,0)}(u)du \\ &= \int_{c_1}^{c_2} \frac{\alpha_0^{s+1}}{s!}(-u)^s e^{\alpha_0 u}\mathbb{1}_{(-\infty,0)}(u)du, \end{aligned}$$

we have to solve simultaneously

$$1-\alpha \stackrel{!}{=} P(-c_2 < \Gamma(s, \alpha_0^{-1}) < -c_1)$$

and

$$1-\alpha \stackrel{!}{=} P(-c_2 < \Gamma(s+1, \alpha_0^{-1}) < -c_1),$$

or, with the denotations introduced above,

$$1-\alpha \stackrel{!}{=} F_{\alpha_0,s}(-c_1) - F_{\alpha_0,s}(-c_2) \tag{3.3.4}$$

and

$$1-\alpha \overset{!}{=} F_{\alpha_0,s+1}(-c_1) - F_{\alpha_0,s+1}(-c_2). \qquad (3.3.5)$$

W.l.o.g., let $c_1 < c_2 < 0$ (otherwise a simultaneous solution of (3.3.4) and (3.3.5) do not exist). Then, setting $(c_1, c_2) = \tau^{(\alpha,\alpha_0,s)}(\beta^*)$ yields the assertion. ■

We now turn to tests with interval hypotheses.

Theorem 3.3.6
For $\alpha \in (0,1)$, $\alpha_0^{(1)}, \alpha_0^{(2)} \in \mathbb{R}_+$ with $\alpha_0^{(1)} < \alpha_0^{(2)}$, and the test problem $H_0 : \alpha_0^{(1)} \leq \alpha_1 \leq \alpha_0^{(2)} \leftrightarrow H_1 : \alpha_1 < \alpha_0^{(1)}$ or $\alpha_1 > \alpha_0^{(2)}$,

$$\varphi^* : \quad (\mathbb{R}^r_<)^{1\times s} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto 1 - \mathbb{1}_{\left(\tau_1^{(\alpha,\alpha_0^{(1)},s)}(\beta^*), \tau_2^{(\alpha,\alpha_0^{(1)},s)}(\beta^*)\right)}(T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on s independent observations of $\boldsymbol{X}$, where β^* is the unique solution of the equation

$$F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta)) - F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta)) \overset{!}{=} 1-\alpha \qquad (3.3.6)$$

with respect to $\beta \in (0,\alpha)$.

Proof. Let $\boldsymbol{V} = (T_2^{(s)}, \dots, T_r^{(s)})'$. Applying Thm. 2.3.4, a UMPU level-α test is given by $\varphi^* = \Psi^* \circ (T_1^{(s)}, \boldsymbol{V})$, $\Psi^*(u, \boldsymbol{v}) = 1 - \mathbb{1}_{[c_1(\boldsymbol{v}), c_2(\boldsymbol{v})]}(u) + \sum_{i=1}^2 \gamma_i(\boldsymbol{v}) \mathbb{1}_{\{c_i(\boldsymbol{v})\}}(u)$, where $c_1, c_2, \gamma_1, \gamma_2 : (\mathbb{R}^{r-1}, \mathbb{B}^{r-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$, $0 \leq \gamma_1, \gamma_2 \leq 1$, are such that

$$\alpha \overset{!}{=} \int \Psi^*(u,\boldsymbol{v}) dP^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}_{\alpha_0^{(1)},\bullet}(u) \overset{!}{=} \int \Psi^*(u,\boldsymbol{v}) dP^{T_1^{(s)}|(T_2^{(s)},\dots,T_r^{(s)})'=\boldsymbol{v}}_{\alpha_0^{(2)},\bullet}(u).$$

By inspecting the proof of Thm. 3.3.4, the problem remains to finding the constants, respectively constant functions, $c_1 \equiv c_1(\boldsymbol{v})$ and $c_2 \equiv c_2(\boldsymbol{v})$, $\boldsymbol{v} \in \mathbb{R}^{r-1}$, with

$$1-\alpha \overset{!}{=} P(-c_2 < \Gamma(s, (\alpha_0^{(1)})^{-1}) < -c_1) \overset{!}{=} P(-c_2 < \Gamma(s, (\alpha_0^{(2)})^{-1}) < -c_1),$$

or, with the denotations introduced above,

$$1-\alpha \overset{!}{=} F_{\alpha_0^{(1)},s}(-c_1) - F_{\alpha_0^{(1)},s}(-c_2) \overset{!}{=} F_{\alpha_0^{(2)},s}(-c_1) - F_{\alpha_0^{(2)},s}(-c_2).$$

Again, w.l.o.g., we assume that $c_1 < c_2 < 0$. Then, setting $(c_1, c_2) = \tau^{(\alpha,\alpha_0^{(1)},s)}(\beta^*)$ yields the assertion. ■

Theorem 3.3.7
For $\alpha \in (0,1)$, $\alpha_0^{(1)}, \alpha_0^{(2)} \in \mathbb{R}_+$ with $\alpha_0^{(1)} < \alpha_0^{(2)}$, and the test problem $H_0 : \alpha_1 \leq \alpha_0^{(1)}$ or $\alpha_1 \geq \alpha_0^{(2)} \leftrightarrow H_1 : \alpha_0^{(1)} < \alpha_1 < \alpha_0^{(2)}$,

$$\varphi^* : \quad (\mathbb{R}^r_<)^{1\times s} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto \mathbb{1}_{\left(\tau_1^{(1-\alpha,\alpha_0^{(1)},s)}(\beta^*), \tau_2^{(1-\alpha,\alpha_0^{(1)},s)}(\beta^*)\right)}(T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on s independent observations of $\boldsymbol{X}$, where β^* is the unique solution of the equation

$$F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\alpha+\beta)) - F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta)) \stackrel{!}{=} \alpha \tag{3.3.7}$$

with respect to $\beta \in (0, 1-\alpha)$.

Proof. Applying Thm. 2.3.5, the statement can be shown similarly to the proof of Thm. 3.3.6, replacing α by $1-\alpha$. ■

Remark 3.3.8

Again, by means of simple analysis, it can be shown that the solution of (3.3.6) and (3.3.7), respectively, is uniquely determined. For fixed $\alpha \in (0,1)$, $\alpha_0^{(1)}, \alpha_0^{(2)} \in \mathbb{R}_+^r$ with $\alpha_0^{(1)} < \alpha_0^{(2)}$ and $s \in \mathbb{N}$, we define the mapping

$$g(\beta) = F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta)) - F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta)), \quad \beta \in (0,\alpha).$$

Then,

$$\begin{aligned}
\lim_{\beta \searrow 0} g(\beta) &= F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha)) > F_{\alpha_0^{(1)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha)) = 1-\alpha, \\
\lim_{\beta \nearrow \alpha} g(\beta) &= 1 - F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\alpha)) < 1 - F_{\alpha_0^{(1)},s}(F^{-1}_{\alpha_0^{(1)},s}(\alpha)) = 1-\alpha.
\end{aligned}$$

g is differentiable with

$$\begin{aligned}
g'(\beta) < 0 \quad &\Leftrightarrow \quad \frac{f_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta))}{f_{\alpha_0^{(1)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta))} - \frac{f_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta))}{f_{\alpha_0^{(1)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta))} < 0 \\
&\Leftrightarrow \quad \frac{f_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta))}{f_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta))} < \frac{f_{\alpha_0^{(1)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta))}{f_{\alpha_0^{(1)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta))} \\
&\Leftrightarrow \quad \exp\left\{-\alpha_0^{(2)}\left(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta) - F^{-1}_{\alpha_0^{(1)},s}(\beta)\right)\right\} \\
&\qquad < \exp\left\{-\alpha_0^{(1)}\left(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta) - F^{-1}_{\alpha_0^{(1)},s}(\beta)\right)\right\} \\
&\Leftrightarrow \quad \alpha_0^{(1)} < \alpha_0^{(2)}.
\end{aligned}$$

Again, it follows from the intermediate value theorem that there exists one and only one $\beta \in (0,\alpha)$ with $g(\beta) = 1-\alpha$.

For fixed $\alpha \in (0,1)$, $\alpha_0^{(1)}, \alpha_0^{(2)} \in \mathbb{R}_+$ and $s \in \mathbb{N}$, we introduce the mapping $\tilde{N}^{(\alpha,\alpha_0^{(1)},\alpha_0^{(2)},s)}$ defined by

$$\tilde{N}^{(\alpha,\alpha_0^{(1)},\alpha_0^{(2)},s)}(\beta) = F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta)) - F_{\alpha_0^{(2)},s}(F^{-1}_{\alpha_0^{(1)},s}(\beta)) - (1-\alpha)$$

for $\beta \in (0,\alpha)$ which is differentiable with respect to β and its derivative is given by

$$\frac{d}{d\beta}\tilde{N}^{(\alpha,\alpha_0^{(1)},\alpha_0^{(2)},s)}(\beta) = \left(\frac{\alpha_0^{(2)}}{\alpha_0^{(1)}}\right)^s \Big(\exp\{(\alpha_0^{(1)}-\alpha_0^{(2)})F^{-1}_{\alpha_0^{(1)},s}(1-\alpha+\beta)\} - \exp\{(\alpha_0^{(1)}-\alpha_0^{(2)})F^{-1}_{\alpha_0^{(1)},s}(\beta)\}\Big)$$

for $\beta \in (0,\alpha)$. Again, Newton's procedure can be applied to numerically obtain solutions of equation (3.3.6) and (3.3.7).

We end this section pointing out two important additional results and another application.

Remark 3.3.9
It should be noted that the UMPU level-α tests defined in Thm.s 3.3.1, 3.3.2 and 3.3.7 can even be considered as UMP level-α tests in the following sense. In general, the conditional distribution of T_1 given $(T_2^{(s)},\dots,T_r^{(s)}) = (t_2^{(s)},\dots,t_r^{(s)})$ forms a one-parameter exponential family in α_1 and the identity (see Lehmann & Romano (2005), La. 2.7.2 (ii), p. 48), and, hence, in this situation, in case of the test problems (I), (II) and (V), UMP tests on α_1 can be established from general results concerning one*-parameter exponential families (e.g., in Lehmann & Romano (2005), Shao (2003) or Witting (1985)). Here, these conditional UMP tests do not depend on the value of $(t_2^{(s)},\dots,t_r^{(s)})$ and coincide with the UMPU tests derived above.*

Remark 3.3.10
As it is seen from the corresponding proofs, Thm.s 3.3.1-3.3.7 remain valid if α_1 and $T_1^{(s)}$ are replaced by any other choice of α_j and its respective $T_j^{(s)}$, $2 \le j \le r$. Hence, a variety of useful tests are near at hand. For $2 \le j \le r$, the test problem $H_0 : \alpha_j = 1 \leftrightarrow H_1 : \alpha_j \neq 1$ is of particular interest and corresponds to the question, whether, upon failure of the $(j-1)^{th}$ component, the lifetime distribution of the remaining $n-j+1$ components is given by the belonging conditional distribution according to the baseline distribution function F. Tables 7.1-7.3 of Chapter 7 show solutions β^* of (3.3.2) and critical values $\tau_1^{(\alpha,1,s)}(\beta^*)$ and $\tau_2^{(\alpha,1,s)}(\beta^*)$ of the corresponding UMPU tests (cf. Thm. 3.3.4) with respect to the sample size s, where $\alpha \in \{0.01, 0.05, 0.1\}$. The solutions β^* are obtained with Newton's procedure (iteration until $|N^{(\alpha,s)}(\cdot)| < 10^{-15}$ with $N^{(\alpha,s)}$ as in (3.3.3)).

Finally, the derived tests are also helpful in the context of progressively type-II censored lifetime experiments. In such an experiment, N units are put on a lifetime test, where the failure times are described by iid random variables. Upon the first failure of some component, R_1 of the surviving units are randomly chosen and removed from the experiment, and it continues with $N - R_1 - 1$ remaining units on test. Upon the second failure of some unit, R_2 of the surviving units are randomly removed and so on. After the r^{th} failure, all remaining R_r units are removed from the experiment. Hence, r failure times are observed, $\sum_{i=1}^r R_i$ units are progressively censored and the number N of all units equals $r + \sum_{i=1}^r R_i$. For a detailed account of the model of progressive type-II censoring, we refer to Balakrishnan & Aggarwala (2000).
If the failure time of every unit is distributed according to the baseline distribution function F, the above model of progressive type-II censoring is contained in the model of SOSs with conditional

proportional hazard rates in the distribution theoretical sense. More precisely, the joint density of all r progressively type-II censored order statistics can be obtained by setting on the right-hand side of (3.0.2) $n = r$ and

$$\alpha_j = \frac{N - j + 1 - \sum_{i=1}^{j-1} R_i}{r - j + 1} = 1 + \frac{\sum_{i=j}^{r} R_i}{r - j + 1}$$

for $1 \leq j \leq r$ (cf., e.g., Balakrishnan & Aggarwala (2000), p. 8). Hence, the derived UMPU tests on the model parameters $\alpha_1, \ldots, \alpha_r$ can be rewritten and interpreted in the sense of progressive type-II censoring in the following way. We consider a progressively type-II censored lifetime experiment, where $N = r + \sum_{i=1}^{r} R_i$ units have been placed on a lifetime test and r failure times have been observed. However, suppose that the R's and, thus, the number N of all involved units are not available. Moreover, we assume to have s iid observations of that lifetime experiment. For fixed $1 \leq j \leq r$ and $m \in \mathbb{N}_0$, we consider as an example the two-sided test problem

$$H_0 : \sum_{i=j}^{r} R_i = m \quad \leftrightarrow \quad H_1 : \sum_{i=j}^{r} R_i \neq m, \tag{3.3.8}$$

concerning the question, whether, from the j^{th} failure of some unit to the end of the experiment, exactly m units have been progressively censored. Obviously, (3.3.8) can be rewritten as

$$H_0 : \alpha_j = 1 + \frac{m}{r - j + 1} \quad \leftrightarrow \quad H_1 : \alpha_j \neq 1 + \frac{m}{r - j + 1}. \tag{3.3.9}$$

In particular, if $j = 1$, the null hyothesis of (3.3.8) corresponds to the case that the number of all censored units equals m and, if additionally $m = 0$ is assumed, test problem (3.3.8) is related to the question, whether the lifetime test has been censored at all. In the latter case, the respective test problem (3.3.9) simplifies to $H_0 : \alpha_1 = 1$ against $H_1 : \alpha_1 \neq 1$, and Tables 7.1-7.3 of Chapter 7 show again critical values for some levels and sample sizes.

3.3.3 Model Tests with Simple Null Hypothesis

In what follows, we are concerned with the question whether for a given system the model of common OSs based on F is in fact appropriate to describe the lifetime of the system and its components, or if the model of SOSs might be the more adequate one. The model of common OSs based on F is included within the model of SOSs with conditional proportional hazard rates in the distribution theoretical sense by setting $\alpha_1 = \cdots = \alpha_r = 1$. Hence, we consider the test problem

$$H_0 : \alpha_1 = \cdots = \alpha_r = 1 \quad \leftrightarrow \quad H_1 : \exists j \in \{1, \ldots, r\} : \alpha_j \neq 1. \tag{3.3.10}$$

Concerning these hypotheses, two clarifying remarks should be made.
Firstly, it should be noted that H_0 corresponds to the assumption of common OSs *based on the underlying distribution function* F. If, based on some experiment, H_0 is rejected, the model of common OSs may not be dropped at all. Suppose that all α's equal the same positive number $\tilde{\alpha}$. Then, the considered model coincides with the model of common OSs based on the distribution function $\tilde{F} = 1 - (1 - F)^{\tilde{\alpha}}$.

Secondly, one might argue that the alternative should only consist of parameter vectors with positive components arranged in ascending order of magnitude, i.e. $\alpha_1 \leq \cdots \leq \alpha_r$, with at least one strict inequality, which seems to be the more natural alternative by modeling sequential systems and increasing load on remaining components. Nevertheless, we shall first of all consider the more general test problem (3.3.10) which can also be used in situations where apriori assumptions on the model parameters cannot be made.

In order to derive the test statistic of the LR test, we continue by computing the Kullback-Leibler distance of two vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)'$, $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, ..., \tilde{\alpha}_r)' \in \mathbb{R}_+^r$ of model parameters (cf. Subsection 2.3.2). We obtain, by application of La. 2.1.30 with κ and π as in (3.1.6) and (3.1.7),

$$\begin{aligned} d_{KL}(\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}) &= \kappa(\tilde{\boldsymbol{\alpha}}) - \kappa(\boldsymbol{\alpha}) + (\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}})'\pi(\boldsymbol{\alpha}) \\ &= -\sum_{j=1}^{r} \ln(\tilde{\alpha}_j) + \sum_{j=1}^{r} \ln(\alpha_j) - \sum_{j=1}^{r} (\alpha_j - \tilde{\alpha}_j)\alpha_j^{-1} \\ &= \sum_{j=1}^{r} \left[\frac{\tilde{\alpha}_j}{\alpha_j} - \ln\left(\frac{\tilde{\alpha}_j}{\alpha_j}\right) - 1 \right]. \end{aligned}$$

In particular, by setting $\tilde{\boldsymbol{\alpha}} = \underline{\mathbf{1}} = (1, \ldots, 1)'$,

$$d_{KL}(\boldsymbol{\alpha}, \underline{\mathbf{1}}) = \sum_{j=1}^{r} \left[\frac{1}{\alpha_j} - \ln\left(\frac{1}{\alpha_j}\right) - 1 \right].$$

Regarding La. 2.3.8, plugging in the MLE $\boldsymbol{\alpha}^{*(s)}$ of $\boldsymbol{\alpha}$ based on s independent observations yields

$$\begin{aligned} T_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= d_{KL}(\boldsymbol{\alpha}^{*(s)}, \underline{\mathbf{1}}) \\ &= \sum_{j=1}^{r} \left[\frac{1}{\alpha_j^{*(s)}} - \ln\left(\frac{1}{\alpha_j^{*(s)}}\right) - 1 \right] \\ &= \sum_{j=1}^{r} \left[-\frac{1}{s} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - \ln\left(-\frac{1}{s} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\right) - 1 \right] \\ &= \sum_{j=1}^{r} [Y_j - \ln(Y_j) - 1], \end{aligned}$$

where $Y_j = -s^{-1}T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$, $1 \leq j \leq r$, are jointly independent random variables with $Y_j \sim \Gamma(s, (s\alpha_j)^{-1})$, $1 \leq j \leq r$.
Hence, by noticing that the test statistic is continuously distributed, the level-α LR test corresponding to test problem (3.3.10) based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is given by

$$\varphi_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(d_s(\alpha),\infty)}\left(\sum_{j=1}^{r} \left[-\frac{1}{s} T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - \ln\left(-\frac{1}{s} T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right) - 1 \right] \right), \quad \tilde{\boldsymbol{x}}^{(s)} \in (\mathbb{R}_<^r)^{1\times s}, \tag{3.3.11}$$

and the critical value $d_s(\alpha)$ is derived from the equation

$$P\left(\sum_{j=1}^{r}[Z_j - \ln(Z_j) - 1] \leq d_s(\alpha)\right) = 1 - \alpha \tag{3.3.12}$$

where $Z_1, \ldots, Z_r$ are iid random variables having distribution $\Gamma(s, s^{-1})$.
By usage of (3.1.10) and (3.1.12), we calculate the test statistics of Wald's test and Rao's score test (cf. Subsection 2.3.3, representations (2.3.12) and (2.3.14)) and obtain

$$\begin{aligned}
T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= s(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}})'\mathbf{I}_f(\boldsymbol{\alpha}^{*(s)})(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}}) \\
&= s\sum_{j=1}^{r}\left(\frac{1}{\alpha_j^{*(s)}} - 1\right)^2 \\
&= s\sum_{j=1}^{r}\left(\frac{1}{s}T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + 1\right)^2 \\
&= s\sum_{j=1}^{r}(Y_j - 1)^2\,,
\end{aligned}$$

and

$$\begin{aligned}
T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= s^{-1}\boldsymbol{U}_{\underline{\mathbf{1}}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})'\mathbf{I}_f(\underline{\mathbf{1}})^{-1}\boldsymbol{U}_{\underline{\mathbf{1}}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \\
&= s^{-1}\sum_{j=1}^{r}\left(T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + s\right)^2 \\
&= s\sum_{j=1}^{r}\left(\frac{1}{s}T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + 1\right)^2 \\
&= s\sum_{j=1}^{r}(Y_j - 1)^2\,,
\end{aligned}$$

with the Y's as above. Here, Wald's statistic and Rao's score statistic coincide, and the respective level-α test for test problem (3.3.10) based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is given by

$$\varphi_{W,R}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(r_s(\alpha),\infty)}\left(s\sum_{j=1}^{r}\left(\frac{1}{s}T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) + 1\right)^2\right), \quad \tilde{\boldsymbol{x}}^{(s)} \in (\mathbb{R}^r_<)^{1\times s}, \tag{3.3.13}$$

where the critical value $r_s(\alpha)$ is obtained from the equation

$$P\left(s\sum_{j=1}^{r}(Z_j - 1)^2 \leq r_s(\alpha)\right) = 1 - \alpha, \tag{3.3.14}$$

with the Z's as defined above.
Wald's modified test statistic turns out to be

$$\begin{aligned} T_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= s(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}})'\mathbf{I}_f(\underline{\mathbf{1}})(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}}) \\ &= s\sum_{j=1}^{r}\left(\alpha_j^{*(s)} - 1\right)^2 \\ &= s\sum_{j=1}^{r}\left(\frac{s}{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})} + 1\right)^2 \\ &= s\sum_{j=1}^{r}\left(\frac{1}{Y_j} - 1\right)^2, \end{aligned}$$

with the Y's as above. Hence, Wald's modified level-α test for test problem (3.3.10) based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is given by

$$\varphi_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(\tilde{w}_s(\alpha),\infty)}\left(s\sum_{j=1}^{r}\left(\frac{s}{T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})} + 1\right)^2\right), \quad \tilde{\boldsymbol{x}}^{(s)} \in (\mathbb{R}_<^r)^{1\times s}, \tag{3.3.15}$$

where the critical value $\tilde{w}_s(\alpha)$ is determined by

$$P\left(s\sum_{j=1}^{r}\left(\frac{1}{Z_j} - 1\right)^2 \leq \tilde{w}_s(\alpha)\right) = 1 - \alpha, \tag{3.3.16}$$

with the Z's as above.

Although all of the three test statistics have a simple representation and can easily be calculated for some given vector of observations, an analytical determination of the critical values is in each case difficult (e.g., for the distribution of the LR test statistic, see Stehlik (2003)). However, by means of simulations, respective empirical quantiles can readily be obtained. Tables 7.4-7.12 of Chapter 7 show the critical values of the three tests above for different levels and sample sizes.

We go on by deriving asymptotic results related to the test statistics above. Firstly, from Thm.s 2.3.9 and 2.3.12, we obtain that $2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ and $T_{W,R}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ are asymptotically $\chi^2(r)$-distributed, i.e. chi-square distributed with r degrees of freedom, if H_0 is true. By noticing that $T_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = [\sqrt{s}(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}})]'\sqrt{s}(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}})$ and $\sqrt{s}(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}}) \xrightarrow{\mathcal{D}} \mathcal{N}_r(\mathbf{0}, \mathbf{I}_r)$ under H_0, this statement is also established for Wald's modified test statistic by using the continuous mapping theorem (e.g., in Billingsley (1999), Thm. 2.7, p. 21).
In the following, we focus on asymptotic efficiency properties, and, for this, we consider the exponential family $\mathfrak{P}^{\boldsymbol{T}}$ of continuous distributions on $(\mathbb{R}_-^r, \mathbb{R}_-^r \cap \mathbb{B}^r)$ with densities given by (3.1.4) and the corresponding independent observations $\boldsymbol{t}_{i;-} = (t_{i;1}, \ldots, t_{i;r})'$, $i = 1, \ldots, s$, obtained by the

transformation $t_{i;j} = T_j(\boldsymbol{x}^{(i)})$, $1 \leq i \leq s$, $1 \leq j \leq r$. Moreover, let $\tilde{\boldsymbol{t}}^{(s)} = (\boldsymbol{t}_{1;-}, \ldots, \boldsymbol{t}_{s;-})$, and $\boldsymbol{t}^{(s)} = (t_1^{(s)}, \ldots, t_r^{(s)})'$ with $t_j^{(s)} = \sum_{i=1}^{s} t_{i;j}$, $1 \leq j \leq r$.
In general, the sequence of LR tests fulfils the conditions $(i) - (iv)$ of Thm. 2.3.17. However, since a proof of this assertion is not completely given in Subsection 2.3.6, we briefly show that these conditions hold in the considered case. Here, from representation (3.3.11), the statistic W on $(\mathbb{R}_-^r, \mathbb{R}_-^r \cap \mathbb{B}^r)$ is given by

$$W(\boldsymbol{t}) = \sum_{j=1}^{r} \left[-t_j - \ln(-t_j) - 1\right], \quad \boldsymbol{t} = (t_1, \ldots, t_r)' \in \mathbb{R}_-^r.$$

For every $s \in \mathbb{N}$ and $d_s \in \mathbb{R}$, $\{\boldsymbol{t} \in \mathbb{R}_-^r : W(\boldsymbol{t}) \leq d_s\}$ is a convex set in $\mathbb{R}_-^r$, since $\mathbf{H}_W(\boldsymbol{t}) = \mathbf{diag}(t_1^{-2}, \ldots, t_r^{-2}) > 0$ and, hence, W is a convex function on $\mathbb{R}_-^r$. Using this fact in combination with $\nabla_j W(\boldsymbol{t}) = -1 - t_j^{-1}$, $1 \leq j \leq r$, we obtain that W is minimal at $\pi(\underline{\mathbf{1}}) = -\underline{\mathbf{1}}$. Hence, (i), (ii) and (iv) are clear. It is left to show that W strictly increases on rays from $-\underline{\mathbf{1}}$, as we will demonstrate in the following.
Let $\boldsymbol{t} \in \mathbb{R}_-^r$, $\boldsymbol{t}_\rho = -\underline{\mathbf{1}} + \rho(\boldsymbol{t} + \underline{\mathbf{1}}) \in \mathbb{R}_-^r$ for $\rho \in (0,1)$, and define the mapping $g_{\boldsymbol{t}}(\rho) = W(\boldsymbol{t}_\rho)$, $\rho \in (0,1)$. Then,

$$g_{\boldsymbol{t}}(\rho) = \sum_{j=1}^{r} [-\rho(t_j + 1) - \ln(1 - \rho(t_j + 1))], \quad \rho \in (0,1),$$

and

$$g'_{\boldsymbol{t}}(\rho) = \sum_{j=1}^{r} \left[-(t_j + 1) + \frac{t_j + 1}{1 - \rho(t_j + 1)}\right] = \sum_{j=1}^{r} \frac{\rho(t_j + 1)^2}{-(\boldsymbol{t}_\rho)_j} > 0, \quad \rho \in (0,1).$$

Application of Thm. 2.3.17 yields that the sequence $\{\varphi_{LR}^{(s)}\}_{s \in \mathbb{N}}$ is asymptotically optimal in the sense of Bahadur with exact slope

$$c_{\varphi_{LR}}(\boldsymbol{\alpha}) = 2 d_{KL}(\boldsymbol{\alpha}, \underline{\mathbf{1}}) = 2 \sum_{j=1}^{r} \left[\alpha_j^{-1} - \ln(\alpha_j^{-1}) - 1\right], \quad \boldsymbol{\alpha} \in \Theta_1. \tag{3.3.17}$$

Moreover, it follows from Thm. 2.3.19, (i), since $\Theta = \mathbb{R}_+^r$, $U_\epsilon(\Theta_0)$ a compact (proper) subset of $\mathbb{R}_+^r$ for every $\epsilon > 0$ and, thus, $\iota(\Theta_0) = \infty$, that, at every $\boldsymbol{\alpha} \in \Theta_1$, the sequence $\{\varphi_{LR}^{(s)}\}_{s \in \mathbb{N}}$ of LR tests is deficient in the sense of Bahadur of order $O(\ln(N^+(\alpha, \beta, \boldsymbol{\alpha})))$ as $\alpha \searrow 0$. Finally, from Thm. 2.3.21, the sequence is also strongly i-efficient.

Wald's test/Rao's score test satisfies the conditions of Thm. 2.3.17, too, as we will show. From (3.3.13), using the denotations introduced above, the respective statistic $\tilde{W}$ on $(\mathbb{R}_-^r, \mathbb{R}_-^r \cap \mathbb{B}^r)$ is

$$\tilde{W}(\boldsymbol{t}) = \sum_{j=1}^{r} (t_j + 1)^2, \quad \boldsymbol{t} = (t_1, \ldots, t_r)' \in \mathbb{R}_-^r.$$

Here, $\mathbf{H}_{\tilde{W}}(\boldsymbol{t}) = 2\mathbf{I}_r > 0$ and $\nabla \tilde{W}(\boldsymbol{t}) = 2(\boldsymbol{t} + \underline{\mathbf{1}})$, and, hence, (i), (ii) and (iv) are shown. For $\boldsymbol{t} \in \mathbb{R}_-^r$ and $\rho \in (0,1)$, let $\boldsymbol{t}_\rho$ be defined as above. Then, $\tilde{W}(\boldsymbol{t}_\rho) = \rho^2(\boldsymbol{t} + \underline{\mathbf{1}})'(\boldsymbol{t} + \underline{\mathbf{1}})$ strictly increases in $\rho \in (0,1)$. Hence, application of Thm. 2.3.17 yields that the exact Bahadur slope of Wald's test/Rao's score test at $\boldsymbol{\alpha} \in \Theta_1$ is given by

$$c_{\varphi_{W,R}}(\boldsymbol{\alpha}) = \inf\left\{ c_{\varphi_{LR}}(\tilde{\boldsymbol{\alpha}}) : \sum_{j=1}^{r}(\tilde{\alpha}_j^{-1}-1)^2 = \sum_{j=1}^{r}(\alpha_j^{-1}-1)^2,\ \tilde{\boldsymbol{\alpha}} \in \Theta_1 \right\}.$$

It is not difficult to see that, for $r > 1$, $c_{\varphi_{LR}}(\boldsymbol{\alpha}) > c_{\varphi_{W,R}}(\boldsymbol{\alpha})$ for every $\boldsymbol{\alpha} \in \Theta_1$. Let $\boldsymbol{\alpha} \in \Theta_1$ be arbitrary, and $M^2 = \sum_{j=1}^{r}(\alpha_j^{-1}-1)^2 > 0$. For $\phi \in (0,\pi/2)$, let $\tilde{\boldsymbol{\alpha}}^{\phi} = ((M\cos(\phi)+1)^{-1}, (M\sin(\phi)+1)^{-1}, 1, \ldots, 1)' \in \Theta_1$. Then, $\sum_{j=1}^{r}((\tilde{\alpha}_j^{\phi})^{-1}-1)^2 = M^2$ for all $\phi \in (0,\pi/2)$. Moreover, we define the mapping $v_{\tilde{\boldsymbol{\alpha}}}(\phi) = c_{\varphi_{LR}}(\tilde{\boldsymbol{\alpha}}^{\phi})/2$, $\phi \in (0,\pi/2)$. Then,

$$\begin{aligned} v_{\tilde{\boldsymbol{\alpha}}}(\phi) &= \sum_{j=1}^{r}((\tilde{\alpha}_j^{\phi})^{-1} - \ln((\tilde{\alpha}_j^{\phi})^{-1}) - 1) \\ &= M\cos(\phi) - \ln(M\cos(\phi)+1) + M\sin(\phi) - \ln(M\sin(\phi)+1), \end{aligned}$$

and the derivative of $v_{\tilde{\boldsymbol{\alpha}}}$ is given by

$$\begin{aligned} v'_{\tilde{\boldsymbol{\alpha}}}(\phi) &= -M\sin(\phi) + \frac{M\sin(\phi)}{M\cos(\phi)+1} + M\cos(\phi) - \frac{M\cos(\phi)}{M\sin(\phi)+1} \\ &= M^2\sin(\phi)\cos(\phi)\left(\frac{1}{M\sin(\phi)+1} - \frac{1}{M\cos(\phi)+1}\right). \end{aligned}$$

Hence, since $v'_{\tilde{\boldsymbol{\alpha}}}(\phi) = 0$ if and only if $\phi = \pi/4$, we conclude that $v_{\tilde{\boldsymbol{\alpha}}}$ is not constant on $(0,\pi/2)$ and, thus, $c_{\varphi_{W,R}}(\boldsymbol{\alpha}) < c_{\varphi_{LR}}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha} \in \Theta_1$.

We summarize our findings in the following theorem.

Theorem 3.3.11
For test problem (3.3.10) with a simple null hypothesis, the following assertions hold true:

(i) *Wald's test and Rao's score test based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ coincide and are given in virtue of (3.3.13) and (3.3.14).*

(ii) *Wald's modified test based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is represented by (3.3.15) and (3.3.16).*

(iii) *The LR test based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is obtained from (3.3.11) and (3.3.12).*

(iv) *If H_0 is true, $2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$, $T_{W,R}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ and $T_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ are each asymptotically $\chi^2(r)$-distributed. Moreover, the sequence $\{\tilde{\varphi}_{LR}^{(s)}\}_{s\in\mathbb{N}}$ of tests defined by $\tilde{\varphi}_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(\chi^2_{1-\alpha}(r),\infty)}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$, $s \in \mathbb{N}$, has asymptotic level α and is consistent.*

(v) *The sequence of LR tests is asymptotically optimal and deficient in the sense of Bahadur of order $O(\ln(N^+(\alpha,\beta,\boldsymbol{\alpha})))$ as $\alpha \searrow 0$ at every $\boldsymbol{\alpha} \in \Theta_1$. Its exact slope is given by* (3.3.17) *and satisfies $c_{\varphi_{LR}}(\boldsymbol{\alpha}) > c_{\varphi_{W,R}}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha} \in \Theta_1$, in case of $r > 1$. Moreover, the sequence of LR tests is strongly i-efficient.*

3.3.4 Model Tests with Composite Null Hypothesis

In Subsection 3.3.3, we have discussed the question whether a given system is of the common type, and the observed data can be considered as realizations of common OSs based on the underlying distribution function F. Now, suppose a system which is described by the model of SOSs with conditional proportional hazard rates based on F, where all of the model parameters equal $\tilde{\alpha} \neq 1$. Then, the system can be considered as a common system and be modelled by common OSs based on the distribution function $\tilde{F} = 1 - (1 - F)^{\tilde{\alpha}}$. Hence, if, based on some experiment, one of the tests of Subsection 3.3.3 rejects the null hypothesis, this does not give evidence for a sequential system with some change in load, but it ensures (with respect to the level) that common OSs *based on* F are inappropriate to model the lifetime of the system and its components.
In this section, we replace the simple null hypothesis (3.3.10) by $\Theta_0 = \Theta_= = \{\boldsymbol{\alpha} \in \mathbb{R}_+^r : \alpha_1 = \cdots = \alpha_r\}$ and discuss the test problem

$$H_0 : \boldsymbol{\alpha} \in \Theta_= \quad \leftrightarrow \quad H_1 : \boldsymbol{\alpha} \in \Theta_1 = \mathbb{R}_+^r \setminus \Theta_=. \tag{3.3.18}$$

We illustrate the usage of test problem (3.3.18) by means of an example. Suppose we have data from some possibly sequential $(n - r + 1)$-out-of-n system, where, based on some prior experiment, F_j is assumed to be the distribution function of the exponential distribution $Exp(\alpha_j^{-1})$ with unknown scale parameter $\alpha_j^{-1} > 0$, $1 \leq j \leq r$. Defining the baseline distribution as the standard exponential distribution, we obtain $F_j = 1 - (1 - F)^{\alpha_j}$, $1 \leq j \leq r$. Then, if the null hypothesis is rejected based on some experiment and test statistic, this gives evidence against the assumption that the observed data can be described by common OSs based on *any* exponential distribution $Exp(\lambda)$, $\lambda > 0$. In that case, a rejection of the null hypothesis can be interpreted as a change of load on remaining components, and, thus, the model of SOSs is supposed to better fit the data than the model of common OSs.

As in Subsection 3.3.3, we determine respective test statistics for test problem (3.3.18). At first, from La. 2.3.8, the LR test statistic based on $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)})$ is given by $T_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = d_{KL}(\boldsymbol{\alpha}^{*(s)}, \Theta_0) = \inf_{\boldsymbol{\alpha}^{(0)} \in \Theta_0} d_{KL}(\boldsymbol{\alpha}^{*(s)}, \boldsymbol{\alpha}^{(0)})$. For $\delta > 0$, let $\underline{\boldsymbol{\delta}} = (\delta, \ldots, \delta)' \in \mathbb{R}_+^r$ and define

$$\begin{aligned} g(\delta) &= d_{KL}(\boldsymbol{\alpha}^{*(s)}, \underline{\boldsymbol{\delta}}) \\ &= \sum_{j=1}^{r} \left[\frac{\delta}{\alpha_j^{*(s)}} - \ln\left(\frac{\delta}{\alpha_j^{*(s)}} \right) - 1 \right]. \end{aligned}$$

Then, the derivative of g is given by

$$\begin{aligned} g'(\delta) &= \sum_{j=1}^{r} \left[\frac{1}{\alpha_j^{*(s)}} - \frac{1}{\delta} \right] \\ &= r \left[\left(\frac{1}{r} \sum_{j=1}^{r} \frac{1}{\alpha_j^{*(s)}} \right) - \frac{1}{\delta} \right]. \end{aligned}$$

Hence, setting $\alpha_{=}^{*(s)} = \left(\frac{1}{r}\sum_{j=1}^{r}\frac{1}{\alpha_j^{*(s)}}\right)^{-1}$, g is strictly decreasing (increasing) on $(0, \alpha_{=}^{*(s)})$ ($(\alpha_{=}^{*(s)}, \infty)$) and, thus, has a global minimal value at $\alpha_{=}^{*(s)}$. We obtain

$$\begin{aligned} d_{KL}(\boldsymbol{\alpha}^{*(s)}, \Theta_0) &= d_{KL}(\boldsymbol{\alpha}^{*(s)}, \underline{\boldsymbol{\alpha}_{=}^{*(s)}}) \\ &= \sum_{j=1}^{r}\left[\frac{\alpha_{=}^{*(s)}}{\alpha_j^{*(s)}} - \ln\left(\frac{\alpha_{=}^{*(s)}}{\alpha_j^{*(s)}}\right) - 1\right] \\ &= -\sum_{j=1}^{r}\ln\left(\frac{\alpha_{=}^{*(s)}}{\alpha_j^{*(s)}}\right) \\ &= -\sum_{j=1}^{r}\ln\left(\frac{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}{r^{-1}\sum_{k=1}^{r}T_k^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}\right) \\ &= -\sum_{j=1}^{r}\ln\left(\frac{\tilde{Y}_j}{r^{-1}\sum_{k=1}^{r}\tilde{Y}_k}\right), \end{aligned}$$

where $\tilde{Y}_1, \ldots, \tilde{Y}_r$ are jointly independent random variables with $\tilde{Y}_j = -T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \sim \Gamma(s, \alpha_j^{-1})$. If H_0 is true, $\tilde{Y}_1, \ldots, \tilde{Y}_r \overset{iid}{\sim} \Gamma(s, \alpha_{=}^{-1})$ for some (unknown) $\alpha_{=} > 0$. Then, $B_j = \frac{\tilde{Y}_j}{\sum_{k=1}^{r}\tilde{Y}_k}$ has a beta distribution $Beta(s, (r-1)s)$ with shape parameters s and $(r-1)s$, and, in particular, its distribution does not depend on the value of $\alpha_{=}$. Hence, the level-α LR test for test problem (3.3.18) based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is represented by

$$\varphi_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(d_s(\alpha),\infty)}\left(-\sum_{j=1}^{r}\ln\left(\frac{T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}{r^{-1}\sum_{k=1}^{r}T_k^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}\right)\right), \quad \tilde{\boldsymbol{x}}^{(s)} \in (\mathbb{R}_<^r)^{1\times s}, \tag{3.3.19}$$

where $d_s(\alpha)$ is obtained from the equation

$$P\left(-\sum_{j=1}^{r}\ln\left(\frac{\tilde{Z}_j}{r^{-1}\sum_{k=1}^{r}\tilde{Z}_k}\right) \le d_s(\alpha)\right) = 1 - \alpha, \tag{3.3.20}$$

where $\tilde{Z}_j \overset{iid}{\sim} \Gamma(s, 1)$.

We continue by deriving Rao's score statistic for test problem (3.3.18) according to (2.3.17), and, for this, we define the mapping $h : \mathbb{R}_+^r \to \mathbb{R}^{r-1} : \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)' \mapsto (\alpha_2 - \alpha_1, \ldots, \alpha_r - \alpha_{r-1})'$. Then, (3.3.18) is equivalent to test problem

$$H_0 : h(\boldsymbol{\alpha}) = \mathbf{0} \quad \leftrightarrow \quad H_1 : h(\boldsymbol{\alpha}) \neq \mathbf{0}. \tag{3.3.21}$$

From $\frac{d}{d\delta}\ln(f_{\underline{\boldsymbol{\delta}}}^{(s)}) = \sum_{j=1}^{r}T_j^{(s)} + sr\delta^{-1} = rs(\delta^{-1} - r^{-1}\sum_{j=1}^{r}(-s^{-1}T_j))$, $\delta > 0$, it is easily seen that $\underline{\boldsymbol{\alpha}_{=}^{*(s)}}$ is the MLE of $\boldsymbol{\alpha}$ in $\Theta_{=}$, i.e. under the restriction that $h(\boldsymbol{\alpha}) = 0$ and, thus, all α's are equal.

Hence, by notice of (3.1.10) and (3.1.12), Rao's score statistic based on $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(s)})$ is as follows:

$$\begin{aligned} T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= s^{-1}(\boldsymbol{U}^{(s)}_{\underline{\boldsymbol{\alpha}^{*(s)}_{=}}}(\tilde{\boldsymbol{X}}^{(s)}))' \mathbf{I}_f(\underline{\boldsymbol{\alpha}^{*(s)}_{=}})^{-1} \boldsymbol{U}^{(s)}_{\underline{\boldsymbol{\alpha}^{*(s)}_{=}}}(\tilde{\boldsymbol{X}}^{(s)}) \\ &= s^{-1}(\alpha^{*(s)}_{=})^2 \sum_{j=1}^{r} \left(T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + \frac{s}{\alpha^{*(s)}_{=}} \right)^2 \\ &= s(\alpha^{*(s)}_{=})^2 \sum_{j=1}^{r} \left(\frac{1}{s} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + \frac{1}{\alpha^{*(s)}_{=}} \right)^2 \\ &= s \sum_{j=1}^{r} \left(\frac{\alpha^{*(s)}_{=}}{\alpha^{*(s)}_{j}} - 1 \right)^2 \\ &= s \sum_{j=1}^{r} \left(\frac{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}{r^{-1}\sum_{k=1}^{r} T_k^{(s)}(\tilde{\boldsymbol{X}}^{(s)})} - 1 \right)^2 \\ &= s \sum_{j=1}^{r} \left(\frac{\tilde{Y}_j}{r^{-1}\sum_{k=1}^{r} \tilde{Y}_k} - 1 \right)^2, \end{aligned}$$

with $\tilde{Y}_1, \ldots, \tilde{Y}_r$ as above. Thus, based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$, Rao's score test with level α is given by

$$\varphi_R^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(r_s(\alpha),\infty)} \left(s \sum_{j=1}^{r} \left(\frac{T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}{r^{-1}\sum_{k=1}^{r} T_k^{(s)}(\tilde{\boldsymbol{x}}^{(s)})} - 1 \right)^2 \right), \quad \tilde{\boldsymbol{x}}^{(s)} \in (\mathbb{R}^r_<)^{1\times s}, \tag{3.3.22}$$

where the critical value $r_s(\alpha)$ is such that

$$P\left(s \sum_{j=1}^{r} \left(\frac{\tilde{Z}_j}{r^{-1}\sum_{k=1}^{r} \tilde{Z}_k} - 1 \right)^2 \le r_s(\alpha) \right) = 1 - \alpha \tag{3.3.23}$$

with the $\tilde{Z}$'s as above.

Finally, we derive Wald's test statistic for test problem (3.3.18), respectively (3.3.21), according to (2.3.15) and (2.3.16). The Jacobian matrix of h at $\boldsymbol{\alpha} \in \mathbb{R}^r_+$ is given by $\mathbf{D}_h(\boldsymbol{\alpha}) = \mathbf{Q} = [Q_{i,j}]_{1\le i\le r-1, 1\le, j\le r}$ with entries $Q_{i,i} = -1$ and $Q_{i,i+1} = 1$ for $i = 1, \ldots, r-1$, and zero otherwise. Hence,

$$
\begin{aligned}
\tilde{\mathbf{I}}(\boldsymbol{\alpha}) &= \mathbf{Q}\,\mathbf{diag}(\alpha_1^2,\dots,\alpha_r^2)\,\mathbf{Q}' \\
&= \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \mathbf{diag}(\alpha_1^2,\dots,\alpha_r^2) \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}' \\
&= \begin{pmatrix} -\alpha_1^2 & \alpha_2^2 & 0 & \cdots & 0 \\ 0 & -\alpha_2^2 & \alpha_3^2 & \ddots & \dots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\alpha_{r-1}^2 & \alpha_r^2 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}' \\
&= \begin{pmatrix} \alpha_1^2+\alpha_2^2 & -\alpha_2^2 & 0 & \cdots & \cdots & 0 \\ -\alpha_2^2 & \alpha_2^2+\alpha_3^2 & -\alpha_3^2 & \ddots & & \vdots \\ 0 & -\alpha_3^2 & \alpha_3^2+\alpha_4^2 & -\alpha_4^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & -\alpha_{r-1}^2 \\ 0 & \cdots & \cdots & 0 & -\alpha_{r-1}^2 & \alpha_{r-1}^2+\alpha_r^2 \end{pmatrix}.
\end{aligned}
$$

Corollary 3.3.12
For every $\boldsymbol{\alpha} \in \mathbb{R}_+^r$, $\tilde{\mathbf{I}}(\boldsymbol{\alpha})$ is invertible with inverse matrix $\tilde{\mathbf{I}}^{-1}(\boldsymbol{\alpha}) = \mathbf{B} = [B_{i,j}]_{1\le i,j\le r-1}$, where

$$
B_{i,j} = \frac{\tilde{B}_{i,j}}{\sum_{k=1}^r \prod_{\substack{q=1\\ q\neq k}}^r \alpha_q^2}, \quad 1 \le i,j \le r-1,
$$

with

$$
\tilde{B}_{i,j} = \left(\sum_{t=1}^{i} \prod_{\substack{p=1\\ p\neq t}}^{i} \alpha_p^2 \right) \left(\prod_{m=i+1}^{j} \alpha_m^2 \right) \left(\sum_{s=j+1}^{r} \prod_{\substack{l=j+1\\ l\neq s}}^{r} \alpha_l^2 \right)
$$

for $i \le j$ and $\tilde{B}_{i,j} = \tilde{B}_{j,i}$ for $i > j$. Moreover, with h as defined above,

$$
h(\boldsymbol{\alpha})'\tilde{\mathbf{I}}(\boldsymbol{\alpha})^{-1}h(\boldsymbol{\alpha}) = r - \frac{\left(\sum_{j=1}^r \alpha_j^{-1}\right)^2}{\sum_{j=1}^r \alpha_j^{-2}}. \tag{3.3.24}
$$

Proof. Let $\boldsymbol{\alpha} \in \mathbb{R}_+^r$ be fixed. We show that $\tilde{\mathbf{I}}(\boldsymbol{\alpha})\ \mathbf{B} = \mathbf{I}_{r-1}$. Then, since $\tilde{\mathbf{I}}(\boldsymbol{\alpha})$ and $\mathbf{B}$ are symmetric, it follows that $\mathbf{B}\ \tilde{\mathbf{I}}(\boldsymbol{\alpha}) = (\tilde{\mathbf{I}}(\boldsymbol{\alpha})\ \mathbf{B})' = \mathbf{I}_{r-1}$.

The case $i = j = 1$:

$$\begin{aligned}
(\alpha_1^2+\alpha_2^2)\tilde{B}_{1,1} - \alpha_2^2\tilde{B}_{2,1} &= (\alpha_1^2+\alpha_2^2)\tilde{B}_{1,1} - \alpha_2^2\tilde{B}_{1,2} \\
&= (\alpha_1^2+\alpha_2^2)\left(\sum_{s=2}^{r}\prod_{\substack{l=2\\l\neq s}}^{r}\alpha_l^2\right) - \alpha_2^2\alpha_2^2\left(\sum_{s=3}^{r}\prod_{\substack{l=3\\l\neq s}}^{r}\alpha_l^2\right) \\
&= \alpha_1^2\left(\sum_{s=2}^{r}\prod_{\substack{l=2\\l\neq s}}^{r}\alpha_l^2\right) + \alpha_2^2\left(\prod_{\substack{l=2\\l\neq 2}}^{r}\alpha_l^2\right) = \left(\sum_{s=2}^{r}\prod_{\substack{l=1\\l\neq s}}^{r}\alpha_l^2\right) + \left(\prod_{l=2}^{r}\alpha_l^2\right) \\
&= \sum_{s=1}^{r}\prod_{\substack{l=1\\l\neq s}}^{r}\alpha_l^2.
\end{aligned}$$

The case $i = 1, j \in \{2, \ldots, r-1\}$:

$$\begin{aligned}
(\alpha_1^2+\alpha_2^2)\tilde{B}_{1,j} - \alpha_2^2\tilde{B}_{2,j} &= \left[(\alpha_1^2+\alpha_2^2)\left(\prod_{m=2}^{j}\alpha_m^2\right) - \alpha_2^2\left(\alpha_1^2+\alpha_2^2\right)\left(\prod_{m=3}^{j}\alpha_m^2\right)\right]\left(\sum_{s=j+1}^{r}\prod_{\substack{l=j+1\\l\neq s}}^{r}\alpha_l^2\right) \\
&= 0.
\end{aligned}$$

The case $i = j = r-1$:

$$\begin{aligned}
-\alpha_{r-1}^2\tilde{B}_{r-2,r-1} + (\alpha_{r-1}^2+\alpha_r^2)\tilde{B}_{r-1,r-1} &= -\alpha_{r-1}^2\left(\sum_{t=1}^{r-2}\prod_{\substack{p=1\\p\neq t}}^{r-2}\alpha_p^2\right)\alpha_{r-1}^2 + (\alpha_{r-1}^2+\alpha_r^2)\left(\sum_{t=1}^{r-1}\prod_{\substack{p=1\\p\neq t}}^{r-1}\alpha_p^2\right) \\
&= \alpha_r^2\left(\sum_{t=1}^{r-1}\prod_{\substack{p=1\\p\neq t}}^{r-1}\alpha_p^2\right) + \alpha_{r-1}^2\left(\prod_{\substack{p=1\\p\neq r-1}}^{r-1}\alpha_p^2\right) \\
&= \left(\sum_{t=1}^{r-1}\prod_{\substack{p=1\\p\neq t}}^{r}\alpha_p^2\right) + \left(\prod_{p=1}^{r-1}\alpha_p^2\right) \\
&= \sum_{t=1}^{r}\prod_{\substack{p=1\\p\neq t}}^{r}\alpha_p^2.
\end{aligned}$$

The case $i = r-1, j \in \{1, \dots, r-2\}$:

$$\begin{aligned}
-\alpha_{r-1}^2 \tilde{B}_{r-2,j} &+ (\alpha_{r-1}^2 + \alpha_r^2)\tilde{B}_{r-1,j} \\
&= -\alpha_{r-1}^2 \tilde{B}_{j,r-2} + (\alpha_{r-1}^2 + \alpha_r^2)\tilde{B}_{j,r-1} \\
&= \left(\sum_{t=1}^{j} \prod_{\substack{p=1 \\ p\neq t}}^{j} \alpha_p^2\right) \left[-\alpha_{r-1}^2 \left(\prod_{m=j+1}^{r-2} \alpha_m^2\right)(\alpha_{r-1}^2 + \alpha_r^2) + (\alpha_{r-1}^2 + \alpha_r^2)\left(\prod_{m=j+1}^{r-1} \alpha_m^2\right)\right] \\
&= 0.
\end{aligned}$$

The case $2 \leq i = j \leq r-2$:

$$\begin{aligned}
-\alpha_j^2 \tilde{B}_{j-1,j} &+ (\alpha_j^2 + \alpha_{j+1}^2)\tilde{B}_{j,j} - \alpha_{j+1}^2 \tilde{B}_{j+1,j} \\
&= -\alpha_j^2 \tilde{B}_{j-1,j} + (\alpha_j^2 + \alpha_{j+1}^2)\tilde{B}_{j,j} - \alpha_{j+1}^2 \tilde{B}_{j,j+1} \\
&= -\alpha_j^2 \left(\sum_{t=1}^{j-1} \prod_{\substack{p=1 \\ p\neq t}}^{j-1} \alpha_p^2\right) \alpha_j^2 \left(\sum_{s=j+1}^{r} \prod_{\substack{l=j+1 \\ l\neq s}}^{r} \alpha_l^2\right) + (\alpha_j^2 + \alpha_{j+1}^2)\left(\sum_{t=1}^{j} \prod_{\substack{p=1 \\ p\neq t}}^{j} \alpha_p^2\right)\left(\sum_{s=j+1}^{r} \prod_{\substack{l=j+1 \\ l\neq s}}^{r} \alpha_l^2\right) \\
&\quad -\alpha_{j+1}^2 \left(\sum_{t=1}^{j} \prod_{\substack{p=1 \\ p\neq t}}^{j} \alpha_p^2\right) \alpha_{j+1}^2 \left(\sum_{s=j+2}^{r} \prod_{\substack{l=j+2 \\ l\neq s}}^{r} \alpha_l^2\right) \\
&= \alpha_j^2 \left(\prod_{\substack{p=1 \\ p\neq j}}^{j} \alpha_p^2\right)\left(\sum_{s=j+1}^{r} \prod_{\substack{l=j+1 \\ l\neq s}}^{r} \alpha_l^2\right) + \alpha_{j+1}^2 \left(\sum_{t=1}^{j} \prod_{\substack{p=1 \\ p\neq t}}^{j} \alpha_p^2\right)\left(\prod_{\substack{l=j+1 \\ l\neq j+1}}^{r} \alpha_l^2\right) \\
&= \left(\sum_{s=j+1}^{r} \prod_{\substack{l=1 \\ l\neq s}}^{r} \alpha_l^2\right) + \left(\sum_{t=1}^{j} \prod_{\substack{p=1 \\ p\neq t}}^{r} \alpha_p^2\right) \\
&= \left(\sum_{s=1}^{r} \prod_{\substack{l=1 \\ l\neq s}}^{r} \alpha_l^2\right).
\end{aligned}$$

The case $i \in \{2, \ldots, r-2\}, j \in \{1, \ldots, i-1\}$:

$$\begin{aligned}
-\alpha_i^2 \tilde{B}_{i-1,j} &+ (\alpha_i^2 + \alpha_{i+1}^2)\tilde{B}_{i,j} - \alpha_{i+1}^2 \tilde{B}_{i+1,j} \\
&= -\alpha_i^2 \tilde{B}_{j,i-1} + (\alpha_i^2 + \alpha_{i+1}^2)\tilde{B}_{j,i} - \alpha_{i+1}^2 \tilde{B}_{j,i+1} \\
&= \left(\prod_{m=j+1}^{i-1} \alpha_m^2\right)\left(\sum_{t=1}^{j}\prod_{\substack{p=1\\p\neq t}}^{j} \alpha_p^2\right)\Bigg[-\alpha_i^2\left(\sum_{s=i}^{r}\prod_{\substack{l=i\\l\neq s}}^{r}\alpha_l^2\right) \\
&\quad +(\alpha_i^2+\alpha_{i+1}^2)\alpha_i^2\left(\sum_{s=i+1}^{r}\prod_{\substack{l=i+1\\l\neq s}}^{r}\alpha_l^2\right) - \alpha_{i+1}^2\alpha_i^2\alpha_{i+1}^2\left(\sum_{s=i+2}^{r}\prod_{\substack{l=i+2\\l\neq s}}^{r}\alpha_l^2\right)\Bigg] \\
&= \left(\prod_{m=j+1}^{i-1} \alpha_m^2\right)\left(\sum_{t=1}^{j}\prod_{\substack{p=1\\p\neq t}}^{j} \alpha_p^2\right)\left[-\alpha_i^2\left(\prod_{\substack{l=i\\l\neq i}}^{r}\alpha_l^2\right) + \alpha_{i+1}^2\left(\prod_{\substack{l=i\\l\neq i+1}}^{r}\alpha_l^2\right)\right] \\
&= 0.
\end{aligned}$$

The case $i \in \{2, \ldots, r-2\}, j \in \{i+1, \ldots, r-1\}$:

$$\begin{aligned}
-\alpha_i^2 \tilde{B}_{i-1,j} &+ (\alpha_i^2 + \alpha_{i+1}^2)\tilde{B}_{i,j} - \alpha_{i+1}^2 \tilde{B}_{i+1,j} \\
&= \left(\prod_{m=i+2}^{j} \alpha_m^2\right)\left(\sum_{s=j+1}^{r}\prod_{\substack{l=j+1\\l\neq s}}^{r} \alpha_l^2\right)\Bigg[-\alpha_i^2\left(\sum_{t=1}^{i-1}\prod_{\substack{p=1\\p\neq t}}^{i-1}\alpha_p^2\right)\alpha_i^2\alpha_{i+1}^2 \\
&\quad +(\alpha_i^2+\alpha_{i+1}^2)\left(\sum_{t=1}^{i}\prod_{\substack{p=1\\p\neq t}}^{i}\alpha_p^2\right)\alpha_{i+1}^2 - \alpha_{i+1}^2\left(\sum_{t=1}^{i+1}\prod_{\substack{p=1\\p\neq t}}^{i+1}\alpha_p^2\right)\Bigg] \\
&= \left(\prod_{m=i+2}^{j} \alpha_m^2\right)\left(\sum_{s=j+1}^{r}\prod_{\substack{l=j+1\\l\neq s}}^{r} \alpha_l^2\right)\left[\alpha_i^2\alpha_{i+1}^2\left(\prod_{\substack{p=1\\p\neq i}}^{i}\alpha_p^2\right) - \alpha_{i+1}^2\left(\prod_{\substack{p=1\\p\neq i+1}}^{i+1}\alpha_p^2\right)\right] \\
&= 0.
\end{aligned}$$

Hence, the first part of the corollary is shown. Now, let us show representation (3.3.24). Firstly, notice that

$$\frac{\left(\sum_{j=1}^r \alpha_j^{-1}\right)^2}{\sum_{j=1}^r \alpha_j^{-2}} = \frac{\left(\sum_{j=1}^r \alpha_j^{-1}\right)^2}{\sum_{j=1}^r \alpha_j^{-2}} \cdot \frac{\left(\prod_{k=1}^r \alpha_k\right)^2}{\prod_{k=1}^r \alpha_k^2} = \frac{\left(\sum_{j=1}^r \prod_{\substack{k=1\\k\neq j}}^r \alpha_k\right)^2}{\sum_{j=1}^r \prod_{\substack{k=1\\k\neq j}}^r \alpha_k^2}.$$

Hence, equation (3.3.24) is equivalent to

$$\sum_{i=1}^{r-1}\sum_{j=1}^{r-1}(\alpha_{i+1}-\alpha_i)(\alpha_{j+1}-\alpha_j)\tilde{B}_{i,j} = r\sum_{j=1}^{r}\prod_{\substack{k=1\\k\neq j}}^{r}\alpha_k^2 - \left(\sum_{j=1}^{r}\prod_{\substack{k=1\\k\neq j}}^{r}\alpha_k\right)^2. \tag{3.3.25}$$

Regarding the definition of $\tilde{B}_{i,j}$, $1 \leq i,j \leq r-1$, we divide both sides of (3.3.25) by $\prod_{k=1}^{r}\alpha_k^2$ and obtain the equivalent equation

$$\begin{aligned}
&\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)^2\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+1}^{r}\frac{1}{\alpha_s^2}\right)\\
+\ &2\sum_{i=1}^{r-2}\sum_{j=i+1}^{r-1}(\alpha_{i+1}-\alpha_i)(\alpha_{j+1}-\alpha_j)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=j+1}^{r}\frac{1}{\alpha_s^2}\right)\\
=\ &r\left(\sum_{i=1}^{r}\frac{1}{\alpha_i^2}\right)-\left(\sum_{i=1}^{r}\frac{1}{\alpha_i}\right)^2.
\end{aligned} \tag{3.3.26}$$

Now,

$$\begin{aligned}
2\sum_{i=1}^{r-2}\sum_{j=i+1}^{r-1}(\alpha_{i+1} &- \alpha_i)(\alpha_{j+1}-\alpha_j)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=j+1}^{r}\frac{1}{\alpha_s^2}\right)\\
&= 2\sum_{i=1}^{r-2}(\alpha_{i+1}-\alpha_i)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{j=i+1}^{r-1}\sum_{s=j+1}^{r}(\alpha_{j+1}-\alpha_j)\frac{1}{\alpha_s^2}\right)\\
&= 2\sum_{i=1}^{r-2}(\alpha_{i+1}-\alpha_i)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+2}^{r}\frac{1}{\alpha_s^2}\sum_{j=i+1}^{s-1}(\alpha_{j+1}-\alpha_j)\right)\\
&= 2\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+2}^{r}\frac{\alpha_s-\alpha_{i+1}}{\alpha_s^2}\right)\\
&= 2\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+1}^{r}\frac{1}{\alpha_s}-\alpha_{i+1}\sum_{s=i+1}^{r}\frac{1}{\alpha_s^2}\right)\\
&= 2\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+1}^{r}\frac{1}{\alpha_s}\right)\\
&\quad+\sum_{i=1}^{r-1}(2\alpha_i\alpha_{i+1}-2\alpha_{i+1}^2)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+1}^{r}\frac{1}{\alpha_s^2}\right).
\end{aligned}$$

Thus, by inserting this representation in (3.3.26), what is left to show is

$$\sum_{i=1}^{r-1}(\alpha_i^2-\alpha_{i+1}^2)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+1}^{r}\frac{1}{\alpha_s^2}\right) \;+\; 2\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)\left(\sum_{t=1}^{i}\frac{1}{\alpha_t^2}\right)\left(\sum_{s=i+1}^{r}\frac{1}{\alpha_s}\right)$$
$$= r\left(\sum_{i=1}^{r}\frac{1}{\alpha_i^2}\right) \;-\; \left(\sum_{i=1}^{r}\frac{1}{\alpha_i}\right)^2,$$

or, equivalently, by introducing the notation $c_i=\sum_{j=1}^{i}\frac{1}{\alpha_j^2}$, $d_i=\sum_{j=i}^{r}\frac{1}{\alpha_j^2}$ and $\tilde{d}_i=\sum_{j=i}^{r}\frac{1}{\alpha_j}$, $1\le i\le r$,

$$\sum_{i=1}^{r-1}(\alpha_i^2-\alpha_{i+1}^2)c_i d_{i+1}+2\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)c_i\tilde{d}_{i+1}=rc_r-\tilde{d}_1^2. \tag{3.3.27}$$

Obviously,

$$\begin{aligned}
\sum_{i=1}^{r-1}(\alpha_i^2-\alpha_{i+1}^2)c_i d_{i+1} &= \sum_{i=1}^{r-1}\alpha_i^2 c_i d_{i+1}-\sum_{i=2}^{r}\alpha_i^2 c_{i-1}d_i \\
&= \sum_{i=2}^{r-1}\alpha_i^2(c_i d_{i+1}-c_{i-1}d_i)+\alpha_1^2 c_1 d_2-\alpha_r^2 c_{r-1}d_r \\
&= \sum_{i=2}^{r-1}\alpha_i^2((c_{i-1}+\alpha_i^{-2})d_{i+1}-c_{i-1}(d_{i+1}+\alpha_i^{-2}))+d_2-c_{r-1} \\
&= \sum_{i=2}^{r-1}(d_{i+1}-c_{i-1})+d_2-c_{r-1} \\
&= \sum_{i=1}^{r-1}d_{i+1}-\sum_{i=2}^{r}c_{i-1} \\
&= \sum_{i=1}^{r-1}(d_{i+1}-c_i).
\end{aligned} \tag{3.3.28}$$

Similarly,

$$
\begin{aligned}
\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)c_i\tilde{d}_{i+1} &= \sum_{i=2}^{r}\alpha_i c_{i-1}\tilde{d}_i - \sum_{i=1}^{r-1}\alpha_i c_i \tilde{d}_{i+1} \\
&= \sum_{i=2}^{r-1}\alpha_i(c_{i-1}\tilde{d}_i - c_i\tilde{d}_{i+1}) + \alpha_r c_{r-1}\tilde{d}_r - \alpha_1 c_1 \tilde{d}_2 \\
&= \sum_{i=2}^{r-1}\alpha_i(c_{i-1}(\tilde{d}_{i+1}+\alpha_i^{-1}) - (c_{i-1}+\alpha_i^{-2})\tilde{d}_{i+1}) + c_{r-1} - \alpha_1^{-1}\tilde{d}_2 \\
&= \sum_{i=2}^{r-1}(c_{i-1}-\alpha_i^{-1}\tilde{d}_{i+1}) + c_{r-1} - \alpha_1^{-1}\tilde{d}_2 \\
&= \sum_{i=2}^{r} c_{i-1} - \sum_{i=1}^{r-1}\alpha_i^{-1}\tilde{d}_{i+1} \\
&= \sum_{i=1}^{r-1}(c_i - \alpha_i^{-1}\tilde{d}_{i+1}). \qquad (3.3.29)
\end{aligned}
$$

Hence, from (3.3.28) and (3.3.29), we conclude that

$$
\begin{aligned}
\sum_{i=1}^{r-1}(\alpha_i^2 &- \alpha_{i+1}^2)c_i d_{i+1} + 2\sum_{i=1}^{r-1}(\alpha_{i+1}-\alpha_i)c_i\tilde{d}_{i+1} \\
&= \sum_{i=1}^{r-1}(d_{i+1} - c_i + 2(c_i - \alpha_i^{-1}\tilde{d}_{i+1})) \\
&= \sum_{i=1}^{r-1}(d_{i+1} + c_i - 2\alpha_i^{-1}\tilde{d}_{i+1}) \\
&= \sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\frac{1}{\alpha_j^2} + \sum_{i=1}^{r-1}\sum_{j=1}^{i}\frac{1}{\alpha_j^2} - 2\sum_{i=1}^{r-1}\frac{1}{\alpha_i}\sum_{j=i+1}^{r}\frac{1}{\alpha_j} \\
&= \sum_{j=2}^{r}\frac{1}{\alpha_j^2}\sum_{i=1}^{j-1}1 + \sum_{j=1}^{r-1}\frac{1}{\alpha_j^2}\sum_{i=j}^{r-1}1 - 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\frac{1}{\alpha_i}\frac{1}{\alpha_j} \\
&= \sum_{j=2}^{r}\frac{1}{\alpha_j^2}(j-1) + \sum_{j=1}^{r-1}\frac{1}{\alpha_j^2}(r-j) - \sum_{i=1}^{r-1}\sum_{\substack{j=1 \\ j\neq i}}^{r}\frac{1}{\alpha_i}\frac{1}{\alpha_j}
\end{aligned}
$$

$$\begin{aligned}
&= \sum_{j=1}^{r} \frac{1}{\alpha_j^2}(r-1) - \sum_{i=1}^{r-1} \sum_{\substack{j=1 \\ j \neq i}}^{r} \frac{1}{\alpha_i} \frac{1}{\alpha_j} \\
&= r \sum_{j=1}^{r} \frac{1}{\alpha_j^2} - \sum_{i=1}^{r} \sum_{j=1}^{r} \frac{1}{\alpha_i} \frac{1}{\alpha_j} \\
&= r \sum_{j=1}^{r} \frac{1}{\alpha_j^2} - \left(\sum_{i=1}^{r} \frac{1}{\alpha_i} \right)^2 \\
&= r c_r - \tilde{d}_1^2,
\end{aligned}$$

and equation (3.3.27) and, thus, equation (3.3.24) are established. ■

Application of Cor. 3.3.12 yields that Wald's test statistic equals

$$\begin{aligned}
T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= s h(\boldsymbol{\alpha}^{*(s)})' \tilde{\mathbf{I}}(\boldsymbol{\alpha}^{*(s)})^{-1} h(\boldsymbol{\alpha}^{*(s)}) = s \left(r - \frac{\left(\sum_{j=1}^{r} (\alpha_j^{*(s)})^{-1} \right)^2}{\sum_{j=1}^{r} (\alpha_j^{*(s)})^{-2}} \right) \\
&= s \left(r - \frac{\left(\sum_{j=1}^{r} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \right)^2}{\sum_{j=1}^{r} (T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}))^2} \right) = s \left(r - \frac{\left(\sum_{j=1}^{r} \tilde{Y}_j \right)^2}{\sum_{j=1}^{r} \tilde{Y}_j^2} \right)
\end{aligned}$$

with the random variables $\tilde{Y}_j$, $1 \le j \le r$, as in the beginning of this subsection.
As it is the case for the LR test statistic and Rao's test statistic, if H_0 is true and, thus, $\alpha_1 = \cdots = \alpha_r = \alpha_=$ for some (unknown) $\alpha_= > 0$, the distribution of $T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ does not depend on the value of $\alpha_=$. Hence, based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$, Wald's test with level α is given by

$$\varphi_W^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(w_s(\alpha),\infty)} \left(s \left(r - \frac{\left(\sum_{j=1}^{r} T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \right)^2}{\sum_{j=1}^{r} (T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))^2} \right) \right), \quad \tilde{\boldsymbol{x}}^{(s)} \in (\mathbb{R}_<^r)^{1 \times s}, \tag{3.3.30}$$

where the critical value $w_s(\alpha)$ is derived from the equation

$$P \left(s \left(r - \frac{\left(\sum_{j=1}^{r} \tilde{Z}_j \right)^2}{\sum_{j=1}^{r} \tilde{Z}_j^2} \right) \le w_s(\alpha) \right) = 1 - \alpha \tag{3.3.31}$$

with the $\tilde{Z}$'s as they are introduced in the beginning of this subsection.

We summarize our findings in the following theorem, where the asymptotical results are obtained from Thm.'s 2.3.11 and 2.3.13.

Theorem 3.3.13
For test problem (3.3.18) with a composite null hypothesis, the following assertions hold true:

(i) Wald's test based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is given in virtue of (3.3.30) and (3.3.31).

(ii) Rao's score test based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is represented by (3.3.22) and (3.3.23).

(iii) The LR test based on s independent observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(s)}$ is obtained from (3.3.19) and (3.3.20).

(iv) If H_0 is true, $2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$, $T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ and $T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ are each asymptotically $\chi^2(r-1)$-distributed.

For small sample sizes and different levels, critical values of the test statistics are contained in Tables 7.13-7.21.

3.3.5 Multivariate Tests with Nuisance Parameters

In this subsection, we briefly discuss the case, where only some of the α's are of interest and the remaining ones are considered as fixed (possibly unknown) nuisance parameters. Let $I = \{j_1, \ldots, j_q\} \subseteq \{1, \ldots, r\}$, $|I| = q$, be an index set with interpretation as in Rem. 3.1.4. Then, from Subsection 3.1.3, it is obvious that the results of Subsections 3.3.3 and 3.3.4 remain valid with some minor changes.
We consider the test problems

$$H_0 : \alpha_{j_1} = \cdots = \alpha_{j_q} = 1 \quad \leftrightarrow \quad H_1 : \exists j_0 \in I : \alpha_{j_0} \neq 1, \tag{3.3.32}$$

and

$$H_0 : \alpha_{j_1} = \cdots = \alpha_{j_q} \quad \leftrightarrow \quad H_1 : \exists k_1, k_2 \in I : \alpha_{k_1} \neq \alpha_{k_2}. \tag{3.3.33}$$

Lemma 3.3.14
With the above denotations, we obtain the following statements:

(i) For test problem (3.3.32), the assertions of Thm. 3.3.11 are true, where all appearing sums are taken over the indices in I and the number r of degrees of freedom is replaced by q.

(ii) For test problem (3.3.33), the assertions of Thm. 3.3.13 are true, where all appearing sums are taken over the indices in I and the number $r-1$ of degrees of freedom is replaced by $q-1$.

As a consequence of La. 3.3.14, for small sample sizes and different levels, critical values of the tests can, once again, be obtained from Tables 7.4-7.21 of Chapter 7 replacing r by $q = |I|$.

3.3.6 Testing under Simple Order Restriction

Using the theorem of Subsection 2.3.4, for the model of SOSs with conditional proportional hazard rates, we derive the asymptotic distribution of the LR test statistic for two test problems concerning the simple ordering of the model parameters $\alpha_1, \ldots, \alpha_r$. For this, let

$$\Theta_= = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)' \in \mathbb{R}_+^r : \alpha_1 = \cdots = \alpha_r\}$$

and

$$\Theta_\leq = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)' \in \mathbb{R}_+^r : \alpha_1 \leq \cdots \leq \alpha_r\}.$$

Now, we are interested in testing

$$H_0 : \boldsymbol{\alpha} \in \Theta_= \quad \leftrightarrow \quad H_1 : \boldsymbol{\alpha} \in \Theta_\leq \setminus \Theta_= \tag{3.3.34}$$

and

$$H_0 : \boldsymbol{\alpha} \in \Theta_\leq \quad \leftrightarrow \quad H_1 : \boldsymbol{\alpha} \in \mathbb{R}_+^r \setminus \Theta_\leq. \tag{3.3.35}$$

When the model of SOSs with conditional proportional hazard rates is used to describe sequential systems, test problem (3.3.34) can be used in order to decide whether the system is of common or sequential type, where the prior information of simply ordered model parameters is taken into account.

Theorem 3.3.15
For the above test problems, the following assertions hold true:

(i) In case of test problem (3.3.34),

$$\lim_{s\to\infty} P_{\boldsymbol{\alpha}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) = \sum_{j=1}^{r} \frac{|S_r^j|}{r!} P(\chi^2(j-1) \geq c) \tag{3.3.36}$$

for every $\boldsymbol{\alpha} \in \Theta_=$ and $c \in \mathbb{R}$.

(ii) In case of test problem (3.3.35),

$$\lim_{s\to\infty} P_{\boldsymbol{\alpha}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) \leq \lim_{s\to\infty} P_{\boldsymbol{\eta}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) = \sum_{j=1}^{r} \frac{|S_r^j|}{r!} P(\chi^2(r-j) \geq c)$$

for every $\boldsymbol{\alpha} \in \Theta_\leq$, $\boldsymbol{\eta} \in \Theta_=$ and $c \in \mathbb{R}$.

At this, S_r^j are the Stirling numbers of the first kind and $\chi^2(0) = 0$.

Proof. (i). Using the same denotations as in the proof of Thm. 3.2.8, we consider the independent random variables $T_j(\boldsymbol{X}^{(i)})$, $1 \leq i \leq s$, $1 \leq j \leq r$. Again, we reparametrize the model parameters via $\tilde{\alpha}_j = \alpha_{r-j+1}^{-1}$, $1 \leq j \leq r$, and, moreover, define the probability measures $\tilde{P}_{\tilde{\boldsymbol{\alpha}}} = P_{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_r)' \in \mathbb{R}_+^r$. Then, for fixed $1 \leq j \leq r$, $T_j(\boldsymbol{X}^{(1)}), \ldots, T_j(\boldsymbol{X}^{(s)})$ are s iid random variables having λ^1-density $\tilde{f}_{\tilde{\alpha}_{r-j+1}}$. Obviously, (3.3.34) is equivalent to testing $H_0 : \tilde{\boldsymbol{\alpha}} \in \Theta_=$ against $H_1 : \tilde{\boldsymbol{\alpha}} \in \Theta_\leq \setminus \Theta_=$. Moreover, setting $\tilde{\boldsymbol{T}}^{(rs)}(\tilde{\boldsymbol{X}}^{(s)}) = (T_1(\boldsymbol{X}^{(1)}), \ldots, T_1(\boldsymbol{X}^{(s)}), \ldots, T_r(\boldsymbol{X}^{(1)}), \ldots, T_r(\boldsymbol{X}^{(s)}))$

and denoting the LR test statistic based on $\tilde{\boldsymbol{T}}^{(rs)}(\tilde{\boldsymbol{X}}^{(s)})$ by $\tilde{T}_{LR}^{(rs)}(\tilde{\boldsymbol{T}}^{(rs)}(\tilde{\boldsymbol{X}}^{(s)}))$, it is easily seen that, for $\boldsymbol{\eta} \in \Theta_=$ and $c \in \mathbb{R}$,

$$\begin{aligned} P_{\boldsymbol{\eta}}(2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \geq c) &= P_{\boldsymbol{\eta}}\left(-2\ln\left(\frac{\sup_{\boldsymbol{\alpha}\in\Theta_=} f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}{\sup_{\boldsymbol{\alpha}\in\Theta_\leq} f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}\right) \geq c\right) \\ &= \tilde{P}_{\tilde{\boldsymbol{\eta}}}\left(-2\ln\left(\frac{\sup_{\tilde{\boldsymbol{\alpha}}\in\Theta_=} \tilde{f}_{\tilde{\boldsymbol{\alpha}}}^{(rs)}(\tilde{\boldsymbol{T}}^{(rs)}(\tilde{\boldsymbol{X}}^{(s)}))}{\sup_{\tilde{\boldsymbol{\alpha}}\in\Theta_\leq} \tilde{f}_{\tilde{\boldsymbol{\alpha}}}^{(rs)}(\tilde{\boldsymbol{T}}^{(rs)}(\tilde{\boldsymbol{X}}^{(s)}))}\right) \geq c\right) \\ &= \tilde{P}_{\tilde{\boldsymbol{\eta}}}(2rs\tilde{T}_{LR}^{(rs)}(\tilde{\boldsymbol{T}}^{(rs)}(\tilde{\boldsymbol{X}}^{(s)})) \geq c). \end{aligned} \tag{3.3.37}$$

Since $\tilde{\boldsymbol{\eta}} \in \Theta_=$, it follows from Thm. 2.3.14 with $n_1^{(rs)} = \cdots = n_r^{(rs)} = s$, $s \in \mathbb{N}$, and $a_1 = \cdots = a_r = r^{-1} \in (0,1)$, that the probability in (3.3.37) converges to the right-hand side of (3.3.36) as s tends to infinity.

(ii). Replacing in the proof of (i) $\Theta_=$ by $\Theta_\leq$ and $\Theta_\leq$ by $\mathbb{R}_+^r$, the assertion follows similarly. ■

Chapter 4

Generalizations and Extensions

In Chapter 3, we have shown that the joint density of the first r SOSs with conditional proportional hazard rates forms a multivariate exponential family in the model parameters. Using the fact that this structure was preserved by considering product probability measures and densities, inferential issues have been discussed, and optimal estimators and tests have been established based on a sample of s iid vectors of SOSs with conditional proportional hazard rates, where the underlying distribution was assumed to be known.
With some minor changes only, most of the results remain true if we consider the less restrictive case of inid vectors of SOSs. The respective statements are then applicable, e.g., when independent observations are made from differently structured sequential systems, assuming that the systems have some of the model parameters in common. This case is discussed in Section 4.1.
In Section 4.2, SOSs with conditional proportional hazard rates are considered based on a parametric distribution function, which is not completely known.

4.1 INID Vectors of SOSs

4.1.1 Basic Properties

In what follows, we extend fundamental findings of Section 3.1 to the more general case of inid vectors of SOSs.
We assume to have s independent vectors of SOSs $\boldsymbol{X}^{(1;1)}, \ldots, \boldsymbol{X}^{(m_1;1)}, \ldots, \boldsymbol{X}^{(1;p)}, \ldots, \boldsymbol{X}^{(m_p;p)}$ with conditional proportional hazard rates based on a known absolutely continuous baseline distribution function F with corresponding density function f, where, for fixed $i \in \{1, \ldots, p\}$, $\boldsymbol{X}^{(1;i)}, \ldots, \boldsymbol{X}^{(m_i;i)}$ are identically distributed having $\lambda^{r_i}|_{\mathbb{R}^{r_i}_{<}}$-density

$$f^{\boldsymbol{X}^{(1;i)}}_{\boldsymbol{\alpha}^{(i)}}(\boldsymbol{x}) = \left(\prod_{j=1}^{r_i} \alpha_j\right) \exp\left\{\sum_{j=1}^{r_i} \alpha_j T_{j;i}(\boldsymbol{x})\right\} \left(\frac{n_i!}{(n_i - r_i)!} \prod_{j=1}^{r_i} \frac{f(x_j)}{1 - F(x_j)}\right),$$

$\boldsymbol{x} = (x_1, \ldots, x_{r_i})' \in \mathbb{R}^{r_i}_{<}$, with vector $\boldsymbol{\alpha}^{(i)} = (\alpha_1, \ldots, \alpha_{r_i})'$ of model parameters and statistics defined as

$$\begin{aligned} T_{1;i}(\boldsymbol{x}) &= n_i \ln(1 - F(x_1)), \\ T_{j;i}(\boldsymbol{x}) &= (n_i - j + 1) \ln\left(\frac{1 - F(x_j)}{1 - F(x_{j-1})}\right), \quad 2 \le j \le r_i, \end{aligned}$$

for $\boldsymbol{x} = (x_1, \ldots, x_{r_i})' \in \mathbb{R}^{r_i}_{<}$. At this, $r_1, \ldots, r_p, n_1, \ldots, n_p \in \mathbb{N}$ are integers satisfying $r_1 \ge \cdots \ge r_p$ and $r_i \le n_i$, $1 \le i \le p$.
Interpreting the above sample situation in terms of sequential systems, we consider independent observations from p possibly differently structured systems, where, for $1 \le i \le p$, the i^{th} system is assumed to be a sequential $(n_i - r_i + 1)$-out-of-n_i-system with conditional proportional hazard rates $\alpha_1 f/(1 - F), \ldots, \alpha_{r_i} f/(1 - F)$. Collectively, $s = \sum_{i=1}^{p} m_i$ independent observations are assumed from the systems, where m_i observations are made of the failure time of the i^{th} system and its components, $1 \le i \le p$. By assumption, the systems have one or more of the model parameters $\alpha_1, \ldots, \alpha_{r_1}$ in common.

We define the integers $w_j = |\{i \in \{1, \ldots, p\} : r_i \ge j\}| = \max\{i \in \{1, \ldots, p\} : r_i \ge j\}$ and $c_j = \sum_{i=1}^{w_j} m_i$ for $j \in \{1, \ldots, r_1\}$. In the context of sequential systems, w_j is the number of systems, where α_j is involved, and c_j is the total number of the corresponding observations, $1 \le j \le r_1$. For every $i \in \{1, \ldots, p\}$, the family $\{P^{\boldsymbol{X}^{(1;i)}}_{\boldsymbol{\alpha}^{(i)}} = f^{\boldsymbol{X}^{(1;i)}}_{\boldsymbol{\alpha}^{(i)}} \lambda^{r_i}|_{\mathbb{R}^{r_i}_{<}} : \boldsymbol{\alpha}^{(i)} \in \mathbb{R}^{r_i}_{+}\}$ forms a r_i-parametrical exponential family in $\alpha_1, \ldots, \alpha_{r_i}$ and statistics $T_{1;i}, \ldots, T_{r_i;i}$ and, thus, by setting $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1;1)}, \ldots, \boldsymbol{X}^{(m_1;1)}, \ldots, \boldsymbol{X}^{(1;p)}, \ldots, \boldsymbol{X}^{(m_p;p)})$, $r = r_1$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(1)}$, we obtain from La. 2.1.25 that the family of joint distributions $\mathfrak{P}^{\tilde{\boldsymbol{X}}^{(s)}}_{mix} = \{P^{\tilde{\boldsymbol{X}}^{(s)}}_{\boldsymbol{\alpha}} = f^{(s)}_{\boldsymbol{\alpha}} \otimes_{i=1}^{p} \otimes_{l=1}^{m_i} \lambda^{r_i}|_{\mathbb{R}^{r_i}_{<}}\}$ of $\boldsymbol{X}^{(1;1)}, \ldots, \boldsymbol{X}^{(m_1;1)}, \ldots, \boldsymbol{X}^{(1;p)}, \ldots, \boldsymbol{X}^{(m_p;p)}$ forms a r-parametrical exponential family in $\alpha_1, \ldots, \alpha_r$ and statistics $T^{(s)}_1, \ldots, T^{(s)}_r$ with

$$\begin{aligned} f^{(s)}_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{x}}^{(s)}) &= \prod_{i=1}^{p} \prod_{l=1}^{m_i} f^{\boldsymbol{X}^{(1;i)}}_{\boldsymbol{\alpha}^{(i)}}(\boldsymbol{x}^{(l;i)}) \\ &= C(\boldsymbol{\alpha}) \exp\left\{\sum_{j=1}^{r} \alpha_j T^{(s)}_j(\tilde{\boldsymbol{x}}^{(s)})\right\} h(\tilde{\boldsymbol{x}}^{(s)}), \end{aligned} \tag{4.1.1}$$

$\tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1;1)}, \ldots, \boldsymbol{x}^{(m_1;1)}, \ldots, \boldsymbol{x}^{(1;p)}, \ldots, \boldsymbol{x}^{(m_p;p)}) \in \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}^{r_i}_{<}$, $\boldsymbol{x}^{(l;i)} = (x^{(l;i)}_1, ..., x^{(l;i)}_{r_i}) \in \mathbb{R}^{r_i}_{<}$, $1 \le i \le p$, $1 \le l \le m_i$, where

$$C(\boldsymbol{\alpha}) = \prod_{j=1}^{r} \alpha_j^{c_j}, \quad \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)' \in \mathbb{R}^{r}_{+},$$

and

$$
\begin{aligned}
h(\tilde{\boldsymbol{x}}^{(s)}) &= \prod_{i=1}^{p}\prod_{l=1}^{m_i}\frac{n_i!}{(n_i-r_i)!}\prod_{j=1}^{r_i}\frac{f(x_j^{(l;i)})}{1-F(x_j^{(l;i)})},\\
T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= \sum_{i=1}^{w_1}\sum_{l=1}^{m_i} n_i \ln(1-F(x_1^{(l;i)})),\\
T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= \sum_{i=1}^{w_j}\sum_{l=1}^{m_i}(n_i-j+1)\ln\left(\frac{1-F(x_j^{(l;i)})}{1-F(x_{j-1}^{(l;i)})}\right), \quad 2\le j\le r,
\end{aligned}
$$

$\tilde{\boldsymbol{x}}^{(s)} = (\boldsymbol{x}^{(1;1)},\dots,\boldsymbol{x}^{(m_1;1)},\dots,\boldsymbol{x}^{(1;p)},\dots,\boldsymbol{x}^{(m_p;p)}) \in \times_{i=1}^{p}\times_{l=1}^{m_i}\mathbb{R}_<^{r_i}$, $\boldsymbol{x}^{(l;i)} = (x_1^{(l;i)},...,x_{r_i}^{(l;i)}) \in \mathbb{R}_<^{r_i}$, $1\le i\le p$, $1\le l\le m_i$.
The natural parameter space of $\mathfrak{P}_{mix}^{\tilde{\boldsymbol{X}}^{(s)}}$ is given by $\mathbb{R}_+^r$, and the densities are given in the canonical form.
As in the iid case, we obtain the distribution of $-T_j^{(s)}$, $1\le j\le r$, by deriving the moment generating function of $-\boldsymbol{T}^{(s)} = (-T_1^{(s)},\dots,-T_r^{(s)})'$ according to La. 2.1.18, i.e., for fixed $\boldsymbol{\alpha}\in\mathbb{R}_+^r$,

$$
m_{-\boldsymbol{T}^{(s)}}(\boldsymbol{t}) = m_{\boldsymbol{T}^{(s)}}(-\boldsymbol{t}) = \frac{C(\boldsymbol{\alpha})}{C(\boldsymbol{\alpha}-\boldsymbol{t})} = \prod_{j=1}^{r}\left(\frac{\alpha_j}{\alpha_j-t_j}\right)^{c_j},
$$

for $\boldsymbol{t} = (t_1,\dots,t_r)' \in (-\delta,\delta)^r$, where $\delta = \min\{\alpha_1,\dots,\alpha_r\}/2$. Hence, $-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \sim \Gamma(c_j,\alpha_j^{-1})$ has a gamma distribution with shape parameter c_j and scale parameter α_j^{-1}, $1\le j\le r$, and $T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)}),\dots,T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ are jointly independent, where $\tilde{\boldsymbol{X}}^{(s)} = (\boldsymbol{X}^{(1;1)},\dots,\boldsymbol{X}^{(m_1;1)},\dots,\boldsymbol{X}^{(1;p)},\dots,\boldsymbol{X}^{(m_p;p)})$. We use this finding to prove the following lemma related to the class $\mathfrak{P}_{mix}^{\tilde{\boldsymbol{X}}^{(s)}}$ of distributions.

Lemma 4.1.1
In the above situation, the following assertions hold true:

(i) $\mathfrak{P}_{mix}^{\tilde{\boldsymbol{X}}^{(s)}}$ *is strictly* r*-parametrical.*

(ii) $\boldsymbol{T}^{(s)} = (T_1^{(s)},\dots,T_r^{(s)})'$ *is minimal sufficient and complete for* $\mathfrak{P}_{mix}^{\tilde{\boldsymbol{X}}^{(s)}}$.

(iii) The Fisher information matrix of $\mathfrak{P}_{mix}^{\tilde{\boldsymbol{X}}^{(s)}}$ *is given by*

$$
\mathbf{I}_f^{(s)}(\boldsymbol{\alpha}) = \mathbf{diag}\left(\frac{c_1}{\alpha_1^2},...,\frac{c_r}{\alpha_r^2}\right), \quad \boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_r)'\in\mathbb{R}_+^r. \tag{4.1.2}
$$

Proof. (i). The result is obvious from Thm. 2.1.9, since

$$
\mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}^{(s)}) = \mathbf{diag}\left(\frac{c_1}{\alpha_1^2},\dots,\frac{c_r}{\alpha_r^2}\right) > 0, \quad \boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_r)'\in\mathbb{R}_+^r.
$$

(ii). The statement follows from (i) and application of La. 2.1.20.
(iii). From Thm. 2.1.22, we conclude that $\mathbf{I}_f^{(s)}(\boldsymbol{\alpha}) = \mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}^{(s)})$, $\boldsymbol{\alpha}\in\mathbb{R}_+^r$, and, thus, all assertions are shown. ■

4.1.2 Estimation

We consider the sample situation of Subsection 4.1.1 and assume that the model parameters $\alpha_1, \ldots, \alpha_r$ are unknown. Then, the MLEs and UMVUEs of the model parameters can similarly be obtained as in the iid case.
Firstly, we define the mapping

$$\kappa : \quad \mathbb{R}_+^r \to \mathbb{R} : \quad (\alpha_1, \ldots, \alpha_r)' \mapsto -\sum_{j=1}^{r} c_j \ln(\alpha_j), \tag{4.1.3}$$

and rewrite the density (4.1.1) in the form

$$f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \exp\left\{\sum_{j=1}^{r} \alpha_j T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - \kappa(\boldsymbol{\alpha})\right\} h(\tilde{\boldsymbol{x}}^{(s)}), \quad \tilde{\boldsymbol{x}}^{(s)} \in \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}_<^{r_i}. \tag{4.1.4}$$

By application of Thm. 2.1.15 (i), the Hessian matrix of κ is given by $\mathbf{H}_\kappa(\boldsymbol{\alpha}) = \mathbf{Cov}_{\boldsymbol{\alpha}}(\boldsymbol{T}^{(s)}) > 0$ and, moreover, the mapping $\pi : \mathbb{R}_+^r \to \pi(\mathbb{R}_+^r) : \boldsymbol{\alpha} \mapsto E_{\boldsymbol{\alpha}}[\boldsymbol{T}^{(s)}]$ is represented by

$$\pi(\boldsymbol{\alpha}) = \nabla\kappa(\boldsymbol{\alpha}) = -\left(\frac{c_1}{\alpha_1}, \ldots, \frac{c_r}{\alpha_r}\right)', \quad \boldsymbol{\alpha} \in \mathbb{R}_+^r, \tag{4.1.5}$$

where $\pi(\mathbb{R}_+^r) = \mathbb{R}_-^r$. π is continuously differentiable and bijective with inverse function

$$\pi^{-1} : \quad \mathbb{R}_-^r \to \mathbb{R}_+^r : \quad \boldsymbol{t} \mapsto -\left(\frac{c_1}{t_1}, \ldots, \frac{c_r}{t_r}\right)'. \tag{4.1.6}$$

Theorem 4.1.2
In the above sample situation, the following statements hold true:

(i) The unique MLE of $\boldsymbol{\alpha}$ based on $\tilde{\boldsymbol{X}}^{(s)}$ is given by

$$\boldsymbol{\alpha}^{*(s)} = \left(-\frac{c_1}{T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \ldots, -\frac{c_r}{T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}\right)'.$$

Moreover, if $g : \mathbb{R}_+^r \to \Gamma$ is a bijective function, $g(\boldsymbol{\alpha}^{*(s)})$ is the MLE of $g(\boldsymbol{\alpha})$ based on $\tilde{\boldsymbol{X}}^{(s)}$.

(ii) The unique MLE of α_j based on $\tilde{\boldsymbol{X}}^{(s)}$ is given by

$$\alpha_j^{*(s)} = -\frac{c_j}{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \quad 1 \le j \le r.$$

$\alpha_1^{*(s)}, \ldots, \alpha_r^{*(s)}$ are jointly independent, and $\alpha_j^{*(s)}$ is inverted gamma distributed with shape parameter c_j and scale parameter $c_j\alpha_j$, $1 \le j \le r$. Moreover, if $g : \mathbb{R}_+ \to \Gamma$ is a bijective function, $g(\alpha_j^{*(s)})$ is the MLE of $g(\alpha_j)$ based on $\tilde{\boldsymbol{X}}^{(s)}$.

(iii) The UMVUE of α_j *based on* $\tilde{\boldsymbol{X}}^{(s)}$ *is given by*

$$\alpha_j^{**(s)} = -\frac{c_j - 1}{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})}, \quad 1 \le j \le r.$$

Proof. (i). From the preliminaries (4.1.4)-(4.1.6), it is easily seen, that $\pi^{-1}(\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$ is the unique solution of the likelihood equation $l_s(\boldsymbol{\alpha}) = \nabla \ln(f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})) \overset{!}{=} 0$ based on the observation $\tilde{\boldsymbol{x}}^{(s)}$. Since l_s is strictly concave on $\mathbb{R}_+^r$, the MLE of $\boldsymbol{\alpha}$ is found (and uniquely determined). The remaining part of the statement follows by application of Thm. 2.2.4 where $\tilde{\boldsymbol{X}}^{(s)}$ is considered a single observation with distribution in the exponential family $\mathfrak{P}_{mix}^{\tilde{\boldsymbol{X}}^{(s)}}$.
(ii). The assertion follows by the same arguments as in (i) by treating all other model parameters as fixed nuisance parameters.
(iii). The statement is obvious from (ii), since $T_j^{(s)}$ is a sufficient and complete statistic for the family $\mathfrak{P}_{j;mix}^{\tilde{\boldsymbol{X}}^{(s)}} = \{P_{\boldsymbol{\alpha}}^{\tilde{\boldsymbol{X}}^{(s)}} : \alpha_j > 0\}$ of distributions, $1 \le j \le r$. ■

Strong consistency of the sequences of estimators is easily obtained by assuming that, roughly speaking, the number of observations which provide information on the respective model parameters tends to infinity when the total number s of observations tends to infinity. In the remaining part of this subsection, the integers m_j and c_j are replaced by $m_j^{(s)}$ and $c_j^{(s)}$, $1 \le j \le r$, in order to express their dependence on the sample size.

Theorem 4.1.3
In the above situation, we find:

(i) If $c_r^{(s)} \overset{s\to\infty}{\longrightarrow} \infty$, *the sequence* $\{\boldsymbol{\alpha}^{*(s)}\}_{s\in\mathbb{N}}$ *of MLEs and the sequence* $\{\boldsymbol{\alpha}^{**(s)}\}_{s\in\mathbb{N}}$, $\boldsymbol{\alpha}^{**(s)} = (\alpha_1^{**(s)}, ..., \alpha_r^{**(s)})'$, $s \in \mathbb{N}$, *of UMVUEs are strongly consistent for estimating* $\boldsymbol{\alpha}$. *Moreover, if* g *in Thm. 4.1.2* (i) *is continuous, the sequences* $\{g(\boldsymbol{\alpha}^{*(s)})\}_{s\in\mathbb{N}}$ *and* $\{g(\boldsymbol{\alpha}^{**(s)})\}_{s\in\mathbb{N}}$ *are strongly consistent for estimating* $g(\boldsymbol{\alpha})$.

(ii) If $c_j^{(s)} \overset{s\to\infty}{\longrightarrow} \infty$, *the sequence* $\{\alpha_j^{*(s)}\}_{s\in\mathbb{N}}$ *of MLEs and the sequence* $\{\alpha_j^{**(s)}\}_{s\in\mathbb{N}}$ *of UMVUEs of* α_j *are strongly consistent for estimating* α_j, $1 \le j \le r$. *Moreover, if* g *in Thm. 4.1.2* (ii) *is continuous, the sequences* $\{g(\alpha_j^{*(s)})\}_{s\in\mathbb{N}}$ *and* $\{g(\alpha_j^{**(s)})\}_{s\in\mathbb{N}}$ *are strongly consistent for estimating* $g(\alpha_j)$, $1 \le j \le r$.

Proof. All statements are clear by showing strong consistency of $\{\alpha_j^{*(s)}\}_{s\in\mathbb{N}}$ if $c_j^{(s)} \overset{s\to\infty}{\longrightarrow} \infty$ is assumed, $1 \le j \le r$. From (3.1.5) in Subsection 3.1, for fixed $j \in \{1, \ldots, r\}$, $-T_{j;i}(\boldsymbol{X}^{(l;i)})$, $1 \le i \le w_j$, $1 \le l \le m_i^{(s)}$, are $\sum_{i=1}^{w_j} m_i^{(s)} = c_j^{(s)}$ iid random variables having exponential distribution with scale parameter α_j^{-1}. Hence, from the strong law of large numbers (e.g, in Shao (2003), Thm. 1.13. (ii), p. 62), we obtain that $-(c_j^{(s)})^{-1} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \overset{s\to\infty}{\longrightarrow} \alpha_j^{-1}$ $P_{\boldsymbol{\alpha}}$-a.s., and the respective convergence of the ratios. ■

The concept of asymptotic efficiency of sequences of estimators introduced in Subsection 2.2.5 (and discussed in 3.2.2 for the iid case) can be extended to the inid case in the following sense (cf. Lehmann & Casella (1998), pp. 475/476, and Bahadur (1964), Section 4). Let $\frac{m_i^{(s)}}{s} \stackrel{s\to\infty}{\longrightarrow} a_i > 0$, $1 \le i \le p$, where $\sum_{i=1}^p a_i = 1$. Then, a sequence $\{\hat{\boldsymbol{\alpha}}^{(s)}\}_{s\in\mathbb{N}}$ of estimators of $\boldsymbol{\alpha}$ is said to be asymptotically efficient provided that

$$\sqrt{s}(\hat{\boldsymbol{\alpha}}^{(s)} - \boldsymbol{\alpha}) \stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}_r(\mathbf{0}, \mathbf{J}(\boldsymbol{\alpha})^{-1}),$$

where

$$J(\boldsymbol{\alpha}) = \sum_{i=1}^{p} a_i \mathbf{I}_{f;i}(\boldsymbol{\alpha})$$

is a convex combination of the Fisher information matrices $\mathbf{I}_{f;i}(\boldsymbol{\alpha})$ of $\{f_{\boldsymbol{\alpha}^{(i)}}^{\boldsymbol{X}^{(1;i)}} \lambda^{r_i}|_{\mathbb{R}^{r_i}_<} : \boldsymbol{\alpha} \in \mathbb{R}^r_+\}$ at $\boldsymbol{\alpha} \in \mathbb{R}^r_+$, $1 \le i \le p$. Here, for technical reasons, the density $f_{\boldsymbol{\alpha}^{(i)}}^{\boldsymbol{X}^{1;i}}$, $1 \le i \le p$, is assumed to formally depend on all model parameters $\alpha_1, \ldots, \alpha_r$.
Now, in the actual case of inid vectors of SOSs, the following (asymptotic) properties of the (sequences of) estimators can be shown.

Theorem 4.1.4
In the above sample situation, we find:

(i) $(-T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)})/c_1^{(s)}, \ldots, -T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})/c_r^{(s)})'$ *is an efficient estimator of* $(\alpha_1^{-1}, \ldots, \alpha_r^{-1})'$ *for every* $s \in \mathbb{N}$.

(ii) The sequence $\{\boldsymbol{\alpha}^{*(s)}\}_{s\in\mathbb{N}}$ *of MLEs and the sequence* $\{\boldsymbol{\alpha}^{**(s)}\}_{s\in\mathbb{N}}$ *of UMVUEs are asymptotically efficient for estimating* $\boldsymbol{\alpha}$*, i.e.*

$$\begin{aligned}\sqrt{s} \quad (\boldsymbol{\alpha}^{*(s)} - \boldsymbol{\alpha}) &\stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}_r(\mathbf{0}, \mathbf{J}(\boldsymbol{\alpha})^{-1}),\\ \sqrt{s} \quad (\boldsymbol{\alpha}^{**(s)} - \boldsymbol{\alpha}) &\stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}_r(\mathbf{0}, \mathbf{J}(\boldsymbol{\alpha})^{-1}),\end{aligned}$$

where

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \sum_{i=1}^{p} a_i \mathbf{diag}\left(\frac{1}{\alpha_1^2}, \ldots, \frac{1}{\alpha_{r_i}^2}, 0, \ldots, 0\right)\\ &= \mathbf{diag}\left(\frac{\sum_{i=1}^{w_1} a_i}{\alpha_1^2}, \ldots, \frac{\sum_{i=1}^{w_r} a_i}{\alpha_r^2}\right). \end{aligned}$$

Moreover, if in Thm. 4.1.2 (i) g *is continuously differentiable with* $|\mathbf{D}_g(\boldsymbol{\alpha})| \ne 0\ \forall \boldsymbol{\alpha} \in \mathbb{R}^r_+$*, then the sequences* $\{g(\boldsymbol{\alpha}^{*(s)})\}_{s\in\mathbb{N}}$ *and* $\{g(\boldsymbol{\alpha}^{**(s)})\}_{s\in\mathbb{N}}$ *are asymptotically efficient for estimating* $g(\boldsymbol{\alpha})$*.*

(iii) $\sqrt{c_j^{(s)}}(\alpha_j^{*(s)} - \alpha_j)$ *and* $\sqrt{c_j^{(s)}}(\alpha_j^{**(s)} - \alpha_j)$ *are asymptotically normal distributed with mean zero and variance* α_j^2, $1 \le j \le r$.

Proof. (i). For every $s \in \mathbb{N}$, the estimator is unbiased for estimating $(\alpha_1^{-1}, \ldots, \alpha_r^{-1})'$, and its covariance matrix is given by $\mathbf{diag}(1/(c_1^{(s)}\alpha_1^2), \ldots, 1/(c_r^{(s)}\alpha_r^2))$ and, thus, it attains the lower bound of the Rao-Cramér inequality (with $g(\boldsymbol{\alpha}) = (\alpha_1^{-1}, \ldots, \alpha_r^{-1})'$, $\boldsymbol{\alpha} \in \mathbb{R}_+^r$).
(ii). For the sequence of MLEs, a proof of the assertion is given in Bradley & Gart (1962). Applying the multivariate version of Slutsky's theorem, the respective result is also obtained for the sequence of UMVUEs. Then, what is left to show follows by the same arguments as in the last part of the proof of La. 2.2.15 (in the Appendix).
(iii). For $1 \leq j \leq r$, it follows from (ii) that

$$\sqrt{s}(\alpha_j^{*(s)} - \alpha_j) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\alpha_j^2}{\sum_{i=1}^{w_j} a_i}\right).$$

Moreover, $\frac{c_j^{(s)}}{s} = \sum_{i=1}^{w_j} \frac{m_i^{(s)}}{s} \xrightarrow{s \to \infty} \sum_{i=1}^{w_j} a_i$ and with Slutsky's theorem

$$\sqrt{c_j^{(s)}}(\alpha_j^{*(s)} - \alpha_j) = \sqrt{\frac{c_j^{(s)}}{s}}\sqrt{s}(\alpha_j^{*(s)} - \alpha_j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha_j^2).$$

Finally, using Slutsky's theorem once again, the remaining statement is obtained. ■

4.1.3 Testing

The UMPU tests derived in Subsections 3.3.1 and 3.3.2 can be established for the inid case as well. In the sample situation of Subsections 4.1.1 and 4.1.2, let $\tilde{\boldsymbol{x}}^{(s)}$ be an observation of the random vector $\tilde{\boldsymbol{X}}^{(s)}$ having $\otimes_{i=1}^{p} \otimes_{l=1}^{m_i} \lambda^{r_i}|_{\mathbb{R}_<^{r_i}}$-density given by (4.1.1).

For fixed $j \in \{1, \ldots, r\}$, we consider the five following test problems.

(I) $H_0 : \alpha_j \leq \alpha_0 \quad \leftrightarrow \quad H_1 : \alpha_j > \alpha_0,$

(II) $H_0 : \alpha_j \geq \alpha_0 \quad \leftrightarrow \quad H_1 : \alpha_j < \alpha_0,$

(III) $H_0 : \alpha_j = \alpha_0 \quad \leftrightarrow \quad H_1 : \alpha_j \neq \alpha_0,$

(IV) $H_0 : \alpha_0^{(1)} \leq \alpha_j \leq \alpha_0^{(2)} \quad \leftrightarrow \quad H_1 : \alpha_j < \alpha_0^{(1)}$ or $\alpha_j > \alpha_0^{(2)},$

(V) $H_0 : \alpha_j \leq \alpha_0^{(1)}$ or $\alpha_j \geq \alpha_0^{(2)} \quad \leftrightarrow \quad H_1 : \alpha_0^{(1)} < \alpha_j < \alpha_0^{(2)},$

where $\alpha_0, \alpha_0^{(1)}$ and $\alpha_0^{(2)}$ are positive constants with $\alpha_0^{(1)} < \alpha_0^{(2)}$.

In the iid case, the proofs of Subsections 3.3.1 and 3.3.2 were essentially based on the fact that the statistics $T_1^{(s)}(\tilde{\boldsymbol{X}}^{(s)}), \ldots, T_r^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ were independent with $-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ having a gamma distribution with shape parameter s and scale parameter α_j^{-1}, $1 \leq j \leq r$. Regarding the inid case, these findings remain true if we replace the shape parameter s of the gamma distribution of $-T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$ by c_j, $1 \leq j \leq r$. Then, respective assertions can similarly be shown. We state our findings but, for the named reasons, the proofs are omitted.

Theorem 4.1.5
Let $\alpha \in (0,1)$ and $j \in \{1,\dots,r\}$. Then, the following assertions hold true:

(i) For test problem (I),

$$\varphi^* : \quad \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}_{<}^{r_i} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto \mathbb{1}_{\left(-\frac{\chi^2_{\alpha}(2c_j)}{2\alpha_0},\infty\right)}(T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on $\tilde{\boldsymbol{x}}^{(s)}$, where $\chi^2_{\alpha}(2c_j)$ denotes the α-quantile of the χ^2-distribution with $2c_j$ degrees of freedom.

(ii) For test problem (II),

$$\varphi^* : \quad \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}_{<}^{r_i} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto \mathbb{1}_{\left(-\infty,-\frac{\chi^2_{1-\alpha}(2c_j)}{2\alpha_0}\right)}(T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on $\tilde{\boldsymbol{x}}^{(s)}$, where $\chi^2_{1-\alpha}(2c_j)$ denotes the $(1-\alpha)$-quantile of the χ^2-distribution with $2c_j$ degrees of freedom.

(iii) For test problem (III),

$$\varphi^* : \quad \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}_{<}^{r_i} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto 1 - \mathbb{1}_{\left(\tau_1^{(\alpha,\alpha_0,c_j)}(\beta^*),\tau_2^{(\alpha,\alpha_0,c_j)}(\beta^*)\right)}(T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on $\tilde{\boldsymbol{x}}^{(s)}$, where β^* is the unique solution of the equation

$$F_{\alpha_0,c_j+1}(F^{-1}_{\alpha_0,c_j}(1-\alpha+\beta)) - F_{\alpha_0,c_j+1}(F^{-1}_{\alpha_0,c_j}(\beta)) \overset{!}{=} 1-\alpha$$

with respect to $\beta \in (0,\alpha)$, and the mapping $\tau^{(\alpha,\alpha_0,c_j)}$ is given by (3.3.1).

(iv) For test problem (IV),

$$\varphi^* : \quad \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}_{<}^{r_i} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto 1 - \mathbb{1}_{\left(\tau_1^{(\alpha,\alpha_0^{(1)},c_j)}(\beta^*),\tau_2^{(\alpha,\alpha_0^{(1)},c_j)}(\beta^*)\right)}(T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on $\tilde{\boldsymbol{x}}^{(s)}$, where β^* is the unique solution of the equation

$$F_{\alpha_0^{(2)},c_j}(F^{-1}_{\alpha_0^{(1)},c_j}(1-\alpha+\beta)) - F_{\alpha_0^{(2)},c_j}(F^{-1}_{\alpha_0^{(1)},c_j}(\beta)) \overset{!}{=} 1-\alpha$$

with respect to $\beta \in (0,\alpha)$, and $\tau^{(\alpha,\alpha_0^{(1)},c_j)}$ is given by (3.3.1).

(v) For test problem (V),

$$\varphi^* : \quad \times_{i=1}^{p} \times_{l=1}^{m_i} \mathbb{R}_{<}^{r_i} \to [0,1] : \quad \tilde{\boldsymbol{x}}^{(s)} \mapsto \mathbb{1}_{\left(\tau_1^{(1-\alpha,\alpha_0^{(1)},c_j)}(\beta^*),\tau_2^{(1-\alpha,\alpha_0^{(1)},c_j)}(\beta^*)\right)}(T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))$$

is a UMPU level-α test based on $\tilde{\boldsymbol{x}}^{(s)}$, where β^* is the unique solution of the equation

$$F_{\alpha_0^{(2)},c_j}(F^{-1}_{\alpha_0^{(1)},c_j}(\alpha+\beta)) - F_{\alpha_0^{(2)},c_j}(F^{-1}_{\alpha_0^{(1)},c_j}(\beta)) \overset{!}{=} \alpha$$

with respect to $\beta \in (0,1-\alpha)$, and $\tau^{(1-\alpha,\alpha_0^{(1)},c_j)}$ is given by (3.3.1).

We end this subsection by deriving the LR test statistic, Wald's (modified) statistic and Rao's score statistic for the test problem

$$H_0 : \alpha_1 = \cdots = \alpha_r = 1 \quad \leftrightarrow \quad H_1 : \exists j_0 \in \{1, \ldots, r\} : \alpha_{j_0} \neq 1 \tag{4.1.7}$$

with a simple null hypothesis. As in the iid case, this test problem tackles the question whether the data is obtained from OSs based on F or not. When dealing with (possibly sequential) systems, the null hypothesis coincides with the assumption that all p systems are of common type with underlying distribution function F.

Firstly, we obtain from La. 2.1.30, (4.1.3) and (4.1.6) that the Kullback-Leibler distance of $\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}} \in \mathbb{R}_+^r$ equals

$$\begin{aligned} d_{KL}(\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}) &= \kappa(\tilde{\boldsymbol{\alpha}}) - \kappa(\boldsymbol{\alpha}) + (\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}})' \pi(\boldsymbol{\alpha}) \\ &= -\sum_{j=1}^r c_j \ln(\tilde{\alpha}_j) + \sum_{j=1}^r c_j \ln(\alpha_j) - \sum_{j=1}^r (\alpha_j - \tilde{\alpha}_j) \frac{c_j}{\alpha_j} \\ &= \sum_{j=1}^r c_j \left[\frac{\tilde{\alpha}_j}{\alpha_j} - \ln\left(\frac{\tilde{\alpha}_j}{\alpha_j}\right) - 1 \right]. \end{aligned}$$

In particular, by setting $\tilde{\boldsymbol{\alpha}} = \underline{\mathbf{1}} = (1, \ldots, 1)'$,

$$d_{KL}(\boldsymbol{\alpha}, \underline{\mathbf{1}}) = \sum_{j=1}^r c_j \left[\frac{1}{\alpha_j} - \ln\left(\frac{1}{\alpha_j}\right) - 1 \right].$$

According to La. 2.3.8, the test statistic of the LR test based on $\tilde{\boldsymbol{X}}^{(s)}$ can be derived by plugging in the MLE $\boldsymbol{\alpha}^{*(s)}$ of $\boldsymbol{\alpha}$ based on $\tilde{\boldsymbol{X}}^{(s)}$ (cf. Thm. 4.1.2), i.e.

$$\begin{aligned} T_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= d_{KL}(\boldsymbol{\alpha}^{*(s)}, \underline{\mathbf{1}}) \\ &= \sum_{j=1}^r c_j \left[\frac{1}{\alpha_j^{*(s)}} - \ln\left(\frac{1}{\alpha_j^{*(s)}}\right) - 1 \right] \\ &= \sum_{j=1}^r c_j \left[-\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - \ln\left(-\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})\right) - 1 \right] \\ &= \sum_{j=1}^r c_j [Y_j - \ln(Y_j) - 1], \end{aligned}$$

where $Y_j = -c_j^{-1} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$, $1 \leq j \leq r$, are jointly independent and $Y_j \sim \Gamma(c_j, (c_j \alpha_j)^{-1})$, $1 \leq j \leq r$.

Thus, for test problem (4.1.7), the LR test with level α based on the observation $\tilde{\boldsymbol{x}}^{(s)}$ is given by

$$\varphi_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(d_s(\alpha), \infty)} \left(\sum_{j=1}^r c_j \left[-\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - \ln\left(-\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right) - 1 \right] \right), \tag{4.1.8}$$

and the critical value $d_s(\alpha)$ is derived from the equation

$$P\left(\sum_{j=1}^{r} c_j[Z_j - \ln(Z_j) - 1] \leq d_s(\alpha)\right) = 1 - \alpha, \tag{4.1.9}$$

where $Z_1, \ldots, Z_r$ are jointly independent random variables with $Z_j \sim \Gamma(c_j, c_j^{-1})$, $1 \leq j \leq r$. In virtue of (2.3.11) and (4.1.2), the test statistic of Wald's test based on $\tilde{\boldsymbol{x}}^{(s)}$ is represented by

$$\begin{aligned} T_W^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= (\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}})'\mathbf{diag}\left(\frac{c_1}{(\alpha_1^{*(s)})^2}, \ldots, \frac{c_r}{(\alpha_r^{*(s)})^2}\right)(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}}) \\ &= \sum_{j=1}^{r} c_j \left(\frac{1}{\alpha_j^{*(s)}} - 1\right)^2 \\ &= \sum_{j=1}^{r} c_j \left(\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + 1\right)^2 \\ &= \sum_{j=1}^{r} c_j \left(Y_j - 1\right)^2 \end{aligned}$$

with the Y's as defined above, and, again, it coincides with the test statistic of Rao's score test given by (2.3.13), since, by application of Thm. 2.1.22,

$$\begin{aligned} T_R^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= (\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - E_{\underline{\mathbf{1}}}(\boldsymbol{T}^{(s)}))'\mathbf{diag}\left(\frac{1}{c_1}, \ldots, \frac{1}{c_r}\right)(\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - E_{\underline{\mathbf{1}}}(\boldsymbol{T}^{(s)})) \\ &= \sum_{j=1}^{r} \frac{1}{c_j} \left(T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + c_j\right)^2 \\ &= \sum_{j=1}^{r} c_j \left(\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) + 1\right)^2. \end{aligned}$$

Hence, for test problem (4.1.7), Wald's test, respectively Rao's score test with level α based on the observation $\tilde{\boldsymbol{x}}^{(s)}$ is

$$\varphi_{W,R}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(r_s(\alpha),\infty)}\left(\sum_{j=1}^{r} c_j \left(\frac{1}{c_j} T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) + 1\right)^2\right), \tag{4.1.10}$$

where the critical value $r_s(\alpha)$ is such that

$$P\left(\sum_{j=1}^{r} c_j \left(Z_j - 1\right)^2 \leq r_s(\alpha)\right) = 1 - \alpha, \tag{4.1.11}$$

with the Z's as above.
Finally, Wald's modified test statistic is

$$\begin{aligned} T_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= (\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}})' \mathbf{I}_f^{(s)}(\underline{\mathbf{1}})(\boldsymbol{\alpha}^{*(s)} - \underline{\mathbf{1}}) \\ &= \sum_{j=1}^{r} c_j (\alpha_j^{*(s)} - 1)^2 \\ &= \sum_{j=1}^{r} c_j \left(\frac{c_j}{T_j^{(s)}(\tilde{\boldsymbol{X}}^{(s)})} + 1 \right)^2 \\ &= \sum_{j=1}^{r} c_j \left(\frac{1}{Y_j} - 1 \right)^2, \end{aligned}$$

with the Y's as above, and, thus, Wald's modified test with level α based on the observation $\tilde{\boldsymbol{x}}^{(s)}$ is given by

$$\varphi_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) = \mathbb{1}_{(\tilde{w}_s(\alpha), \infty)} \left(\sum_{j=1}^{r} c_j \left(\frac{c_j}{T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})} + 1 \right)^2 \right), \tag{4.1.12}$$

where the critical value $\tilde{w}_s(\alpha)$ is obtained from the equation

$$P\left(\sum_{j=1}^{r} c_j \left(\frac{1}{Z_j} - 1 \right)^2 \leq \tilde{w}_s(\alpha) \right) = 1 - \alpha \tag{4.1.13}$$

with the Z's as above.

Theorem 4.1.6
For test problem (4.1.7) with a simple null hypothesis, the following assertions hold true:

(i) Wald's test and Rao's score test based on the observation $\tilde{\boldsymbol{x}}^{(s)}$ coincide and are given in virtue of (4.1.10) and (4.1.11).

(ii) Wald's modified test based on the observation $\tilde{\boldsymbol{x}}^{(s)}$ is represented by (4.1.12) and (4.1.13).

(iii) The LR test based on the observation $\tilde{\boldsymbol{x}}^{(s)}$ is obtained from (4.1.8) and (4.1.9).

4.2 SOSs with Partially Unknown Baseline Distribution

In this section, we extend the findings of Chapter 3 to the case in which the baseline distribution of SOSs with conditional proportional hazard rates is partially unknown. For this, we assume that the underlying distribution function F is of the form

$$F(t) = 1 - e^{-\lambda g(t-\mu)}, \quad t \geq \mu, \tag{4.2.1}$$

with location parameter $\mu \in \mathbb{R}$ and rate parameter $\lambda > 0$, where the function $g : [0, \infty) \to [0, \infty)$ is strictly increasing, continuously differentiable on $(0, \infty)$ and satisfies $g(0) = 0$ and $\lim_{t\to\infty} g(t) = \infty$ (cf. Cramer & Kamps (1996, 2001a)). Throughout this section, g is assumed to be known. Different well-known distributions are included in this set-up by choosing the respective function g, e.g.,

- two-parameter exponential distribution by setting $g(t) = t$,
- shifted Weibull distribution by setting $g(t) = t^\beta$ for some known $\beta > 0$, and
- shifted Pareto distribution by setting $g(t) = \ln(t + 1)$.

Here, as a consequence of (4.2.1), the distribution function F_j, $1 \leq j \leq n$, defined by (3.0.1), can be represented as

$$F_j(t) = 1 - (1 - F(t))^{\alpha_j} = 1 - e^{-\lambda\alpha_j g(t-\mu)}, \quad t \geq \mu.$$

4.2.1 Known Location Parameter

In the following, we consider the joint density of the first r SOSs with conditional proportional hazard rates based on a partially unknown distribution function F given by (4.2.1) and model parameters $\alpha_1, \ldots, \alpha_n$, where the location parameter μ is assumed to be known, whereas the rate parameter λ and the model parameter $\alpha_1, \ldots, \alpha_n$ are unknown. Then, the statistics given by (3.1.1) simplify to

$$\begin{aligned} T_1(\boldsymbol{x}) &= n \ln(1 - F(x_1)) = -n\lambda g(x_1 - \mu), \\ T_j(\boldsymbol{x}) &= (n - j + 1) \ln\left(\frac{1 - F(x_j)}{1 - F(x_{j-1})}\right) \\ &= -(n - j + 1)\lambda(g(x_j - \mu) - g(x_{j-1} - \mu)), \quad 2 \leq j \leq r, \end{aligned}$$

where $\boldsymbol{x} = (x_1, \ldots, x_r)' \in \mathbb{R}^r_<$. Introducing the parametrization $\eta_j = \lambda\alpha_j$, $1 \leq j \leq n$, and the statistics

$$\begin{aligned} H_1(\boldsymbol{x}) &= \frac{T_1(\boldsymbol{x})}{\lambda} = -ng(x_1 - \mu), \\ H_j(\boldsymbol{x}) &= \frac{T_j(\boldsymbol{x})}{\lambda} = -(n - j + 1)(g(x_j - \mu) - g(x_{j-1} - \mu)), \quad 2 \leq j \leq r, \end{aligned}$$

where $\boldsymbol{x} = (x_1, \ldots, x_r)' \in \mathbb{R}^r_<$, the densities (3.1.2) can be rewritten as

$$\begin{aligned} f^{\boldsymbol{X}}_{\boldsymbol{\alpha}}(\boldsymbol{x}) &= \left(\prod_{j=1}^r \alpha_j\right) \exp\left\{\sum_{j=1}^r \alpha_j \, T_j\,(\boldsymbol{x})\right\} \left(\frac{n!}{(n-r)!} \prod_{j=1}^r \frac{f(x_j)}{1-F(x_j)}\right) \\ &= \left(\prod_{j=1}^r \alpha_j\right) \exp\left\{\sum_{j=1}^r \alpha_j \lambda H_j(\boldsymbol{x})\right\} \left(\frac{n!}{(n-r)!} \prod_{j=1}^r \lambda g'(x_j - \mu)\right) \\ &= \left(\prod_{j=1}^r \eta_j\right) \exp\left\{\sum_{j=1}^r \eta_j H_j(\boldsymbol{x})\right\} \left(\frac{n!}{(n-r)!} \prod_{j=1}^r g'(x_j - \mu)\right). \end{aligned}$$

Obviously, in the considered set-up, the exponential family structure of the densities is preserved, and, by defining $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)'$, $\boldsymbol{H} = (H_1, \ldots, H_r)'$, and

$$\tilde{h}(\boldsymbol{x}) = \frac{n!}{(n-r)!} \prod_{j=1}^r g'(x_j - \mu), \quad \boldsymbol{x} = (x_1, \ldots, x_r)' \in \mathbb{R}^r_<,$$

$\tilde{\mathfrak{P}}^{\boldsymbol{X}} = \{\tilde{P}^{\boldsymbol{X}}_{\boldsymbol{\eta}} = \tilde{f}^{\boldsymbol{X}}_{\boldsymbol{\eta}} \lambda^r|_{\mathbb{R}^r_<} : \boldsymbol{\eta} \in \mathbb{R}^r_+\}$ forms a r-parametrical exponential family in $\eta_1, \ldots, \eta_r$ and $H_1, \ldots, H_r$, where

$$\tilde{f}^{\boldsymbol{X}}_{\boldsymbol{\eta}}(\boldsymbol{x}) = C(\boldsymbol{\eta}) e^{\boldsymbol{\eta}' \boldsymbol{H}(\boldsymbol{x})} \tilde{h}(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, \ldots, x_r)' \in \mathbb{R}^r_<,$$

with the same C as introduced in Chapter 3, i.e.

$$C(\boldsymbol{\eta}) = \prod_{j=1}^r \eta_j, \quad \boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)' \in \mathbb{R}^r_+.$$

Hence, replacing α_j by η_j and T_j by H_j, $1 \leq j \leq r$, all mathematical results derived in Chapter 3, in particular, the inferential statements of Sections 3.2 and 3.3, remain true in the actual case with some minor changes in interpretation.

Chapter 5

Simulation Study

5.1 Univariate Tests

To illustrate the theoretical results of Subsection 3.3.2, we consider a 3-out-of-5 system with baseline distribution function

$$F(t) = 1 - e^{-1000^{-1}t}, \quad t > 0, \tag{5.1.1}$$

and model parameters $\alpha_1 = 1$, α_2, $\alpha_3 = 2.5$, and the two-sided test problems

$$H_0 : \ \alpha_2 = 1 \quad \leftrightarrow \quad H_1 : \ \alpha_2 \neq 1 \tag{5.1.2}$$

and

$$H_0 : \ 0.7 \leq \alpha_2 \leq 1.3 \quad \leftrightarrow \quad H_1 : \ \alpha_2 < 0.7 \text{ or } \alpha_2 > 1.3, \tag{5.1.3}$$

where, in the latter case, the alternative is given by the complement of an interval. Figure 5.1 shows the empirical power functions of the corresponding UMPU tests at $\alpha_2 \in (0, 2)$ based on 1,000,000 simulated samples of size $s = 40$, where the level is given by $\alpha = 0.05$. Throughout this section, in all figures, the legends are omitted, since the assignments are clear.
Similarly, the alternative may be chosen as an open interval, e.g.,

$$H_0 : \ \alpha_2 \leq 0.3 \text{ or } \alpha_2 \geq 1.7 \quad \leftrightarrow \quad H_1 : \ 0.3 < \alpha_2 < 1.7, \tag{5.1.4}$$

$$H_0 : \ \alpha_2 \leq 0.5 \text{ or } \alpha_2 \geq 1.5 \quad \leftrightarrow \quad H_1 : \ 0.5 < \alpha_2 < 1.5, \tag{5.1.5}$$

$$H_0 : \ \alpha_2 \leq 0.7 \text{ or } \alpha_2 \geq 1.3 \quad \leftrightarrow \quad H_1 : \ 0.7 < \alpha_2 < 1.3 \tag{5.1.6}$$

or

$$H_0 : \ \alpha_2 \leq 0.9 \text{ or } \alpha_2 \geq 1.1 \quad \leftrightarrow \quad H_1 : \ 0.9 < \alpha_2 < 1.1. \tag{5.1.7}$$

Figure 5.2 shows the empirical power functions of the corresponding UMPU tests at $\alpha_2 \in (0, 2)$ based on 1,000,000 simulated samples of size $s = 40$ (level $\alpha = 0.05$). Note, that for sufficiently large intervals, the power of this kind of test is satisfactory and, hence, it can actually be used in applications.

Figure 5.1: Empirical power functions of the UMPU tests corresponding to the hypotheses (5.1.2) and (5.1.3) based on 1,000,000 simulated samples of size $s = 40$, where the level α is given by 0.05.

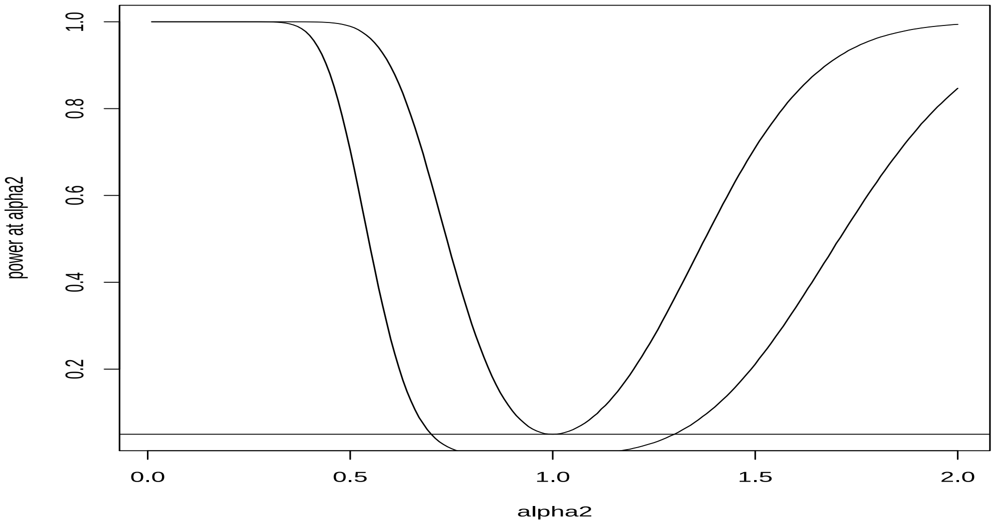
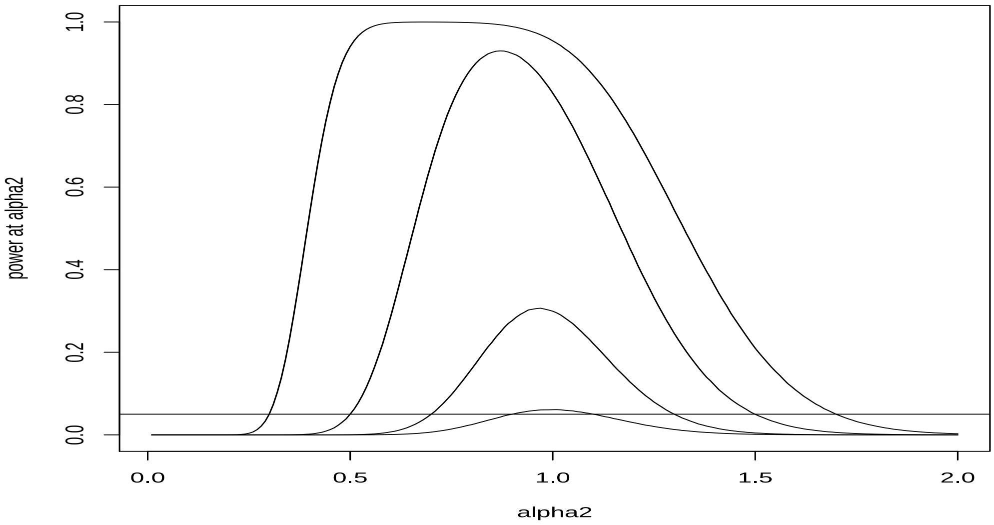


Figure 5.2: Empirical power functions of the UMPU tests corresponding to the hypotheses (5.1.4)-(5.1.7) based on 1,000,000 simulated samples of size $s = 40$, where the level α is given by 0.05.



Certainly, it is of some interest to compare the derived UMPU tests with the short-cut tests in Cramer and Kamps (1996,2001) in terms of the power function. For this, we consider, once again, a 3-out-of-5 system with the above baseline distribution function F and model parameters $\alpha_1 = \alpha_2 = 1$ and α_3. In the test problem

$$H_0 : \ \alpha_3 = 1 \quad \leftrightarrow \quad H_1 : \ \alpha_3 \neq 1, \tag{5.1.8}$$

the null hypothesis coincides with the assumption of a common 3-out-of-5 system with underlying baseline distribution function F. For short, let $\underline{\mathbf{1}} = (1, 1, 1)'$.
Using the denotations of Chapter 3, Cramer & Kamps (1996) proposed the following test statistics and decision rules based on a sample of size s:

(i) *Test A:* Reject H_0 if the ratio

$$Q_A(\tilde{\boldsymbol{x}}^{(s)}) = \frac{\max\{T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_2^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\}}{\min\{T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_2^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\}}$$

is too small. The critical value is derived from the equation

$$P_{\underline{\mathbf{1}}}^{\tilde{\boldsymbol{X}}^{(s)}}(Q_A \leq c) = 1 - \frac{3}{(s-1)!} \int_0^\infty \left(\sum_{i=0}^{s-1} \frac{z^i}{i!} (c^i e^{-cz} - e^{-z}) \right)^2 z^{s-1} e^{-z} dz.$$

(ii) *Test B:* Reject H_0 if the ratio

$$Q_B(\tilde{\boldsymbol{x}}^{(s)}) = \frac{T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}{T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) + T_2^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) + T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}$$

is too small or too large. Here, $Q_B(\tilde{\boldsymbol{X}}^{(s)})$ has a beta distribution with shape parameters s and $2s$ if H_0 is true and, thus, the critical values are the respective $\alpha/2$-quantile and $(1-\alpha/2)$-quantile from that distribution.

(iii) *Test C:* Reject H_0 if the difference

$$Q_C(\tilde{\boldsymbol{x}}^{(s)}) = \max\{T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_2^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\} - \min\{T_1^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_2^{(s)}(\tilde{\boldsymbol{x}}^{(s)}), T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\}$$

is too large. The critical value is derived from the equation

$$P_{\underline{\mathbf{1}}}^{\tilde{\boldsymbol{X}}^{(s)}}(Q_C \leq c) = \frac{3}{(s-1)!} \int_0^\infty (g(z) - g(z+c))^2 z^{s-1} e^{-z} dz,$$

where $g(z) = e^{-z} \sum_{i=0}^{s-1} z^i/i!$, $z \in \mathbb{R}$.

(iv) *LR test:* Reject H_0 if

$$Q_{LR}(\tilde{\boldsymbol{x}}^{(s)}) = -\frac{1}{s}\ln\left(\frac{\sup\limits_{\alpha_3=1} f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}{\sup\limits_{\alpha_3>0} f_{\boldsymbol{\alpha}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})}\right) = -\frac{1}{s}T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - \ln\left(-\frac{1}{s}T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right) - 1 \quad (5.1.9)$$

is too large. The representation on the right-hand side of (5.1.9) results by inserting the maximum likelihood estimate $-s/T_3^{(s)}(\tilde{\boldsymbol{x}}^{(s)})$ of α_3 in the denominator. If H_0 is true, $-T_3^{(s)}(\tilde{\boldsymbol{X}}^{(s)})/s$ has a gamma distribution with shape parameter s and scale parameter s^{-1} and, thus, the critical value can, by means of simulations, be computed using the equation

$$P_{\underline{\mathbf{1}}}^{\tilde{\boldsymbol{X}}^{(s)}}(Q_{LR} \le c) = P(Z - \ln(Z) - 1 \le c),$$

where $Z \sim \Gamma(s, s^{-1})$.

In the Figures 5.3-5.5, the empirical power functions of the UMPU test and the short cut tests A, B and C corresponding to test problem (5.1.8) are compared at $\alpha_3 \in (0, 2)$ on the basis of $300,000$ simulated samples of size $s = 40$ (level $\alpha = 0.05$). For $\alpha_3 \in (0, 1)$, the power of tests A and C differ comparatively little from the power of the UMPU test, whereas the power of test B is not satisfying. In the (certainly more interesting) case of $\alpha_3 \in (1, 2)$, test B and C perform very poorly in comparison with the UMPU test, whereas the power of test A behaves again considerably better. Moreover, it turns out that in the given situation, the power functions of the UMPU and the LR test differ from each other only marginally. More precisely, the maximum of the difference of both power functions taken over all values $\alpha_3 \in (0, 2)$ equals $1.33 \cdot 10^{-4}$. Here, these tests might coincide (cf. Shao (2003), Proposition 6.5., p. 429).

Figure 5.3: Empirical power functions of the UMPU test and short-cut test A corresponding to test problem (5.1.8) based on 300,000 simulated samples of size $s = 40$, where the level α is given by 0.05.

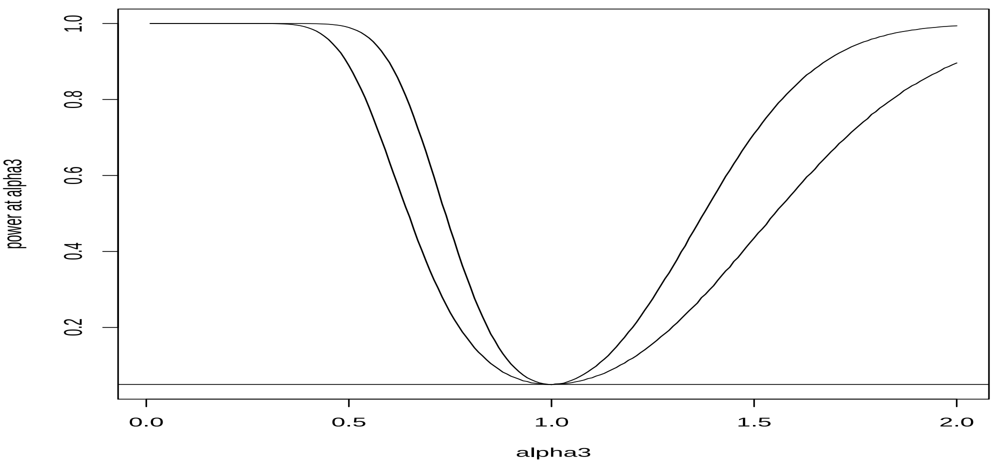


Figure 5.4: Empirical power functions of the UMPU test and short-cut test B corresponding to test problem (5.1.8) based on 300,000 simulated samples of size $s = 40$, where the level α is given by 0.05.

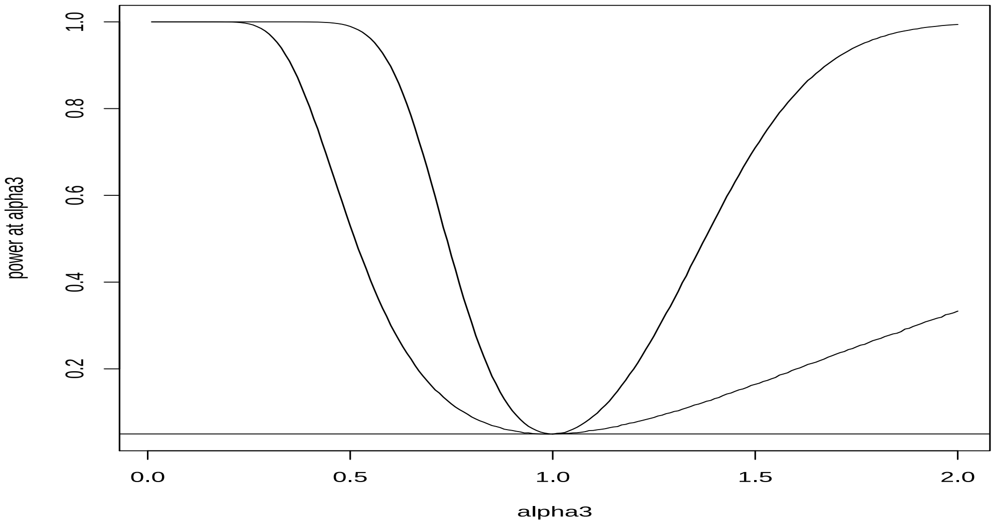
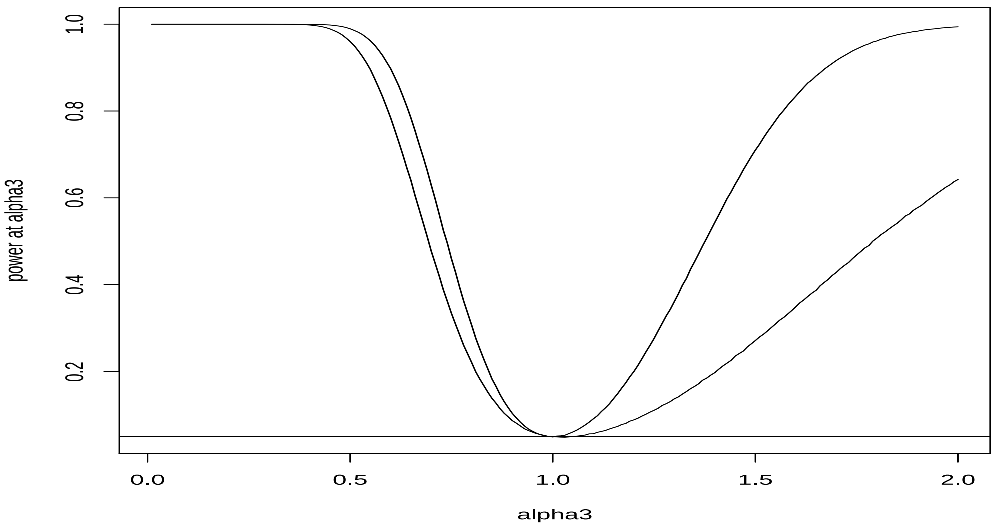


Figure 5.5: Empirical power functions of the UMPU test and short-cut test C corresponding to test problem (5.1.8) based on 300,000 simulated samples of size $s = 40$, where the level α is given by 0.05.



5.2 Multivariate Tests

If in the context of test problem (5.1.8) from Section 5.1 no prior information on α_2 is available, the LR test, Rao's score test and Wald's (modified) test can be applied to test the type of the system.
We consider, once again, a 3-out-of-5 system with baseline distribution function F given by (5.1.1) and model parameters $\alpha_1 = 1$, α_2 and α_3. The test problem

$$H_0 : \alpha_2 = \alpha_3 = 1 \quad \leftrightarrow \quad H_1 : \alpha_2 \neq 1 \text{ or } \alpha_3 \neq 1, \tag{5.2.1}$$

corresponds to the question whether the system is of the usual or the sequential type, and, hence, whether the model of common OSs or the more flexible model of SOSs should be preferred to describe the lifetime of the system and its components.
Based on a sample of size $s \in \{10, 50, 90\}$, we consider the LR test, Wald's or, which is the same here, Rao's score test, and Wald's modified test, which are given by

$$\begin{aligned}
\varphi_{LR}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= \mathbb{1}_{(d_s(\alpha),\infty)}\left(\sum_{j=2}^{3}\left[-\frac{1}{s}T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - \ln\left(-\frac{1}{s}T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right) - 1\right]\right), \\
\varphi_{W,R}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= \mathbb{1}_{(r_s(\alpha),\infty)}\left(s\sum_{j=2}^{3}\left(\frac{1}{s}T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) + 1\right)^2\right), \\
\varphi_{\tilde{W}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) &= \mathbb{1}_{(\tilde{w}_s(\alpha),\infty)}\left(s\sum_{j=2}^{3}\left(\frac{s}{T_j^{(s)}(\tilde{\boldsymbol{x}}^{(s)})} + 1\right)^2\right)
\end{aligned}$$

(cf. (3.3.11), (3.3.13) and (3.3.15)). According to (3.3.12), (3.3.14) and (3.3.16) with level $\alpha = 0.05$, the corresponding critical values of the tests are each computed empirically by 10,000,000 realizations of the respective test statistic. We obtain

$$\begin{aligned}
\tilde{d}_{10} &= 6.086159, \quad r_{10} = 6.05324, \quad \tilde{w}_{10} = 12.85457, \\
\tilde{d}_{50} &= 6.011769, \quad r_{50} = 5.986285, \quad \tilde{w}_{50} = 7.074217, \\
\tilde{d}_{90} &= 6.006873, \quad r_{90} = 5.989459, \quad \tilde{w}_{90} = 6.551305,
\end{aligned}$$

where $\tilde{d}_s(\alpha) = 2sd_s(\alpha)$, $s \in \mathbb{N}$. If H_0 is true, all three test statistics are asymptotically $\chi^2(2)$-distributed, i.e. exponentially distributed with scale parameter 2, and the respective 0.95-quantile of the asymptotic distribution is given by $\chi^2_{0.95}(2) = -2\ln(1-0.95) = 5.991465$.
Figures 5.6-5.8 show the empirical power functions of the LR test, Wald's test/Rao's score test and Wald's modified test at $(\alpha_2, \alpha_3)'$ in a neighbourhood of $(1,1)'$, where the graphics are each based on 200,000 simulated samples of size $s = 10$, respectively $s = 50$ and $s = 90$, and the level is given by $\alpha = 0.05$. For fixed $\alpha_2 \in \{0.6, 0.8, 1, 1.2, 1.4\}$, the corresponding cross sections of the empirical power functions are illustrated in Figures 5.12-5.14. Related power maps can be found in Figures 5.9-5.11 and show which test is best with respect to a fixed alternative $(\alpha_2, \alpha_3)' \in \{0.6, 0.65, \dots, 1.4\}^2$.

Figure 5.6: Empirical power functions of the LR test with level $\alpha = 0.05$ corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s \in \{10, 50, 90\}$.

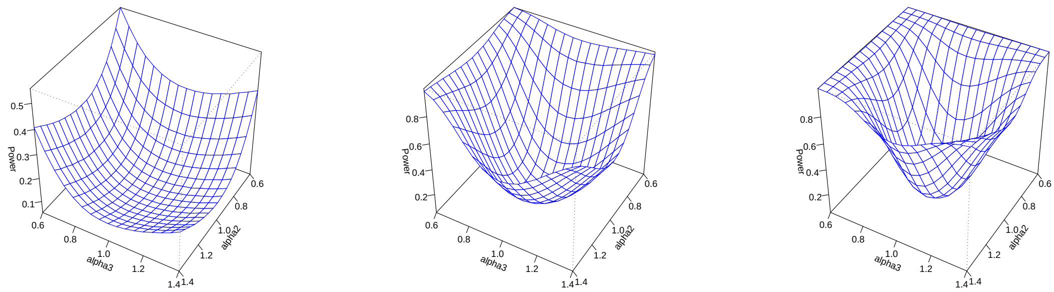


Figure 5.7: Empirical power functions of Wald's test/Rao's score test with level $\alpha = 0.05$ corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s \in \{10, 50, 90\}$.

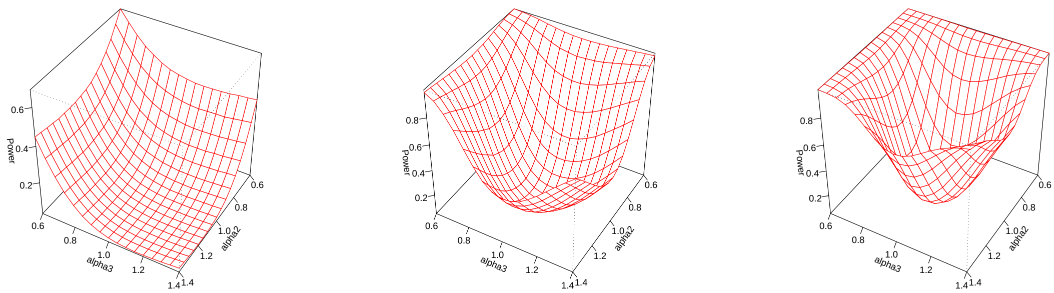


Figure 5.8: Empirical power functions of Wald's modified test with level $\alpha = 0.05$ corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s \in \{10, 50, 90\}$.

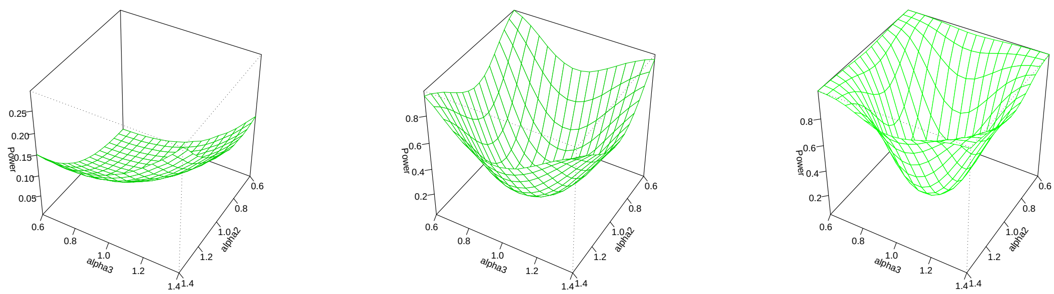


Figure 5.9: Power map of the LR test, Wald's/Rao's score test and Wald's modified test corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s = 10$, where the level α is given by 0.05.

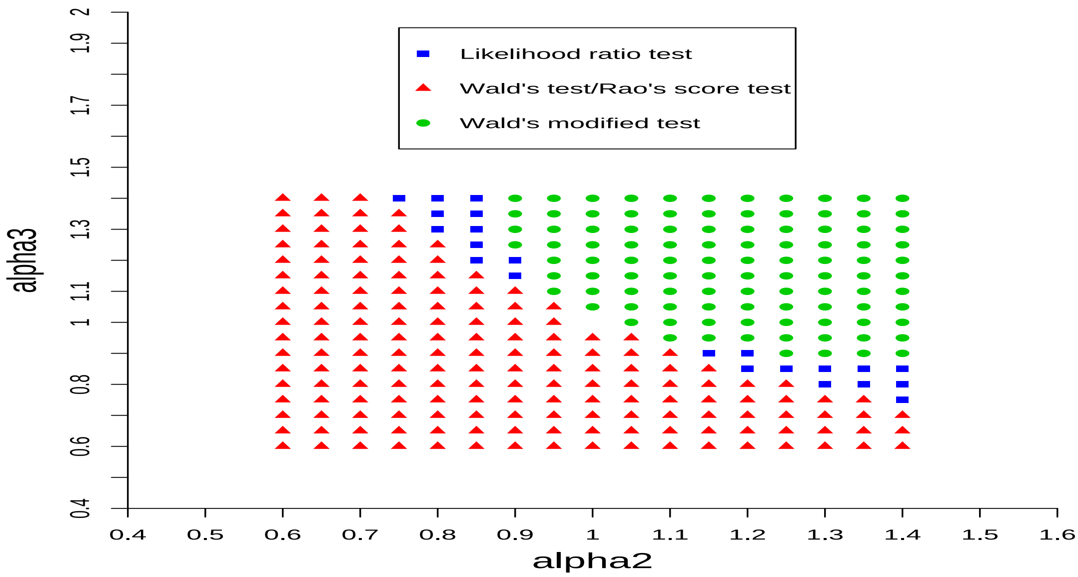


Figure 5.10: Power map of the LR test, Wald's/Rao's score test and Wald's modified test corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s = 50$, where the level α is given by 0.05.

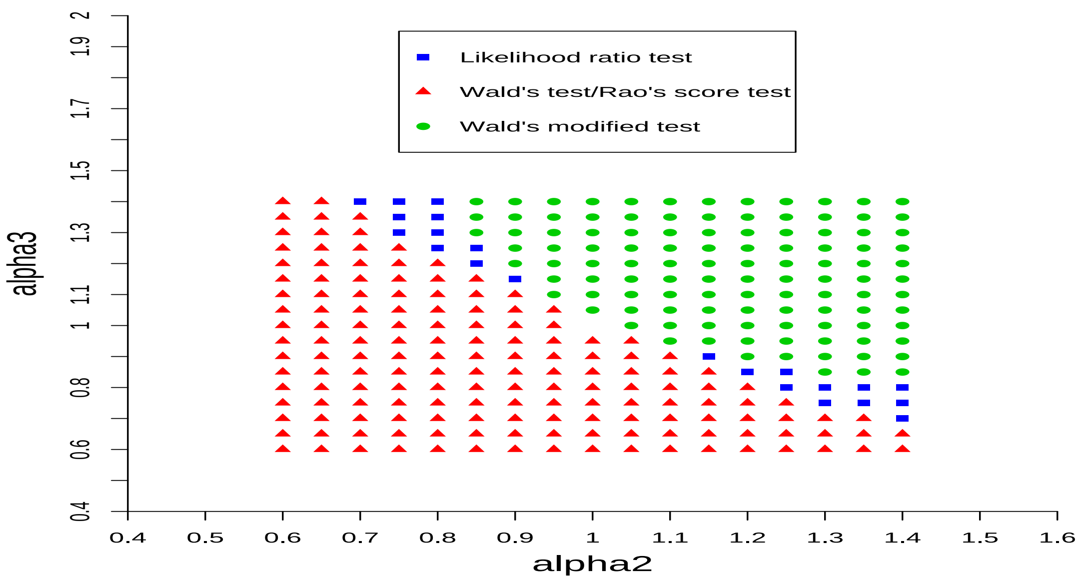
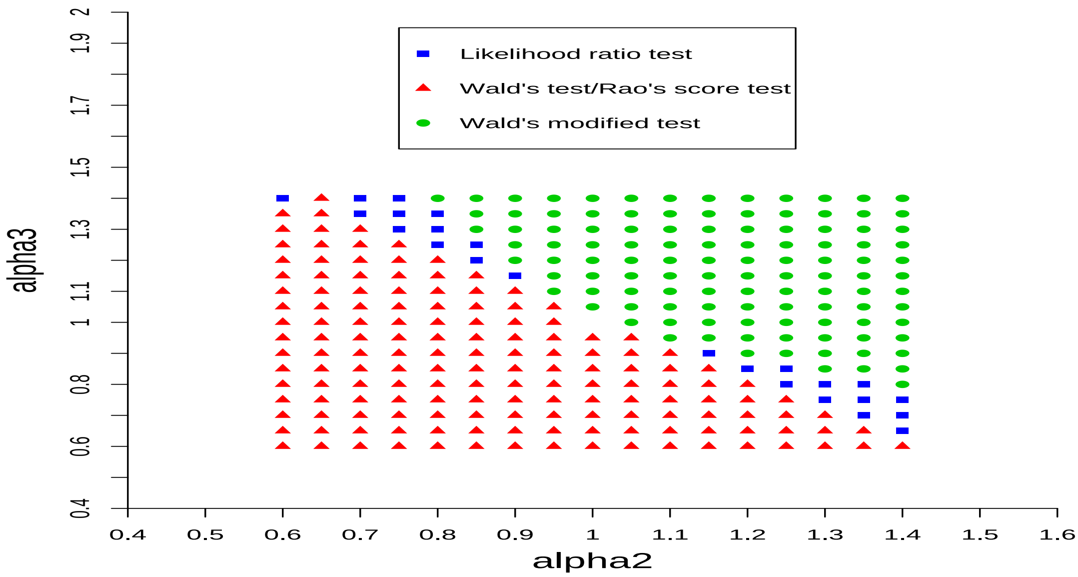


Figure 5.11: Power map of the LR test, Wald's/Rao's score test and Wald's modified test corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s = 90$, where the level α is given by 0.05.



From the cross sections in Figures 5.12-5.14, it is seen that Wald's/Rao's score test and Wald's modified test are biased if the sample size is small, whereas the LR test seems to be unbiased. Moreover, the simulation study suggests that, in the considered case, all three tests are consistent and thus, in particular, asymptotically unbiased. This result might also be true in general (see Rao (2005), p. 10). From the power maps in Figures 5.9-5.11, we conclude that neither the LR test nor Wald's/Rao's score test or Wald's modified test uniformly dominates the other ones in terms of power. If α_2 and α_3 are both less than 1, Wald's/Rao's score test is best, but the test is worst if the α's are greater than 1. For Wald's modified test the contrary assertion is true. Hence, in applications where one of these conditions is supposed to be true, the respective test can be chosen to let the error probabilities of the second kind at the more likely alternatives be preferably small. Here, α_2 and α_3 model (possibly increasing) load on remaining components of a system, where, by assumption, $\alpha_1 = 1$. Thus, in the actual case, $(\alpha_2, \alpha_3)'$ with both α's greater than 1 are the alternatives we are interested in, and Wald's modified test is adequate. If from the context of the experiment no alternatives are more likely to be true, the LR test seems to be a proper choice since it performs uniformly well over all alternatives (see Figures 5.12-5.14). However, it does not have best (but also not worst) power at most alternatives when the three tests are compared.

Figure 5.12: Cross sections at $\alpha_2 \in \{0.6, 0.8, 1, 1.2, 1.4\}$ of the empirical power functions of the LR test, Wald's/Rao's score test and Wald's modified test corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s = 10$, where the level α is given by 0.05.

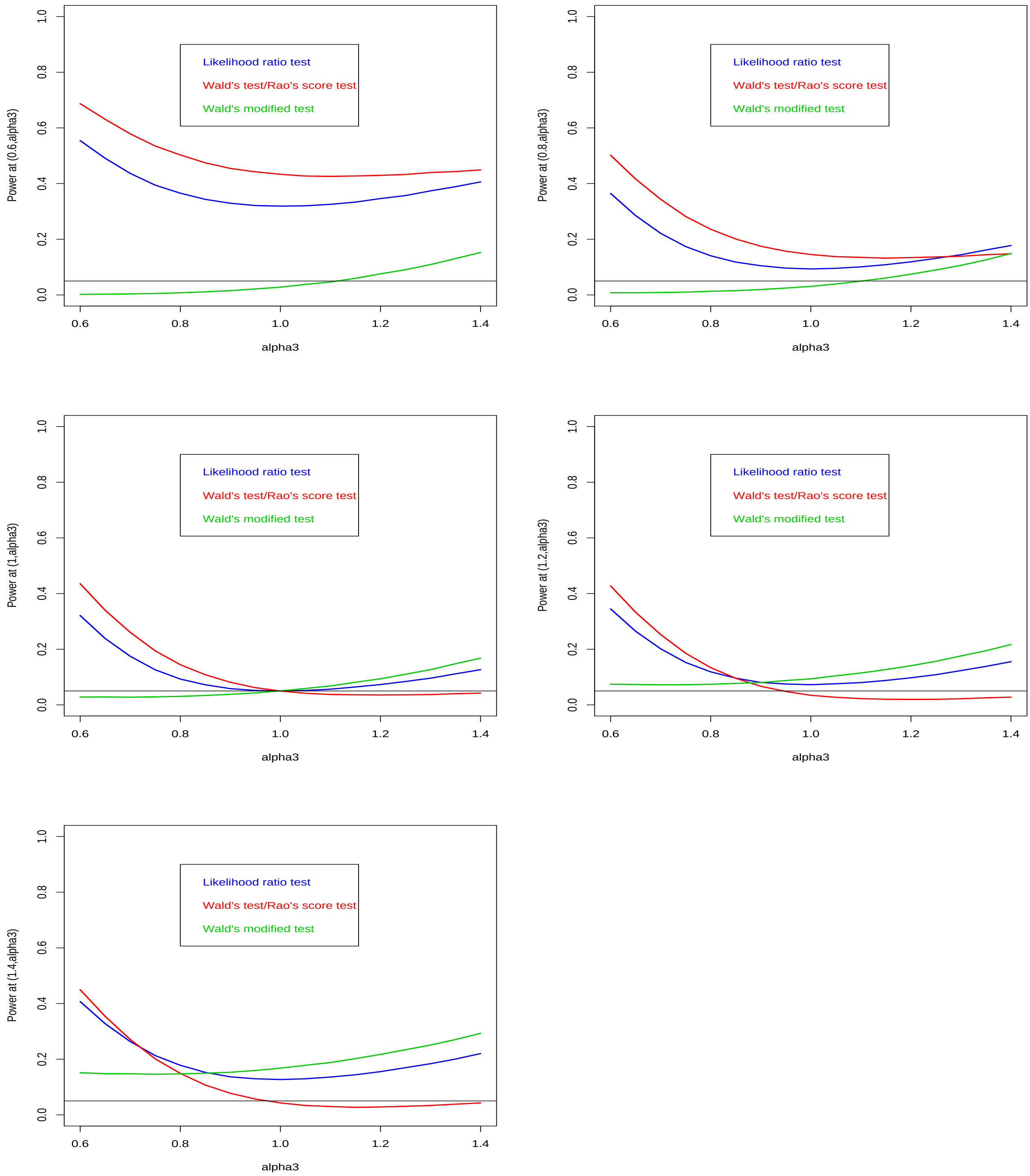


Figure 5.13: Cross sections at $\alpha_2 \in \{0.6, 0.8, 1, 1.2, 1.4\}$ of the empirical power functions of the LR test, Wald's/Rao's score test and Wald's modified test corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s = 50$, where the level α is given by 0.05.

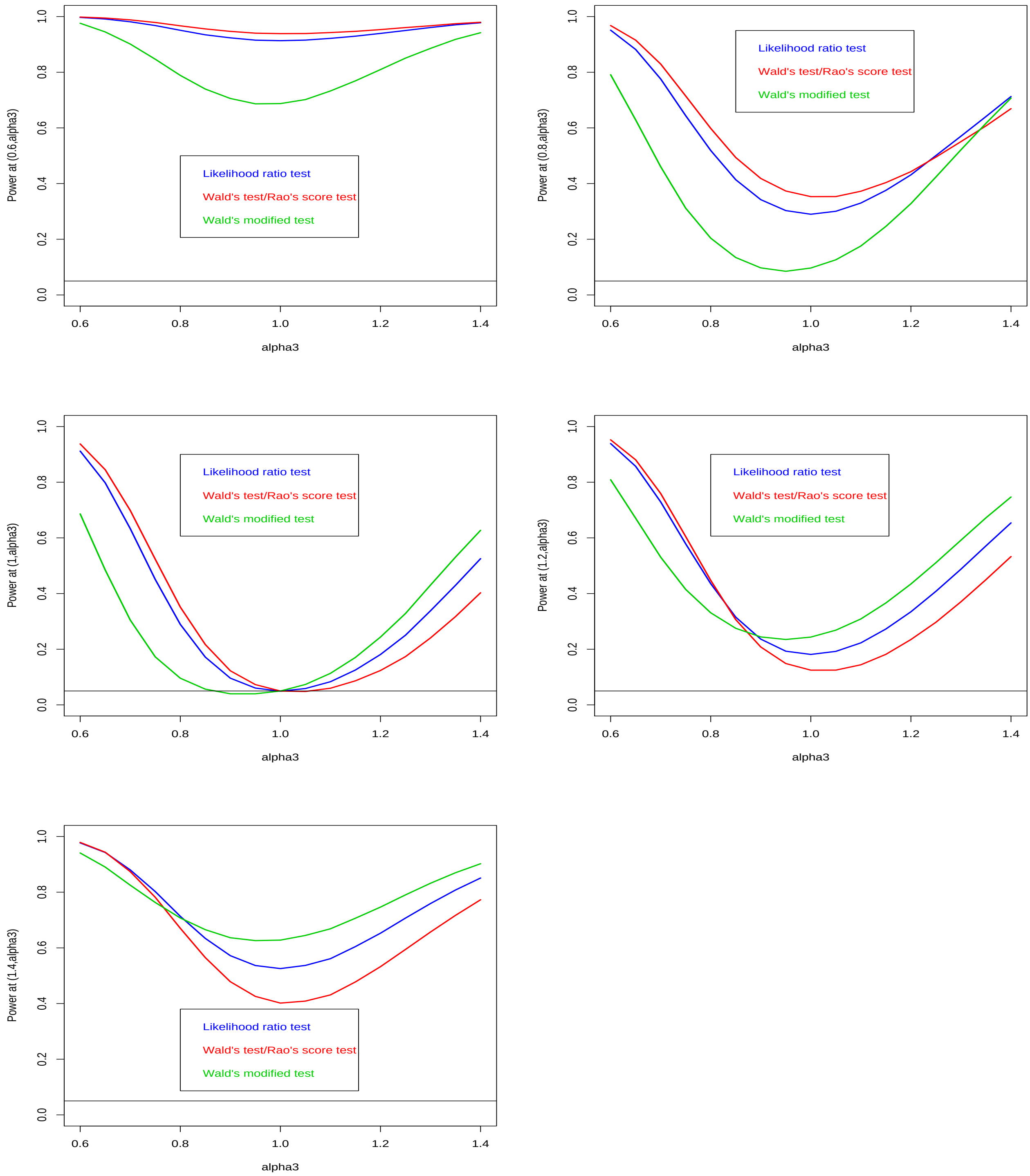
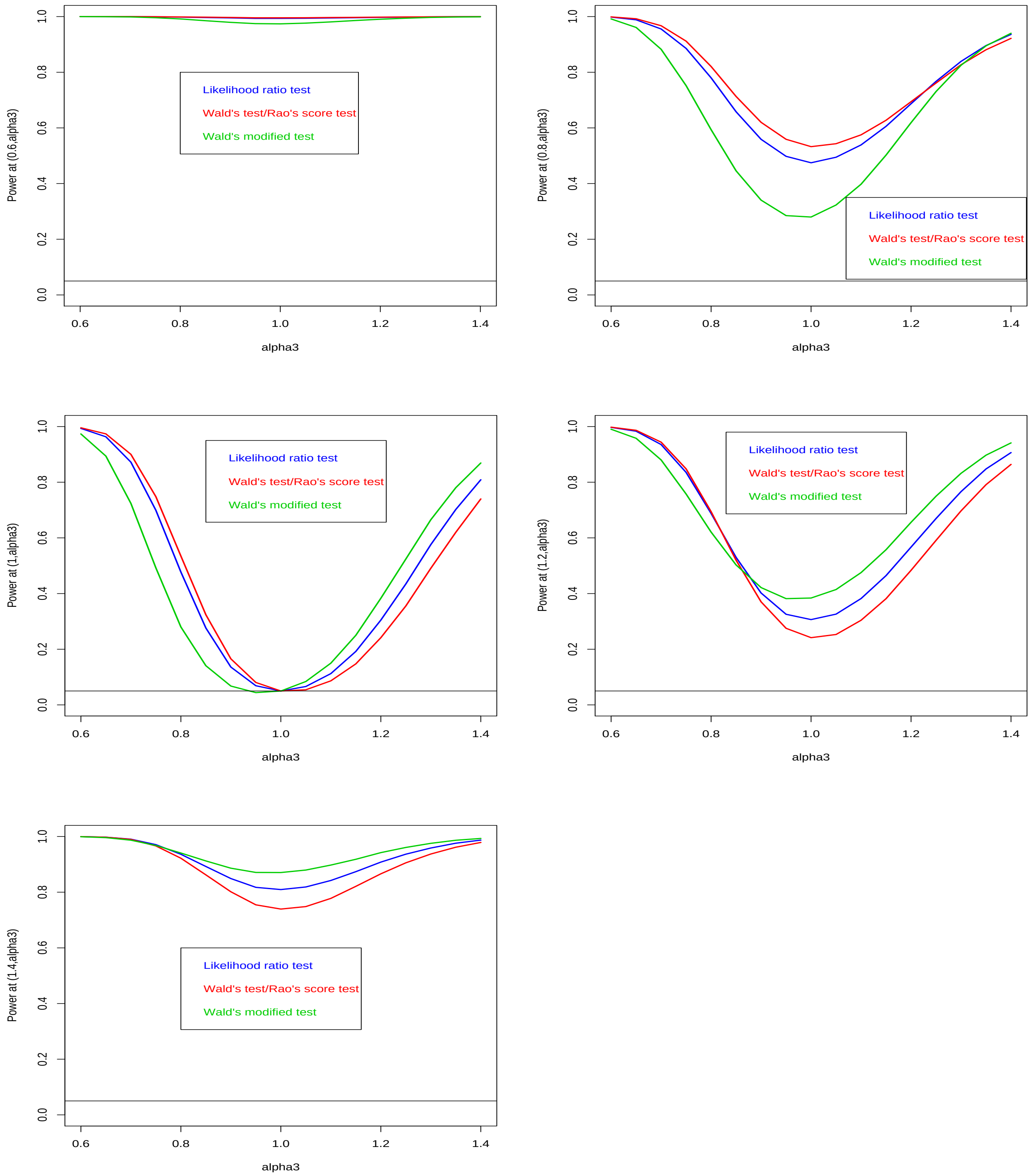


Figure 5.14: Cross sections at $\alpha_2 \in \{0.6, 0.8, 1, 1.2, 1.4\}$ of the empirical power functions of the LR test, Wald's/Rao's score test and Wald's modified test corresponding to test problem (5.2.1) based on 200,000 simulated samples of size $s = 90$, where the level α is given by 0.05.



Chapter 6

Conclusion

In this doctoral thesis, we have pointed out the new finding that the joint distribution of gOSs or, which is of the same structure, of SOSs with conditional proportional hazard rates based on a known underlying baseline distribution function F, forms a multivariate exponential family in the respective model parameters.
Using this structural insight, many results in literature, in particular those related to inferential issues, have readily and briefly been shown in a general and elegant way. This has been demonstrated, e.g., by deriving the MLEs of the model parameters.
Moreover, the structural insight has led to a variety of new useful results that have not been observed by now and that might have been hard or almost impossible to derive without recognizing this structure. Based on well-known results concerning exponential families, optimality properties of estimators on the one hand and statistical tests on the other hand can be shown. This has been demonstrated, once again, in the context of maximum likelihood estimation in terms of efficiency notions and, moreover, in the derivation of UMPU tests on single model parameters.
To sum up, the present thesis has opened the wide and extensively examined field of exponential families to models of ordered random variables, in particular, to SOSs and their applications (e.g. sequential k-out-of-n systems). Having said that this work provides many new statements and several generalizations, it can also be considered the potential basis of further studies. The observed structure gives rise to extended models with helpful properties which provide an even more flexible modeling.

Chapter 7

Tables

Solutions of (3.3.2) and critical values of the UMPU test corresponding to the hypotheses $H_0 : \alpha_j = 1 \leftrightarrow H_1 : \alpha_j \neq 1$ with respect to the sample size s, where $j \in \{1, \ldots, r\}$ is fixed and the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.1: ($\alpha = 0.01$)

s	β^*	$\tau_1^{(0.01,1,s)}(\beta^*)$	$\tau_2^{(0.01,1,s)}(\beta^*)$
1	0.008697	-6.643	-0.009
2	0.00798	-8.451	-0.132
3	0.007547	-10.148	-0.393
4	0.007256	-11.766	-0.749
5	0.007046	-13.327	-1.172
6	0.006884	-14.842	-1.646
7	0.006756	-16.321	-2.158
8	0.00665	-17.77	-2.702
9	0.006561	-19.195	-3.272
10	0.006485	-20.598	-3.864
20	0.006064	-33.897	-10.547
30	0.005873	-46.453	-17.983
40	0.005757	-58.617	-25.816
50	0.005678	-70.527	-33.904

Table 7.2: ($\alpha = 0.05$)

s	β^*	$\tau_1^{(0.05,1,s)}(\beta^*)$	$\tau_2^{(0.05,1,s)}(\beta^*)$
1	0.041479	-4.765	-0.042
2	0.037717	-6.401	-0.304
3	0.03568	-7.948	-0.713
4	0.034376	-9.43	-1.207
5	0.033453	-10.864	-1.758
6	0.032757	-12.262	-2.35
7	0.032208	-13.632	-2.974
8	0.031761	-14.977	-3.623
9	0.031388	-16.304	-4.292
10	0.03107	-17.613	-4.979
20	0.029325	-30.137	-12.439
30	0.02854	-42.089	-20.482
40	0.028069	-53.739	-28.829
50	0.027747	-65.196	-37.372

Table 7.3: ($\alpha = 0.1$)

s	β^*	$\tau_1^{(0.1,1,s)}(\beta^*)$	$\tau_2^{(0.1,1,s)}(\beta^*)$
1	0.080398	-3.932	-0.084
2	0.072964	-5.479	-0.441
3	0.069137	-6.946	-0.937
4	0.066735	-8.355	-1.509
5	0.065053	-9.723	-2.129
6	0.063792	-11.059	-2.785
7	0.062802	-12.371	-3.467
8	0.061998	-13.663	-4.171
9	0.061329	-14.938	-4.893
10	0.06076	-16.199	-5.629
20	0.057648	-28.323	-13.493
30	0.056256	-39.963	-21.849
40	0.055422	-51.349	-30.461
50	0.054852	-62.572	-39.236

Critical values of the LR test statistic $2sT_{LR}^{(s)}$ for test problem $H_0 : \alpha_1 = \cdots = \alpha_r = 1 \leftrightarrow H_1 : \exists j \in \{1, \ldots, r\} : \alpha_j \neq 1$ based on 5,000,000 samples of size s, where the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.4: ($\alpha = 0.01$)

s/r	2	3	4	5
1	10.428	12.853	15.037	17.111
2	9.896	12.198	14.285	16.249
3	9.698	11.924	13.988	15.872
4	9.582	11.787	13.789	15.687
5	9.515	11.715	13.703	15.578
6	9.456	11.638	13.628	15.499
7	9.423	11.612	13.589	15.438
8	9.397	11.597	13.528	15.4
9	9.384	11.536	13.503	15.377
10	9.36	11.513	13.494	15.325
20	9.27	11.436	13.375	15.234
30	9.273	11.406	13.358	15.182
40	9.246	11.403	13.337	15.122
50	9.238	11.378	13.318	15.15

Table 7.5: ($\alpha = 0.05$)

s/r	2	3	4	5
1	6.841	8.927	10.841	12.653
2	6.463	8.432	10.236	11.934
3	6.314	8.234	9.995	11.669
4	6.227	8.132	9.872	11.524
5	6.186	8.068	9.801	11.438
6	6.155	8.023	9.752	11.379
7	6.136	7.99	9.716	11.331
8	6.109	7.985	9.677	11.29
9	6.105	7.959	9.66	11.282
10	6.081	7.945	9.65	11.25
20	6.043	7.879	9.569	11.163
30	6.028	7.855	9.537	11.138
40	6.016	7.85	9.532	11.117
50	6.016	7.839	9.521	11.111

Table 7.6: ($\alpha = 0.1$)

s/r	2	3	4	5
1	5.293	7.176	8.931	10.6
2	4.977	6.757	8.404	9.972
3	4.852	6.591	8.201	9.741
4	4.796	6.514	8.103	9.617
5	4.752	6.459	8.033	9.539
6	4.732	6.425	7.992	9.486
7	4.719	6.402	7.961	9.454
8	4.696	6.384	7.944	9.431
9	4.691	6.37	7.924	9.409
10	4.683	6.355	7.903	9.389
20	4.645	6.299	7.845	9.308
30	4.629	6.29	7.821	9.295
40	4.626	6.282	7.809	9.275
50	4.623	6.274	7.809	9.272

Critical values of Wald's/Rao's score statistic $T_W^{(s)} = T_R^{(s)}$ for test problem $H_0 : \alpha_1 = \cdots = \alpha_r = 1 \leftrightarrow H_1 : \exists j \in \{1, \ldots, r\} : \alpha_j \neq 1$ based on 5,000,000 samples of size s, where the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.7: ($\alpha = 0.01$)

s/r	2	3	4	5
1	19.378	24.14	28.137	31.645
2	15.749	19.506	22.603	25.381
3	14.144	17.436	20.281	22.858
4	13.221	16.31	18.925	21.39
5	12.612	15.597	18.117	20.393
6	12.15	14.987	17.496	19.673
7	11.812	14.582	17.01	19.153
8	11.519	14.264	16.597	18.734
9	11.313	13.956	16.285	18.419
10	11.126	13.765	16.054	18.119
20	10.209	12.626	14.783	16.762
30	9.858	12.193	14.285	16.214
40	9.669	12.002	14.038	15.917
50	9.574	11.842	13.884	15.78

Table 7.8: ($\alpha = 0.05$)

s/r	2	3	4	5
1	7.813	10.949	13.696	16.171
2	7.087	9.735	12.104	14.266
3	6.705	9.189	11.375	13.379
4	6.477	8.841	10.941	12.892
5	6.325	8.62	10.661	12.557
6	6.216	8.464	10.466	12.309
7	6.157	8.336	10.298	12.126
8	6.106	8.26	10.196	11.984
9	6.085	8.195	10.111	11.891
10	6.044	8.142	10.027	11.793
20	5.989	7.942	9.733	11.409
30	5.986	7.893	9.629	11.288
40	5.987	7.875	9.603	11.226
50	5.994	7.853	9.565	11.203

Table 7.9: ($\alpha = 0.1$)

s/r	2	3	4	5
1	4.401	6.772	8.966	11.001
2	4.313	6.475	8.428	10.253
3	4.24	6.299	8.164	9.916
4	4.261	6.211	8.013	9.708
5	4.279	6.166	7.922	9.575
6	4.316	6.144	7.863	9.493
7	4.347	6.129	7.831	9.426
8	4.364	6.141	7.802	9.395
9	4.389	6.139	7.791	9.368
10	4.41	6.144	7.776	9.335
20	4.502	6.184	7.754	9.26
30	4.532	6.205	7.756	9.256
40	4.553	6.214	7.766	9.243
50	4.562	6.225	7.769	9.247

Critical values of Wald's modified statistic $T_{\tilde{W}}^{(s)}$ for test problem $H_0 : \alpha_1 = \cdots = \alpha_r = 1 \leftrightarrow H_1 : \exists j \in \{1, \dots, r\} : \alpha_j \neq 1$ based on 5,000,000 samples of size s, where the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.10: ($\alpha = 0.01$)

s/r	2	3	4	5
1	39681	89125.1	159013.2	246122.1
2	678.91	1054.34	1453.87	1852.33
3	190.94	267.61	339.49	407.32
4	101.4	135.47	165.63	193.88
5	68.748	89.852	108.16	125.27
6	52.724	67.753	81.086	92.637
7	43.584	55.627	65.71	74.843
8	37.582	47.489	56.175	63.781
9	33.451	42.008	49.253	55.927
10	30.266	37.916	44.551	50.107
20	18.737	23.133	26.928	30.359
30	15.518	19.127	22.242	25.077
40	13.975	17.198	20.001	22.553
50	13.025	16.069	18.672	21.101

Table 7.11: ($\alpha = 0.05$)

s/r	2	3	4	5
1	1456.5	3391.4	6145.3	9672.2
2	107.86	180.09	255.42	331.31
3	46.331	70.295	93.698	116.37
4	30.005	44.115	56.862	69.125
5	23.075	32.856	41.975	50.441
6	19.147	26.932	33.988	40.644
7	16.681	23.252	29.244	34.668
8	14.983	20.865	25.923	30.761
9	13.797	18.996	23.64	27.967
10	12.842	17.664	21.946	25.842
20	9.093	12.275	15.127	17.733
30	7.941	10.678	13.126	15.391
40	7.395	9.909	12.163	14.264
50	7.074	9.454	11.597	13.592

Table 7.12: ($\alpha = 0.1$)

s/r	2	3	4	5
1	333.41	792.45	1452.1	2302.3
2	43.934	77.28	113.33	151.52
3	22.361	36.233	50.028	63.893
4	15.774	24.65	33.13	41.388
5	12.651	19.342	25.638	31.766
6	10.897	16.415	21.561	26.479
7	9.769	14.538	18.968	23.208
8	8.918	13.246	17.223	20.974
9	8.324	12.27	15.924	19.347
10	7.876	11.546	14.922	18.11
20	5.951	8.552	10.957	13.199
30	5.403	7.686	9.782	11.761
40	5.172	7.268	9.216	11.073
50	5.038	7.046	8.898	10.669

Critical values of the LR test statistic $2sT_{LR}^{(s)}$ for test problem $H_0 : \alpha_1 = \cdots = \alpha_r \leftrightarrow H_1 : \exists j, k \in \{1, \ldots, r\} : \alpha_j \neq \alpha_k$ based on 5,000,000 samples of size s, where the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.13: ($\alpha = 0.01$)

s/r	2	3	4	5
1	7.821	10.746	13.133	15.343
2	7.357	10.121	12.397	14.477
3	7.146	9.82	12.088	14.105
4	7.014	9.697	11.909	13.917
5	6.956	9.605	11.793	13.783
6	6.903	9.547	11.737	13.722
7	6.874	9.486	11.673	13.652
8	6.827	9.455	11.634	13.603
9	6.812	9.427	11.614	13.57
10	6.789	9.398	11.58	13.533
20	6.718	9.323	11.467	13.413
30	6.694	9.256	11.428	13.369
40	6.685	9.268	11.411	13.342
50	6.665	9.266	11.412	13.336

Table 7.14: ($\alpha = 0.05$)

s/r	2	3	4	5
1	4.654	7.11	9.183	11.098
2	4.303	6.624	8.581	10.376
3	4.15	6.421	8.335	10.101
4	4.078	6.316	8.205	9.956
5	4.032	6.255	8.139	9.859
6	3.993	6.205	8.088	9.799
7	3.976	6.177	8.037	9.755
8	3.955	6.153	8.023	9.719
9	3.949	6.132	7.991	9.693
10	3.934	6.121	7.969	9.677
20	3.892	6.057	7.897	9.582
30	3.876	6.035	7.872	9.555
40	3.866	6.021	7.862	9.539
50	3.861	6.013	7.847	9.527

Table 7.15: ($\alpha = 0.1$)

s/r	2	3	4	5
1	3.321	5.514	7.397	9.146
2	3.039	5.098	6.876	8.53
3	2.929	4.938	6.676	8.282
4	2.874	4.854	6.572	8.162
5	2.846	4.809	6.509	8.09
6	2.818	4.77	6.464	8.035
7	2.802	4.751	6.437	8.004
8	2.789	4.732	6.411	7.965
9	2.78	4.718	6.396	7.953
10	2.773	4.704	6.383	7.936
20	2.743	4.659	6.316	7.858
30	2.728	4.643	6.297	7.83
40	2.722	4.632	6.282	7.823
50	2.721	4.622	6.276	7.816

Critical values of Rao's score statistic $T_R^{(s)}$ for test problem $H_0 : \alpha_1 = \cdots = \alpha_r \leftrightarrow H_1 : \exists j, k \in \{1, \ldots, r\} : \alpha_j \neq \alpha_k$ based on 5,000,000 samples of size s, where the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.16: ($\alpha = 0.01$)

s/r	2	3	4	5
1	1.96	5.011	8.095	11
2	3.364	6.816	9.727	12.359
3	4.176	7.499	10.263	12.717
4	4.671	7.881	10.523	12.926
5	5.012	8.094	10.679	13.021
6	5.249	8.246	10.803	13.065
7	5.432	8.348	10.871	13.106
8	5.558	8.441	10.915	13.125
9	5.671	8.524	10.964	13.146
10	5.756	8.567	11.007	13.161
20	6.184	8.883	11.16	13.213
30	6.334	8.967	11.223	13.244
40	6.413	9.04	11.254	13.237
50	6.448	9.072	11.287	13.255

Table 7.17: ($\alpha = 0.05$)

s/r	2	3	4	5
1	1.805	3.93	5.872	7.736
2	2.636	4.722	6.626	8.479
3	2.996	5.022	6.954	8.767
4	3.195	5.202	7.121	8.918
5	3.318	5.334	7.246	9.003
6	3.397	5.43	7.322	9.082
7	3.462	5.499	7.377	9.134
8	3.504	5.559	7.432	9.166
9	3.546	5.601	7.466	9.202
10	3.572	5.644	7.494	9.228
20	3.709	5.812	7.651	9.346
30	3.754	5.868	7.703	9.399
40	3.774	5.898	7.731	9.423
50	3.788	5.912	7.747	9.43

Table 7.18: ($\alpha = 0.1$)

s/r	2	3	4	5
1	1.62	3.218	4.718	6.251
2	2.129	3.698	5.308	6.87
3	2.317	3.924	5.561	7.112
4	2.414	4.08	5.701	7.251
5	2.477	4.177	5.793	7.341
6	2.512	4.246	5.859	7.403
7	2.539	4.301	5.911	7.451
8	2.56	4.335	5.948	7.484
9	2.576	4.365	5.982	7.52
10	2.589	4.386	6.007	7.54
20	2.651	4.498	6.126	7.655
30	2.667	4.535	6.17	7.693
40	2.676	4.553	6.186	7.72
50	2.685	4.558	6.197	7.733

Critical values of Wald's statistic $T_W^{(s)}$ for test problem $H_0 : \alpha_1 = \cdots = \alpha_r \leftrightarrow H_1 : \exists j, k \in \{1, \ldots, r\} : \alpha_j \neq \alpha_k$ based on 5,000,000 samples of size s, where the level is given by $\alpha \in \{0.01, 0.05, 0.1\}$.

Table 7.19: ($\alpha = 0.01$)

s/r	2	3	4	5
1	0.99	1.876	2.677	3.437
2	1.827	3.191	4.39	5.528
3	2.462	4.091	5.532	6.882
4	2.949	4.757	6.348	7.852
5	3.339	5.257	6.962	8.562
6	3.652	5.655	7.45	9.101
7	3.913	5.973	7.831	9.535
8	4.125	6.244	8.139	9.882
9	4.313	6.479	8.404	10.174
10	4.47	6.664	8.632	10.419
20	5.356	7.737	9.794	11.671
30	5.73	8.154	10.263	12.17
40	5.937	8.406	10.514	12.415
50	6.057	8.555	10.684	12.588

Table 7.20: ($\alpha = 0.05$)

s/r	2	3	4	5
1	0.949	1.701	2.379	3.037
2	1.589	2.642	3.624	4.588
3	1.998	3.223	4.403	5.533
4	2.283	3.629	4.928	6.168
5	2.491	3.935	5.319	6.619
6	2.647	4.172	5.61	6.971
7	2.775	4.358	5.839	7.244
8	2.874	4.514	6.031	7.457
9	2.962	4.639	6.183	7.64
10	3.03	4.75	6.311	7.791
20	3.394	5.299	6.983	8.547
30	3.533	5.509	7.239	8.845
40	3.604	5.622	7.375	8.999
50	3.649	5.688	7.458	9.087

Table 7.21: ($\alpha = 0.1$)

s/r	2	3	4	5
1	0.895	1.552	2.165	2.778
2	1.389	2.288	3.191	4.072
3	1.672	2.733	3.8	4.825
4	1.855	3.045	4.204	5.321
5	1.985	3.267	4.492	5.675
6	2.077	3.435	4.71	5.938
7	2.149	3.57	4.881	6.144
8	2.207	3.672	5.016	6.305
9	2.253	3.757	5.13	6.443
10	2.292	3.826	5.223	6.552
20	2.486	4.185	5.69	7.111
30	2.554	4.317	5.868	7.317
40	2.59	4.387	5.955	7.433
50	2.614	4.424	6.011	7.501

Chapter 8

Appendix

8.1 Proofs of several Theorems of Chapter 2

Proof of La. 2.1.18. Let $\boldsymbol{t} \in V$, then

$$\begin{aligned} m_{\boldsymbol{T}}(\boldsymbol{t}) &= E_{\boldsymbol{\zeta}}[\exp\{\boldsymbol{t}'\boldsymbol{T}\}] = \int \exp\{\boldsymbol{t}'\boldsymbol{T}(x)\} f_{\boldsymbol{\zeta}}(x) d\mu(x) \\ &= \frac{C(\boldsymbol{\zeta})}{C(\boldsymbol{\zeta}+\boldsymbol{t})} \int C(\boldsymbol{\zeta}+\boldsymbol{t}) \exp\{(\boldsymbol{\zeta}+\boldsymbol{t})'\boldsymbol{T}(x)\} h(x) d\mu(x) \\ &= \frac{C(\boldsymbol{\zeta})}{C(\boldsymbol{\zeta}+\boldsymbol{t})}. \end{aligned}$$

■

Proof of La. 2.1.30. Let $\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)} \in \Theta$, then

$$\begin{aligned} d_{KL}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}) &= \int \left(\ln f_{\boldsymbol{\zeta}^{(1)}}(x) - \ln f_{\boldsymbol{\zeta}^{(2)}}(x)\right) f_{\boldsymbol{\zeta}^{(1)}}(x) d\mu(x) \\ &= \int \left((\boldsymbol{\zeta}^{(1)} - \boldsymbol{\zeta}^{(2)})'\boldsymbol{T}(x) - \kappa(\boldsymbol{\zeta}^{(1)}) + \kappa(\boldsymbol{\zeta}^{(2)})\right) f_{\boldsymbol{\zeta}^{(1)}}(x) d\mu(x) \\ &= \kappa(\boldsymbol{\zeta}^{(2)}) - \kappa(\boldsymbol{\zeta}^{(1)}) + (\boldsymbol{\zeta}^{(1)} - \boldsymbol{\zeta}^{(2)})' \int \boldsymbol{T}(x) f_{\boldsymbol{\zeta}^{(1)}}(x) d\mu(x) \\ &= \kappa(\boldsymbol{\zeta}^{(2)}) - \kappa(\boldsymbol{\zeta}^{(1)}) + (\boldsymbol{\zeta}^{(1)} - \boldsymbol{\zeta}^{(2)})' \pi(\boldsymbol{\zeta}^{(1)}). \end{aligned}$$

■

Proof of La. 2.2.1. Denoting by l_s the log likelihood function based on the observations $x^{(1)}, \ldots, x^{(s)}$, we obtain with (2.2.1)

$$\begin{aligned} \nabla l_s(\boldsymbol{\zeta}) &= \nabla \left[\ln(f_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}))\right] \\ &= \boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - s\nabla\kappa(\boldsymbol{\zeta}) \\ &= s\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) - \pi(\boldsymbol{\zeta})\right), \quad \boldsymbol{\zeta} \in \Theta. \end{aligned}$$

By assumption, $\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)}) \in \pi(\Theta)$, and, hence,

$$\pi^{-1}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{x}}^{(s)})\right)$$

is the unique solution of the likelihood equation. Moreover, (2.1.15) and Thm. 2.1.9 yield that the Hessian matrix of l_s equals $-s\,\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T})$ which is negative definite for all $\boldsymbol{\zeta} \in \Theta$. It follows that l_s is a strictly concave function on Θ and, thus, the proof is completed. ■

Proof of Thm. 2.2.3. The first part of the statement is an immediate consequence of La. 2.2.1. Moreover, for fixed $\boldsymbol{\zeta} \in \Theta$, all moments of $\boldsymbol{T}$ exist (cf. Thm. 2.1.15 (i)), and the strong law of large numbers (e.g, in Shao (2003), Thm. 1.13. (ii), p. 62) yields

$$\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = \frac{1}{s}\sum_{i=1}^{s}\boldsymbol{T}(X^{(i)}) \stackrel{s\to\infty}{\longrightarrow} E_{\boldsymbol{\zeta}}[\boldsymbol{T}(X)] = \pi(\boldsymbol{\zeta}) \quad [P_{\boldsymbol{\zeta}}],$$

which implies that $\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \to \pi(\boldsymbol{\zeta})$ in $P_{\boldsymbol{\zeta}}$-probability (see, e.g., Shao (2003), Thm. 1.8. (i), p. 51). $\pi(\Theta)$ is as pre-image of Θ of the continuous mapping π^{-1} an open subset of $\mathbb{R}^k$. Hence, there exists a positive number $\delta > 0$ with $\boldsymbol{U}_\delta(\pi(\boldsymbol{\zeta})) = \{\boldsymbol{\eta} \in \mathbb{R}^k : ||\boldsymbol{\eta} - \pi(\boldsymbol{\zeta})|| < \delta\} \subseteq \pi(\Theta)$. Clearly,

$$P_{\boldsymbol{\zeta}}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \in \pi(\Theta)\right) \geq P_{\boldsymbol{\zeta}}\left(\frac{1}{s}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \in \boldsymbol{U}_\delta(\pi(\boldsymbol{\zeta}))\right) \stackrel{s\to\infty}{\longrightarrow} 1,$$

and, thus, the assertion is established. ■

Proof of La. 2.2.10. Upon choosing g as $g(\boldsymbol{\zeta}) = E_{\boldsymbol{\zeta}}[\boldsymbol{T}] = \pi(\zeta)$, $\boldsymbol{\zeta} \in \Theta$, we obtain, from (2.1.16) and (2.1.22),

$$\begin{aligned}\mathbf{D}_g(\boldsymbol{\zeta})\mathbf{I}_f^{(s)}(\boldsymbol{\zeta})^{-1}\mathbf{D}_g(\boldsymbol{\zeta})' &= \mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T})(s\,\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}))^{-1}\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}) \\ &= \frac{1}{s}\mathbf{Cov}_{\boldsymbol{\zeta}}(\boldsymbol{T}) = \mathbf{Cov}_{\boldsymbol{\zeta}}\left(\frac{1}{s}\boldsymbol{T}^{(s)}\right).\end{aligned}$$

Clearly, $E_{\boldsymbol{\zeta}}[s^{-1}\boldsymbol{T}^{(s)}] = \pi(\boldsymbol{\zeta})$ and this completes the proof. ■

Proof of La. 2.2.15. At first, we show the statement for the sequence $\{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ of MLEs of $\boldsymbol{\zeta}$. In order to apply Thm. 5.1 in Lehmann & Casella (1998), p. 463, we verify the conditions (A)-(D) given, where $\boldsymbol{\zeta}^{(0)} \in \Theta$ is assumed to be the true parameter vector. The conditions (A0)-(A2) (Lehmann & Casella (1998), pp. 443/444) are, by assumption and Rem. 2.1.4, obvious.

(A) Since Θ is assumed to be open, there exists an open neighbourhood $U_\delta(\boldsymbol{\zeta}^{(0)})$, $\delta = \delta(\boldsymbol{\zeta}^{(0)}) > 0$, of $\boldsymbol{\zeta}^{(0)}$ with the property that its closure lies in Θ, i.e. $cl(U_\delta(\boldsymbol{\zeta}^{(0)})) \subseteq \Theta$. From Thm. 2.1.15 (i) follows that C is infinitely often differentiable in $\boldsymbol{\zeta} \in U_\delta(\boldsymbol{\zeta}^{(0)})$, and so is $f_{\boldsymbol{\zeta}}^X$.

(B) From Thm. 2.1.22, the first part of the condition is obvious. Moreover, from Thm. 2.1.15 (i) and Thm. 2.1.22, we obtain, for $1 \leq i, j \leq k$,

$$\begin{aligned} E_{\boldsymbol{\zeta}}\left[-\frac{d^2}{d\zeta_i d\zeta_j}\ln(f_{\boldsymbol{\zeta}}^X)\right] &= E_{\boldsymbol{\zeta}}\left[-\frac{d^2}{d\zeta_i d\zeta_j}(\boldsymbol{\zeta}'\boldsymbol{T} - \kappa(\boldsymbol{\zeta}) + \ln(h))\right] \\ &= [\mathbf{H}_\kappa(\boldsymbol{\zeta})]_{i,j} \\ &= Cov_{\boldsymbol{\zeta}}(T_i, T_j) \\ &= [\mathbf{I}_f(\boldsymbol{\zeta})]_{i,j}. \end{aligned}$$

(C) By assumption, $\mathfrak{P}^X$ is strictly k-parametrical, and (2.1.19) is fulfilled for every $\boldsymbol{\zeta} \in U_\delta(\boldsymbol{\zeta}^{(0)})$.

(D) For $1 \leq i, j.l \leq k$, the third partial derivatives

$$\frac{d^3}{d\zeta_i d\zeta_j d\zeta_l}\ln(f_{\boldsymbol{\zeta}}^X) = -\frac{d^3}{d\zeta_i d\zeta_j d\zeta_l}\kappa(\boldsymbol{\zeta})$$

do not depend on x. Moreover, κ is infinitely often differentiable and, thus, in particular, for all $\boldsymbol{\zeta} \in U_\delta(\boldsymbol{\zeta}^{(0)})$,

$$\left|\frac{d^3}{d\zeta_i d\zeta_j d\zeta_l}\ln(f_{\boldsymbol{\zeta}}^X(x))\right| \leq \max_{\boldsymbol{\zeta} \in cl(U_\delta(\boldsymbol{\zeta}^{(0)}))}\left|\frac{d^3}{d\zeta_i d\zeta_j d\zeta_l}\kappa(\boldsymbol{\zeta})\right| < \infty, \quad \forall x \in \mathfrak{X},$$

where the upper bounds are $P^X_{\boldsymbol{\zeta}^{(0)}}$-integrable considered as constant functions on $(\mathfrak{X}, \mathfrak{B})$.

Now, by application of the theorem, with $P_{\boldsymbol{\zeta}^{(0)}}$-probability tending to 1 as s tends to infinity, there exists a sequence of solutions of the likelihood equations that is asymptotically efficient. By assumption, the sequence of MLEs of $\boldsymbol{\zeta}$ exists, and it is the unique solution of these equations. Hence, asymptotic efficiency of $\{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ is shown.
We continue by application of the mean value theorem to the real-valued component functions g_j, $1 \leq j \leq k$, yields

$$\sqrt{s}(\boldsymbol{\gamma}^{*(s)} - \boldsymbol{\gamma}) = \sqrt{s}(g(\boldsymbol{\zeta}^{*(s)}) - g(\boldsymbol{\zeta})) = \mathbf{D}_g[\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}]\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}),$$

where $\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}$ are points on the line between $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^{*(s)}$, and $\mathbf{D}_g[\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}]$ is a matrix with rows $\nabla g_j(\boldsymbol{\delta_j}^{(s)})'$, $1 \leq j \leq k$. By assumption, all partial derivatives of g are continuous and, hence, using strong consistency of the sequence $\{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ (cf. La. 2.2.12), we derive $\mathbf{D}_g[\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}] \to \mathbf{D}_g(\boldsymbol{\zeta})$ $P_{\boldsymbol{\zeta}}$-a.s.. We have shown that $\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}) \xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \mathbf{I}_f(\boldsymbol{\zeta})^{-1})$ and, thus, the multivariate version of Slutsky's theorem (see, e.g., Sen & Singer (1993), Thm. 3.4.3, p. 130) yields

$$\sqrt{s}(\boldsymbol{\gamma}^{*(s)} - \boldsymbol{\gamma}) \xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \mathbf{D}_g(\boldsymbol{\zeta})\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\mathbf{D}_g(\boldsymbol{\zeta})').$$

From (2.1.20), by setting $h = g$, the asymptotic covariance matrix is exactly the inverse of the Fisher information matrix of $\tilde{\mathfrak{P}}^{\boldsymbol{X}}$ at $\gamma \in \Gamma$, and this completes the proof. ■

Proof Δ_1. From Thm. 2.1.22, Thm. 2.2.3 and the multivariate mean value theorem, it follows that, for $\boldsymbol{\zeta} \in \Theta$,

$$\begin{aligned}
&\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}) - \frac{1}{\sqrt{s}}\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \\
= \;& \sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}) - \frac{1}{\sqrt{s}}\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\left(\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - E_{\boldsymbol{\zeta}}[\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})]\right) \\
= \;& \sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}) - \sqrt{s}\,\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\left(\pi(\boldsymbol{\zeta}^{*(s)}) - \pi(\boldsymbol{\zeta})\right) \\
= \;& \left(\mathbf{I}_k - \mathbf{I}_f(\boldsymbol{\zeta})^{-1}\mathbf{D}_\pi\left[\boldsymbol{\delta}_1^{(s)}, \ldots, \boldsymbol{\delta}_k^{(s)}\right]\right)\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}),
\end{aligned}$$

where $\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}$ are points on the line between $\boldsymbol{\zeta}^{*(s)}$ and $\boldsymbol{\zeta}$, and $\mathbf{D}_\pi[\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}]$ is a matrix with lines $\nabla\pi_i(\boldsymbol{\delta}_i^{(s)})'$, $1 \leq i \leq k$. Since all partial derivatives of π are continuous, strong consistency of $\{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ implies $\mathbf{D}_\pi[\boldsymbol{\delta}_1^{(s)}, ..., \boldsymbol{\delta}_k^{(s)}] \to \mathbf{D}_\pi(\boldsymbol{\zeta}) = \mathbf{I}_f(\boldsymbol{\zeta})$ $P_{\boldsymbol{\zeta}}$-a.s.. Hence, $\mathbf{I}_k - \mathbf{I}_f(\boldsymbol{\zeta})^{-1}\mathbf{D}_\pi\left[\boldsymbol{\delta}_1^{(s)}, \ldots, \boldsymbol{\delta}_k^{(s)}\right]$ converges to $\mathbf{0} \in \mathbb{R}^{k\times k}$ $P_{\boldsymbol{\zeta}}$-a.s., and, in combination with the asymptotic efficiency of $\{\boldsymbol{\zeta}^{*(s)}\}_{s\in\mathbb{N}}$ (cf. La. 2.2.15), the multivariate version of Slutsky's theorem (e.g., in Sen & Singer (1993), Thm. 3.4.3, p. 130) yields

$$\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}) - \frac{1}{\sqrt{s}}\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \xrightarrow{\mathcal{D}} \mathbf{0}.$$

Application of Thm. 1.8 (vii) in Shao (2003), p. 51, leads to

$$\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}) - \frac{1}{\sqrt{s}}\mathbf{I}_f(\boldsymbol{\zeta})^{-1}\boldsymbol{U}_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) \xrightarrow{P_{\boldsymbol{\zeta}}} \mathbf{0}, \quad \boldsymbol{\zeta} \in \Theta.$$

■

Proof of Thm. 2.3.2. Set $\eta_0 = -\zeta_0$, $\eta_j = -\zeta_j$ and $S_j = -T_j$, $1 \leq j \leq k$. Moreover, define $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)'$, $\boldsymbol{S} = (S_1, \ldots, S_k)'$ and $\tilde{f}_{\boldsymbol{\eta}}^X = f_{-\boldsymbol{\eta}}^X$, $\boldsymbol{\eta} \in -\Theta$. Then, $\mathfrak{P} = \{\tilde{P}_{\boldsymbol{\eta}}^X = \tilde{f}_{\boldsymbol{\eta}}^X\mu : \boldsymbol{\eta} \in -\Theta\}$ forms a strictly k-parametrical exponential family in $\eta_1, \ldots, \eta_k$ and $S_1, \ldots, S_k$, and test problem (II) is equivalent to the test problem $H_0 : \eta_1 \leq \eta_0$ against $H_1 : \eta_1 > \eta_0$. With the denotations $\tilde{\boldsymbol{S}} = (S_2, \ldots, S_k)'$ and $\tilde{\boldsymbol{s}} = (s_2, \ldots, s_k)'$, application of Thm. 2.3.1 yields that $\varphi^* = \tilde{\Psi}^* \circ (S_1, \tilde{\boldsymbol{S}})$ defined by

$$\tilde{\Psi}^*(s_1, \tilde{\boldsymbol{s}}) = \mathbb{1}_{(\tilde{c}(\tilde{\boldsymbol{s}}),\infty)}(s_1) + \tilde{\gamma}(\tilde{\boldsymbol{s}})\mathbb{1}_{\{\tilde{c}(\tilde{\boldsymbol{s}})\}}(s_1)$$

is an UMPU level-α test for that test problem, where $\tilde{c}, \tilde{\gamma} :\ (\mathbb{R}^{k-1}, \mathbb{B}^{k-1}) \to (\mathbb{R}^1, \mathbb{B}^1)$ with $0 \leq \tilde{\gamma} \leq 1$ fulfill

$$\tilde{P}_{\eta_0,\bullet}^{S_1|\tilde{\boldsymbol{S}}=\tilde{\boldsymbol{s}}}((\tilde{c}(\tilde{\boldsymbol{s}}), \infty)) + \tilde{\gamma}(\tilde{\boldsymbol{s}})\tilde{P}_{\eta_0,\bullet}^{S_1|\tilde{\boldsymbol{S}}=\tilde{\boldsymbol{s}}}(\{\tilde{c}(\tilde{\boldsymbol{s}})\}) = \alpha, \quad \tilde{\boldsymbol{s}} \in \mathbb{R}^{k-1}. \tag{8.1.1}$$

Defining $c(\tilde{\boldsymbol{t}}) = -\tilde{c}(-\tilde{\boldsymbol{t}})$ and $\gamma(\tilde{\boldsymbol{t}}) = \tilde{\gamma}(-\tilde{\boldsymbol{t}})$, $\tilde{\boldsymbol{t}} \in \mathbb{R}^{k-1}$, condition (8.1.1) is equivalent to condition (2.3.2) and, by setting Ψ^* as in (2.3.1), it follows that $\Psi^*(t_1, \tilde{\boldsymbol{t}}) = \tilde{\Psi}^*(-t_1, -\tilde{\boldsymbol{t}})$ and, thus,

$\varphi^* = \Psi^* \circ (T_1, \tilde{\boldsymbol{T}})$, which completes the proof. ∎

Proof of La. 2.3.8. According to Thm. 2.2.3, the MLE of $\boldsymbol{\zeta}$ based on $\tilde{\boldsymbol{X}}^{(s)}$ is given by $\boldsymbol{\zeta}^{*(s)} = \pi^{-1}(s^{-1}\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}))$. Plugging in leads to

$$\begin{aligned} T_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) &= -s^{-1}\ln\left(\frac{\sup_{\boldsymbol{\zeta}^{(0)}\in\Theta_0}\exp\{(\boldsymbol{\zeta}^{(0)})'\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - s\kappa(\boldsymbol{\zeta}^{(0)})\}}{\exp\{(\boldsymbol{\zeta}^{*(s)})'\boldsymbol{T}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) - s\kappa(\boldsymbol{\zeta}^{*(s)})\}}\right) \\ &= \ln\left(\frac{\exp\{(\boldsymbol{\zeta}^{*(s)})'\pi(\boldsymbol{\zeta}^{*(s)}) - \kappa(\boldsymbol{\zeta}^{*(s)})\}}{\exp\{\sup_{\boldsymbol{\zeta}^{(0)}\in\Theta_0}\{(\boldsymbol{\zeta}^{(0)})'\pi(\boldsymbol{\zeta}^{*(s)}) - \kappa(\boldsymbol{\zeta}^{(0)})\}\}}\right) \\ &= \inf_{\boldsymbol{\zeta}^{(0)}\in\Theta_0}\{\kappa(\boldsymbol{\zeta}^{(0)}) - \kappa(\boldsymbol{\zeta}^{*(s)}) + (\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)})'\pi(\boldsymbol{\zeta}^{*(s)})\} \\ &= \inf_{\boldsymbol{\zeta}^{(0)}\in\Theta_0} d_{KL}(\boldsymbol{\zeta}^{*(s)}, \boldsymbol{\zeta}^{(0)}), \end{aligned}$$

where the last equality follows from La. 2.1.30. ∎

Proof of Thm. 2.3.9 (i)**.** Let $\boldsymbol{\zeta}^{(0)}$ be the true parameter. Since $\nabla l_s(\boldsymbol{\zeta}^{*(s)}) = 0$, $s \in \mathbb{N}$, the Taylor expansion of the second order around $\boldsymbol{\zeta}^{*(s)}$ at the point $\boldsymbol{\zeta}^{(0)}$ (see, e.g., Heuser (2000), Thm. 168.1, p. 282) of the log likelihood statistic $l_s(\boldsymbol{\zeta}) = l_s(\boldsymbol{\zeta}; \tilde{\boldsymbol{X}}^{(s)}) = \ln f_{\boldsymbol{\zeta}}^{(s)}(\tilde{\boldsymbol{X}}^{(s)})$, $s \in \mathbb{N}$, is given by

$$l_s(\boldsymbol{\zeta}^{(0)}) = l_s(\boldsymbol{\zeta}^{*(s)}) + \frac{1}{2}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)})'\mathbf{H}_{l_s}(\tilde{\boldsymbol{\zeta}}^{(s)})(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)}),$$

where $\{\tilde{\boldsymbol{\zeta}}^{(s)}\}_{s\in\mathbb{N}}$ is a sequence of random vectors with $\tilde{\boldsymbol{\zeta}}^{(s)}$ on the line between $\boldsymbol{\zeta}^{(0)}$ and $\boldsymbol{\zeta}^{*(s)}$, $s \in \mathbb{N}$. Then, since $-sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = l_s(\boldsymbol{\zeta}^{(0)}) - l_s(\boldsymbol{\zeta}^{*(s)})$,

$$2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = \sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)})'[-s^{-1}\mathbf{H}_{l_s}(\tilde{\boldsymbol{\zeta}}^{(s)})]\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)}).$$

Moreover, $-s^{-1}\mathbf{H}_{l_s}(\boldsymbol{\zeta}) = \mathbf{H}_\kappa(\boldsymbol{\zeta}) = \mathbf{I}_f(\boldsymbol{\zeta}) > 0$, $\boldsymbol{\zeta} \in \Theta$, and, hence,

$$2sT_{LR}^{(s)}(\tilde{\boldsymbol{X}}^{(s)}) = ||\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{\frac{1}{2}}\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)})||_2^2,$$

where $\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{\frac{1}{2}} > 0$ is a positive definite matrix satisfying $\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{\frac{1}{2}}\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{\frac{1}{2}} = \mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})$. Applying La. 2.2.12 and using the fact that $\mathbf{I}_f(\bullet)^{\frac{1}{2}}$ is continuous on Θ, we obtain that $\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{\frac{1}{2}} \stackrel{s\to\infty}{\longrightarrow} \mathbf{I}_f(\boldsymbol{\zeta}^{(0)})^{\frac{1}{2}}$ $P_{\boldsymbol{\zeta}^{(0)}}$-a.s., and La. 2.2.15 and the multivariate version of Slutsky's theorem (e.g., in Sen & Singer (1993), Thm. 3.4.3, p. 130) yield

$$\mathbf{I}_f(\tilde{\boldsymbol{\zeta}}^{(s)})^{\frac{1}{2}}\sqrt{s}(\boldsymbol{\zeta}^{*(s)} - \boldsymbol{\zeta}^{(0)}) \stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k).$$

In virtue of the mapping $(y_1, \ldots, y_k)' \mapsto \sum_{j=1}^k y_j^2$, the assertion then follows from the continuous mapping theorem (e.g., in Billingsley (1999), Thm. 2.7, p. 21). ∎

8.2 Notations and Abbreviations

8.2.1 Mathematical Symbols

$\mathbb{N}$	$\{1, 2, ...\}$
$\mathbb{N}_0$	$\{0, 1, 2, ...\}$
$\mathbb{R}$	$(-\infty, \infty)$
$\mathbb{R}_+$	$(0, \infty)$
$\mathbb{R}_-$	$(-\infty, 0)$
$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$
$\mathbb{R}^n \equiv \mathbb{R}^{n\times 1}$	n-dimensional Euklidean space (column vectors)
$\mathbb{R}^n_+$	$\{(y_1, ..., y_n)' \in \mathbb{R}^n : \ y_i > 0, \ i = 1, ..., n\}$
$\mathbb{R}^n_-$	$\{(y_1, ..., y_n)' \in \mathbb{R}^n : \ y_i < 0, \ i = 1, ..., n\}$
$\mathbb{R}^n_<$	(truncated) cone of strictly increasing numbers in $\mathbb{R}^n$
$\mathbb{R}^{n\times k}$	$(n \times k)$-matrices with real-valued entries
$\mathbb{B}^n$	borel sets of $\mathbb{R}^n$
$A \cap \mathbb{B}^n$	borel sets $\{A \cap B : \ B \in \mathbb{B}^n\}$ of $A \in \mathbb{B}^n$
λ^n	Lebesgue measure on $(\mathbb{R}^n, \mathbb{B}^n)$
$\lambda^n\vert_A$	restriction of λ^n to the measurable space $(A, A \cap \mathbb{B}^n)$, $A \in \mathbb{B}^n$
$\times_{i=1}^n \mathfrak{X}_i$	cartesian product of $\mathfrak{X}_1, \ldots, \mathfrak{X}_n$
$\mathfrak{X}^{1\times n} \equiv \times_{i=1}^n \mathfrak{X}$	n-dimensional (row) vectors with entries in $\mathfrak{X}$
$\otimes_{i=1}^n \mathfrak{B}_i$	product sigma algebra of $\mathfrak{B}_1, \ldots, \mathfrak{B}_n$
$\mathfrak{B}^n$	product sigma algebra $\otimes_{i=1}^n \mathfrak{B}$
$\otimes_{i=1}^n \mu_i$	product measure of $\mu_1, \ldots, \mu_n$
$\mu^{(n)}$	product measure $\otimes_{i=1}^n \mu$
$X : \ (\Omega, \mathfrak{A}) \to (\mathfrak{X}, \mathfrak{B})$	$X : \ \Omega \to \mathfrak{X}$ is $\mathfrak{A}$-$\mathfrak{B}$-measurable
$f \equiv a$	function f identically equals the constant a
$[\mu]$, μ-a.e.; $[P]$, P-a.s.	μ-almost everywhere; P-almost sure
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$\mu = h\nu$, $h = \frac{d\mu}{d\nu}$	μ has a ν-density h
$\mathbb{1}_B$	indicator function of a measurable set B, i.e. $\mathbb{1}_B(y) = 1$ if $y \in B$ and 0 otherwise
$int(B)$	interior of set B
$cl(B)$	closure of set B
$\mathbf{I}_n$	n-dimensional unit matrix
$\boldsymbol{v}'$; $\mathbf{A}'$	transpose of vector $\boldsymbol{v}$; transpose of matrix $\mathbf{A}$
$\vert\mathbf{A}\vert$	determinant of matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$
$\mathbf{A}^{-1}$	inverse matrix of $\mathbf{A} \in \mathbb{R}^{n\times n}$
$\nabla_i f$, $\frac{d}{dx_i} f$	partial derivative of $f : \ \mathbb{R}^n \to \mathbb{R}$

∇f	gradient of $f : \mathbb{R}^n \to \mathbb{R}$, i.e. the column vector $\left(\frac{d}{dx_1}f, ..., \frac{d}{dx_n}f\right)' \in \mathbb{R}^n$ of partial derivatives of f
$\mathbf{H}_f$	Hessian matrix of $f : \mathbb{R}^n \to \mathbb{R}$, i.e. the matrix $\left\{\frac{d^2}{dx_i dx_j}f\right\}_{1\leq i,j\leq n} \in \mathbb{R}^{n\times n}$ of second partial derivatives of f
$\mathbf{D}_f$	Jacobian matrix of $f = (f_1, ..., f_m)' : \mathbb{R}^n \to \mathbb{R}^m$, i.e. the matrix $\left\{\frac{d}{dx_j}f_i\right\}_{1\leq i\leq m, 1\leq j\leq n} \in \mathbb{R}^{m\times n}$

8.2.2 Abbreviations

OS	order statistic
gOS	generalized order statistic
SOS	sequential order statistic
iid	independent and identically distributed
inid	independent, not necessarily identically distributed
MLE	maximum likelihood estimator
UMVUE	uniformly minimum variance unbiased estimator
UMP test	uniformly most powerful test
UMPU test	uniformly most powerful unbiased test
LR test	likelihood ratio test
ARE	asymptotic relative efficiency
Def.	Definition
Thm.	Theorem
La.	Lemma
Cor.	Corollary
Rem.	Remark
Ex.	Example
p.; pp.	page; pages
ff.	and following

Bibliography

Abramowitz, M. & Stegun, I. A., eds (1965), *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, Dover, New York.

Bahadur, R. R. (1964), 'On Fisher's bound for asymptotic variances', *Ann. Math. Statist.* **35**(4), 1545–1552.

Bahadur, R. R. (1971), *Some Limit Theorems in Statistics*, SIAM, Philadelphia.

Balakrishnan, N. & Aggarwala, R. (2000), *Progressive Censoring: Theory, Methods, and Applications*, Birkhäuser.

Balakrishnan, N., Beutner, E. & Kamps, U. (2008), 'Order restricted inference for sequential k-out-of-n systems', *J. Multivariate Anal.* **99**(7), 1489–1502.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972), *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*, Wiley, London.

Barndorff-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Bauer, H. (1990), *Maß- und Integrationstheorie*, De Gruyter, Berlin.

Bedbur, S. (2010), 'UMPU tests based on sequential order statistics', *J. Statist. Plann. Inference* **140**(9), 2520–2530.

Bedbur, S., Beutner, E. & Kamps, U. (2010), 'Generalized order statistics: an exponential family in model parameters', *Statistics* (to appear).

Belzunce, F., Ruiz, J. M. & Suárez-Llorens, A. (2008), 'On multivariate dispersion orderings based on the standard construction', *Statist. Probab. Lett.* **78**(3), 271–281.

Beutner, E. (2008), 'Nonparametric inference for sequential k-out-of-n systems', *Ann. Inst. Statist. Math.* **60**(3), 605–626.

Beutner, E. (2010), 'Nonparametric model checking for k-out-of-n systems', *J. Statist. Plann. Inference* **140**(3), 626–639.

Beutner, E. & Kamps, U. (2009), 'Order restricted statistical inference for scale parameters based on sequential order statistics', *J. Statist. Plann. Inference* **139**(9), 2963–2969.

Bieniek, M. (2008), 'Projection bounds on expectations of spacings of generalized order statistics from DFR and DFRA families', *Statistics* **42**(3), 231–243.

Billingsley, P. (1995), *Probability and Measure*, third edn, Wiley, New York.

Billingsley, P. (1999), *Convergence of Probability Measures*, second edn, Wiley, New York.

Birnbaum, A. (1955), 'Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests', *Ann. Math. Statist.* **26**(1), 21–36.

Bradley, R. A. & Gart, J. J. (1962), 'The asymptotic properties of ML estimators when sampling from associated populations', *Biometrika* **49**(1,2), 205–214.

Burkschat, M. (2009), 'Multivariate dependence of spacings of generalized order statistics', *J. Multivariate Anal.* **100**(6), 1093–1106.

Burkschat, M. (2010), 'Linear estimators and predictors based on generalized order statistics from generalized Pareto distributions', *Comm. Statist. Theory Methods* **39**(2), 311–326.

Chen, J. & Rubin, H. (1986), 'Bounds for the difference between median and mean of gamma and poisson distributions', *Statist. Probab. Lett.* **4**(6), 281–283.

Choi, K., P. (1994), 'On the medians of gamma distributions and an equation of Ramanujan', *Proc. Amer. Math. Soc.* **121**(1), 245–251.

Cramer, E. (2006), 'Dependence structure of generalized order statistics', *Statistics* **40**(5), 409–413.

Cramer, E. & Kamps, U. (1996), 'Sequential order statistics and k-out-of-n systems with sequentially adjusted failure rates', *Ann. Inst. Statist. Math.* **48**(3), 535–549.

Cramer, E. & Kamps, U. (2001*a*), 'Estimation with sequential order statistics from exponential distributions', *Ann. Inst. Statist. Math.* **53**(2), 307–324.

Cramer, E. & Kamps, U. (2001*b*), Sequential k-out-of-n systems, *in* N. Balakrishnan & C. R. Rao, eds, 'Advances in Reliability', Vol. 20 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 301–372.

Cramer, E. & Kamps, U. (2003), 'Marginal distributions of sequential and generalized order statistics', *Metrika* **58**(3), 293–310.

David, H. A. & Nagaraja, H. N. (2003), *Order Statistics*, third edn, Wiley, New York.

Heuser, H. (2000), *Lehrbuch der Analysis, Teil 2*, 11th edn, Teubner, Stuttgart.

Hollander, M. & Peña, E. A. (1995), 'Dynamic reliability models with conditional proportional hazards', *Lifetime Data Anal.* **1**(4), 377–401.

Kallenberg, W. C. M. (1978), *Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families*, Mathematical Centre Tracts 77, Amsterdam.

Kallenberg, W. C. M. (1983), 'Intermediate efficiency, theory and examples', *Ann. Statist.* **11**(1), 170–182.

Kamps, U. (1995*a*), 'A concept of generalized order statistics', *J. Statist. Plann. Inference* **48**(1), 1–23.

Kamps, U. (1995*b*), *A Concept of Generalized Order Statistics*, Teubner, Stuttgart.

Keating, J. P., Mason, R. L. & Sen, P. K. (1993), *Pitman's Measure of Closeness: A Comparison of Statistical Estimators*, Siam, Philadelphia.

Kim, G.-W. (1997), 'Exact slopes of test statistics for the multivariate exponential family', *Scand. J. Statist.* **24**(3), 387–406.

Kim, H. & Kvam, P. H. (2004), 'Reliability estimation based on system data with an unknown load share rule', *Lifetime Data Anal.* **10**(1), 83–94.

Kotz, S. & Johnson, N. L., eds (1983), *Encyclopedia of Statistical Sciences*, Vol. 4, Wiley, New York.

Lehmann, E. L. & Casella, G. (1998), *Theory of Point Estimation*, second edn, Springer, New York.

Lehmann, E. L. & Romano, J. P. (2005), *Testing Statistical Hypotheses*, third edn, Springer, New York.

Mathes, T. K. & Truax, D. R. (1967), 'Tests of composite hypotheses for the multivariate exponential family', *Ann. Math. Statist.* **38**(3), 681–697.

Neyman, J. & Pearson, E. S. (1928), 'On the use and interpretation of certain test criteria for purposes of statistical inference', *Biometrika* **20A**(1-4), 175–240, 263–294.

Nikitin, Y. (1995), *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, Cambridge.

Raghavachari, M. (1970), 'On a theorem of Bahadur on the rate of convergence of test statistics', *Ann. Math. Statist.* **41**(5), 1695–1699.

Rao, C. R. (1948), 'Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation', *Math. Proc. Cambridge Philos.* **44**(1), 50–57.

Rao, C. R. (2005), Score Test: Historical Review and Recent Developments, *in* N. Balakrishnan, N. Kannan & H. N. Nagaraja, eds, 'Advances in Ranking and Selection, Multiple Comparisons, and Reliability', Birkhäuser, Boston, pp. 3–20.

Rencher, A. C. (1998), *Multivariate Statistical Inference and Applications*, Wiley, New York.

Robertson, T., Wright, F. T. & Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, Wiley, Chichester.

Sen, P. K. (1986), 'Are BAN estimators the Pitman-closest ones too?', *Sankhyā* **48**(A), 51–58.

Sen, P. K. & Singer, J. M. (1993), *Large Sample Methods in Statistics: An Introduction with Applications*, first edn, Chapmann & Hall/CRC, Boca Raton.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shao, J. (2003), *Mathematical Statistics*, second edn, Springer, New York.

Stehlik, M. (2003), 'Distributions of exact tests in the exponential family', *Metrika* **57**(2), 145–164.

Wald, A. (1943), 'Tests of statistical hypotheses concerning several parameters when the number of observations is large', *Trans. Amer. Math. Soc.* **54**(3), 426–482.

Wellek, S. (2003), *Testing Statistical Hypotheses of Equivalence*, Chapman & Hall/CRC, Boca Raton.

Wilks, S. S. (1962), *Mathematical Statistics*, Wiley, New York.

Witting, H. (1985), *Mathematische Statistik I*, Teubner, Stuttgart.

Wolfowitz, J. (1965), 'Asymptotic efficiency of the maximum likelihood estimator', *Theory Probab. Appl.* **10**(2), 247–260.