UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS

TESIS DOCTORAL

Robust statistical inference for one-shot devices based on divergences

Inferencia estadística robusta basada en divergencias para dispositivos de un sólo uso

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Elena María Castilla González

Directores

Nirian Martín Apaolaza
Leandro Pardo Llorente

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS

TESIS DOCTORAL

Robust statistical inference for one-shot devices based on divergences

Inferencia estadística robusta basada en divergencias para dispositivos de un sólo uso

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Elena María Castilla González

DIRECTORES

Nirian Martín Apaolaza
Leandro Pardo Llorente

Madrid

Universidad Complutense de Madrid

Facultad de Ciencias Matemáticas

Robust Statistical Inference for One-shot devices based on Divergences

Inferencia estadística robusta basada en divergencias para dispositivos de un sólo uso

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics and Statistics

Author:

Elena María Castilla González

Supervisors:

Nirian Martín Apaolaza
Leandro Pardo Llorente

Madrid 2021

Programa de Doctorado en Ingeniería Matemática, Estadística e Investigación Operativa por la Universidad Complutense de Madrid y la Universidad Politécnica de Madrid

Robust Statistical Inference for One-shot devices based on Divergences

Tesis Doctoral

ELENA MARÍA CASTILLA GONZÁLEZ

Directores:

NIRIAN MARTÍN APAOLAZA **LEANDRO PARDO LLORENTE**

Año 2021

A mis abuelos Ramón y José Luis; a mis abuelas Antonia y Mary. Os quiero.

A Pedro. Siempre gracias.

Summary

A one-shot device is a unit that performs its function only once and, after use, the device either gets destroyed or must be rebuilt. For this kind of device, one can only know whether the failure time is either before or after a specific inspection time, and consequently the lifetimes are either left- or right-censored, with the lifetime being less than the inspection time if the test outcome is a failure (resulting in left censoring) and the lifetime being more than the inspection time if the test outcome is a success (resulting in right censoring). An accelerated life test (ALT) plan is usually employed to evaluate the reliability of such products by increasing the levels of stress factors and then extrapolating the life characteristics from high stress conditions to normal operating conditions. This acceleration process will shorten the life span of devices and reduce the costs associated with the experiment. The study of one-shot device from ALT data has been developed considerably recently, mainly motivated by the work of Fan et al. [2009].

In the last decades the use of divergence measures in the resolution of statistical problems has reached a remarkable relevance among the statisticians. It can be seen in Basu et al. [2011] and Pardo [2005] the importance of divergence measures in the areas of parametric estimation and parametric tests of hypotheses, together with many non-parametric uses. In particular, the minimum density power divergence estimators, introduced by Basu et al. [1998], are well-known to have robust statistical properties. Along this Thesis, robust estimators and tests are developed based on density power divergences for one-shot device testing.

In Chapter 2, we consider the problem of one-shot device testing along with an accelerating factor, in which the failure time of the devices is assumed to follow an exponential distribution. In this context, Fan et al. [2009] considered a single stress factor to the accelerated life test plan for one-shot devices, and analyzed the data by using a Bayesian approach in which the model parameters in the prior information were assumed to be close to the true values. In contrast, Balakrishnan and Ling [2012a] developed an EM algorithm for a single stress model, and made a comparative study with the mentioned Bayesian approach, showing that the EM method is more appropriate for moderately and lowly reliable products. Proposed minimum density power divergence estimators and Z-type tests are shown, both theoretically and empirically, to present a much more robust behavior than the classical MLE and Z-test. However, as opposed to a single-stress test by using a high stress level so as to attain the aging within a limited time, some ALTs involve two or more stress factors. Effectively, multiple-stress model becomes better suited for the prediction of lifetimes of products, subjected to, for example, electrical, thermal or mechanical stresses; see, for example, Srinivas and Ramu [1992] and Bartnikas and Morin [2004]. In Balakrishnan and Ling [2012b], an EM algorithm for developing inference is developed, based on one-shot device testing data under the exponential distribution when there are multiple stress factors. In Chapter 3, we extend the results in Chapter 2 to multiple-stress ALTs. In this case, instead of Z-type tests, we muss define Wald-type tests, which are shown, by means of an extensive simulation study, to be much more robust than classical Wald-test. In Chapter 4, we extend the results of Chapter 3 by assuming that the lifetimes follow a gamma distribution. Gamma distribution is commonly used for fitting lifetime data in reliability and survival studies due to its flexibility. Its hazard function can be increasing, decreasing, and constant. When the hazard function of gamma distribution is a constant, it corresponds to the exponential distribution. In addition to the exponential distribution,

the gamma distribution also includes the Chi-square distribution as a special case.

In practice, the Weibull distribution is widely used as a lifetime model in engineering and physical sciences. In fact, the Weibull model is also used extensively in biomedical studies as a proportional hazards model for evaluating the effects of covariates on lifetimes, meaning that the hazard rates of any two products stay in constant ratio over time. See Meeter and Meeker [1994], Meeker et al. [1998], and references therein. However, in some situations,the assumption of constant shape parameters may not be valid; see, for example, Kodell and Nelson [1980], Nogueira et al. [2009] and Vázquez et al. [2010]. In such situations, Balakrishnan and Ling [2013] suggested using a log-link of the stress levels to model the unequal shape parameters. Based on this idea, we develop, in Chapter 5, robust inference for one-shot device testing under the Weibull distribution with scale and shape parameters varying over stress. Other distributions may be considered for modeling the lifetimes. In Chapter 6, we consider the Lindley and lognormal distributions. The Lindley distribution, introduced by Lindley [1958], has shown to give better modeling that the exponential distribution in some contexts (see Ghitany et al. [2008]). On the other hand, the lognormal distribution has been studied in different types of censored data, see, for example, Meeker [1984] and Ng et al. [2002].

Under the classical parametric setup, product lifetimes are assumed to be fully described by a probability distribution involving some model parameters. However, as data from one-shot devices do not contain actual lifetimes, parametric inferential methods can be very sensitive to violations of the model assumption. Ling et al. [2015] proposed a semi-parametric model, in which, under the proportional hazards assumption, the hazard rate is allowed to change in a non-parametric way. However, this methods suffer again from lack of robustness, as it is based on the MLE of model parameters. In Chapter 7, we develop robust estimators and tests for one-shot device testing based on divergence measures under proportional hazards model.

In lifetime data analysis, it is often the case that the products under study can experience one of different types of failure. For example, in the context of survival analysis, we can have several different types of failure (death, relapse, opportunistic infection, etc.) that are of interest to us, leading to the so-called "competing risks" scenario. A competing risk is an event whose occurrence precludes the occurrence of the primary event of interest. Balakrishnan et al. [2015a,b] have discussed the problem of one-shot devices under competing risk for the first time. The main purpose of Chapter 8 is to develop weighted minimum density power divergence estimators as well as Wald-type test statistics under competing risk models for one-shot device testing assuming exponential lifetimes. Chapter 9 finally provides some concluding remarks and also points out some further problems of interest. The Appendix briefly presents some other results, which have also been obtained by the candidate during her Ph.D. studies.

Resumen

Los dispositivos de un sólo uso (*one shot devices* en inglés), son aquellos que, una vez usados, dejan de funcionar. La mayor dificultad a la hora de modelizar su tiempo de vida es que sólo se puede saber si el momento de fallo se produce antes o después de un momento específico de inspección. Así pues, se trata de un caso extremo de censura interválica: si el tiempo de vida es inferior al de inspección observaremos un fallo (censura por la izquierda), mientras que si el tiempo de vida es mayor que el tiempo de inspección, observaremos un éxito (censura por la derecha). Para la observación y modelización de este tipo de dispositivos es común el uso de tests de vida acelerados. Los tests de vida acelerados permiten evaluar la fiabilidad de los productos en menos tiempo, incrementando las condiciones a las que se ven sometidos los dispositivos para extrapolar después estos resultados a condiciones más normales. El estudio de los dispositivos de un sólo uso por medio de tests de vida acelerados se ha incrementado considerablemente en los últimos años motivado, principalmente, por el trabajo de Fan et al. [2009].

Por otra parte, en las últimas décadas, el uso de medidas de divergencia en la resolución de problemas estadísticos ha ganado gran importancia dentro de la investigación. Por ejemplo, en Basu et al. [2011] y Pardo [2005], se puede apreciar la relevancia de las medidas de divergencia en la estimación paramétrica y tests de hipótesis paramétricos, así como para otros muchos usos. En particular, los estimadores de mínima densidad de potencia (*minimum density power divergence estimators* en inglés), introducidos en Basu et al. [1998], son muy importantes debido a su robustez. A lo largo de esta Tesis, se desarrollarán estimadores y tests robustos para los dispositivos de un sólo uso basados en estas divergencias.

En el Capítulo 2, presentamos el problema de los dispositivos de un sólo uso con un único factor de estrés, asumiendo que el tiempo de fallo de los dispositivos sigue una distribución exponencial. En este contexto, Fan et al. [2009] consideraron un único factor de estrés para los tests de vida acelerados en el contexto de dispositivos de un sólo uso, y analizaron los datos usando un enfoque Bayesiano en el que los parámetros de la información a priori se asumían cercanos a los verdaderos valores. Por otro lado, Balakrishnan and Ling [2012a] desarrollaron un algoritmo EM para un factor de estrés, e hicieron un estudio comparativo con el método Bayesiano antes mencionado, mostrando que el método EM es más apropiado para productos con media o baja fiabilidad. Los estimadores y tests de tipo Z que proponemos en este capítulo muestran, de forma tanto teórica como empírica, ser más robustos que el estimador de máxima verosimilitud (EMV) y el clásico test o prueba Z. Sin embargo, muchos tests de vida acelerados constan de más de un factor de estrés, lo cuál puede resultar más preciso para la predicción de los tiempos de vida (Srinivas and Ramu [1992],Bartnikas and Morin [2004]). En Balakrishnan and Ling [2012b], se desarrolla un algoritmo EM para múltiples factores de estrés asumiendo que los tiempos de vida siguen una distribución exponencial. En el Capítulo 3, extendemos los resultados del Capítulo 2 al caso de múltiples factores de estrés. En este caso, en lugar de tests de tipo Z, tenemos que definir tests de tipo Wald, los cuáles generalizan el clásico test de Wald. Con esta idea en mente, en el Capítulo 4, asumimos que los tiempos de vida siguen una distribución gamma. Esta distribución se usa comúnmente en estudios de supervivencia y fiabilidad debido a su flexibilidad. Su función de riesgo puede ser creciente, decreciente o constante. En este último caso, la distribución gamma corresponde a la exponencial. Aparte de ésta, la distribución gamma también contiene a la Chi-cuadrado como caso

particular.

En la práctica, la distribución Weibull es ampliamente utilizada en ingeniería y ciencias físicas. De hecho, esta distribución también se usa habitualmente en estudios biomédicos para un modelo de riesgos proporcional que evalue el efecto de las covariables, ya que los ratios o tasas de riesgo de cualesquiera dos productos se mantienen constantes en el tiempo, véase por ejemplo Meeter and Meeker [1994], Meeker et al. [1998]. Sin embargo, no siempre es válido asumir ésto (Kodell and Nelson [1980], Nogueira et al. [2009] y Vázquez et al. [2010]). En esos casos, Balakrishnan and Ling [2013] sugirieron relacionar los factores de estrés a los parámetros de forma. Basándonos en esta idea, desarrollamos, en el Capítulo 5 inferencia robusta para los dispositivos de un sólo uso bajo la distribución Weibull asumiendo que los parámetros de escala y forma varían con los factores de estrés. También podríamos considerar otras distribuciones para modelizar los tiempos de vida. En el Capítulo 6, consideramos las distribuciones Lindley y lognormales. En algunos contextos se ha demostrado que la distribución Lindley, introducida por Lindley [1958], da mejores resultados que la distribución exponencial (Ghitany et al. [2008]). Por otra parte, la distribución lognormal ha sido estudiada en diferentes tipos de datos censurados, por ejemplo Meeker [1984] y Ng et al. [2002].

En el clásico modelo paramétrico, se asume que las vidas de los productos están completamente descritas por una distribución de probabilidad con ciertos parámetros. Sin emabrgo, como no se pueden conoccer el momento real de fallo de los dispositivos de un sólo uso, los métodos parmétricos pueden ser demasiado sensibles. Ling et al. [2015] propusieron un modelo semi-paramétrico, en el que, bajo el modelo de riesgos proporcionales, el ratio de riesgo puede variar de forma no paramétrica. Sin embargo, este método sufre de nuevo de falta de robustez, ya que se basa en el EMV de los parámetros del modelo. En el Capítulo 7, definimos estimadores y tests robustos basados en medidas de divergencia para el modelo de riesgos proporcionales.

Si bien en los capítulos descritos se asume que hay una sola causa de riesgo, es habitual en el estudio de este tipo de datos, que los dispositivos puedan tener diferentes causas de fallo. Por ejemplo, en el contexto de análisis de supervivencia, podemos tener varias causas de fallo (muerte, recaída, infección, etc.) que son de interés, llevando a lo que llamamos escenario de riesgos competitivos (*competing risks* en inglés). Balakrishnan et al. [2015a,b] trataron por primera vez el problema de dispositivos de un sólo uso bajo riesgos competitivos. El principal objetivo del Capítulo 8 es desarrollar estimadores y tests para este caso asumiendo distribuciones exponenciales. Finalmente, en el Capítulo 9 se presentan algunas conclusiones finales y se dan unas pinceladas sobre posibles líneas de investigación futuras. El Apéndice presenta de forma breve otros resultados, obtenidos por la candidata durante la elaboración de su Tesis.

Table of Contents

Chapter 1

Introduction

1.1 Motivation of the Thesis

On May 2015, Professor N. Balakrishnan (McMaster Uinversity, Ontario, Canada) was invited by Professor L. Pardo and the Department of Statistics and Operational Research at Complutense University of Madrid (Madrid, Spain), to give a talk entitled "One-Shot Device Testing and Analysis". In this talk, Professor N. Balakrishnan first introduced the one-shot devices and the corresponding form of test and data and made an overview of some results related to the EM-algorithm for this kind of devices under different lifetime distribution assumptions. All these results had resulted on the publication of several papers (see, for example, Balakrishnan and Ling [2012a,b, 2013]) and were collected on the Thesis of Dr. Ling on 2012 (Ling [2012]). An extension of these results can be found on several papers collected in the Thesis of Dr. So (So [2016]).

While all these results deal with the efficiency of the estimation on one-shot devices, the robustness of these estimators was not considered. In this regard, Professors N. Martin, L. Pardo and N. Balakrishnan discussed the possibility of applying divergence measures to one-shot device testing to deal with this problem. In particular, the density power divergence was known to have good robustness properties in several statistical models. This idea, which also resulted on the concession of the National Research Project MTM2015-67057-P, can be considered the origin of this Thesis.

This work, developed under the supervision of Professors N. Martín and L. Pardo, has been funded by the Santander Bank Funding Program (Complutense University of Madrid) and by an FPU scholarship (FPU 16/03104). It has also received support from the Research Projects MTM2015-67057-P and PGC2018-095194-B-I00. Three main research stays have been undoubtedly essential in the development of this work. The first two (July-August 2016 and June-August 2018) were carried out in McMaster Uinversity (Ontario, Canada) under the supervision of Professor N. Balakrishnan. The last one (May-July 2019) was carried out in the University of Ioannina (Greece) under the supervision of Professor K. Zografos.

1.2 Divergence measures

In the last decades the use of divergence measures in the resolution of statistical problems has reached a remarkable relevance among the statisticians. It can be seen in Basu et al. [2011] and Pardo [2005] the importance of divergence measures in the areas of parametric estimation and parametric tests of hypotheses, together with many non-parametric uses. In the following, in accordance with the scope of this Thesis we focus on parametric methods. In estimation theory is very intuitive the role of the divergence measures in order to get estimates of the unknown parameters: Minimizing a suitable divergence measure between the data and the assumed model.

From a historical point of view was Wolfowitz [1952, 1953, 1954, 1957] who considered for the first time the possibility to use divergence measures (distances) in statistical inference. The robustness properties of many minimum divergence estimators in relation to the maximum likelihood estimator (MLE), without a significant loss of efficiency, have been one of the most important reasons for which that statistical procedures become more popular every day. Important works in which it is possible to see these facts are, for instance: Beran et al. [1977], Lindsay et al. [1994], Simpson [1987, 1989] and Tamura and Boos [1986]. Based on these minimum divergence estimators has been possible to get test statistics that have better robustness properties that the classical likelihood ratio tests, Wald tests of Rao's tests. In this Thesis, we shall use divergence measures in order to present robust inference procedures for one-shot devices that we shall describe in the next sections.

The statistical distances or measures of divergence can be classified in two different groups:

1. Distances between the distribution function of the data and the model distribution. Examples include the Kolmogorov-Smirnov distance, the Cramer-von Mises distance, see Mises [1936, 1939, 1947], the Anderson-Darling distance (Anderson and Darling [1952]), etc...

2. Distances or divergence measures between the probability density function or probability mass function of data (such as a nonparametric density estimator or the vector of relative frequencies) and the model density. The term "divergence" for a statistical distance was used formally by Bhattacharyya [1943, 1946] and the term was popularized by its use for Kulllback-Leibler divergence in Kullback and Leibler (1951), its use in the textbook Kullback (1959), and then by Ali and Silvey [1966] and Csiszar [1963], for the class of ϕ-divergences. The three more important families of divergences of this type are: ϕ-divergence measures, Bregman divergences and Burbea-Rao divergences.

In this Thesis we pay special attention to some members of the Bregman divergences and ϕ-divergence measures. We are going to describe these two classes of divergence measures. We shall introduce some additional notation. Let $\boldsymbol{X}$ be a random variable taking values on a sample space $\mathcal{X}$ (usually $\mathcal{X}$ will be a subset of $\mathbb{R}^n$, n-dimensional Euclidean space). Suppose that the distribution function F of $\boldsymbol{X}$ depends on a certain number of parameters, and suppose further that the functional form of F is known except perhaps for a finite number of these parameters; we denote by $\boldsymbol{\theta}$ the vector of unknown parameters associated with F. Let $(\mathcal{X}, \beta_{\mathcal{X}}, P_\theta)_{\boldsymbol{\theta}\in\Theta}$ be the statistical space associated with the random variable $\boldsymbol{X}$, where $\beta_{\mathcal{X}}$ is the σ-field of Borel subsets $A \subset \mathcal{X}$ and $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ a family of probability distributions defined on the measurable space $(\mathcal{X}, \beta_{\mathcal{X}})$ with Θ an open subset of $\mathbb{R}^{M_0}$, $M_0 \geq 1$. In the following the support of the probability distribution $P_{\boldsymbol{\theta}}$ is denoted by $S_{\mathcal{X}}$.

We assume that the probability distributions $P_{\boldsymbol{\theta}}$ are absolutely continuous with respect to a σ-finite measure μ on $(\mathcal{X},\, \beta_{\mathcal{X}})$. For simplicity μ is either the Lebesgue measure (i.e., satisfying the condition $P_{\boldsymbol{\theta}}(C) = 0$, whenever C has zero Lebesgue measure), or a counting measure (i.e., there exists a finite or countable set$S_{\mathcal{X}}$ with the property $P_\theta\,(\mathcal{X} - S_{\mathcal{X}}) = 0$). In the following

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{dP_{\boldsymbol{\theta}}}{d\mu}(\boldsymbol{x}) = \left\{ \begin{array}{ll} f_{\boldsymbol{\theta}}(\boldsymbol{x}) & \text{if } \mu \text{ is the Lebesgue measure,} \\ \Pr_{\boldsymbol{\theta}}\left(\boldsymbol{X}{=}\boldsymbol{x}\right) = p_{\boldsymbol{\theta}}(\boldsymbol{x}) & \text{if } \mu \text{ is a counting measure,} \\ & (\boldsymbol{x} \in S_{\mathcal{X}}) \end{array} \right.$$

denotes the family of probability density functions if μ is the Lebesgue measure, or the family of probability mass functions if μ is a counting measure. In the first case $\boldsymbol{X}$ is a random variable with absolutely continuous distribution and in the second case it is a discrete random variable with support $S_{\mathcal{X}}$.

1.2.1 Bregman's divergence measures

Bregman [1967] introduced a family of divergences measures, between the probability distributions $P_{\boldsymbol{\theta}_1}$ and $P_{\boldsymbol{\theta}_2}$, by

$$B_{\varphi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\int_{\mathcal{X}}\left(\varphi\left(f_{\boldsymbol{\theta}_1}\left(\boldsymbol{x}\right)\right)-\varphi\left(f_{\boldsymbol{\theta}_2}\left(\boldsymbol{x}\right)\right)\right)-\varphi'\left(f_{\boldsymbol{\theta}_2}\left(\boldsymbol{x}\right)\right)\left(f_{\boldsymbol{\theta}_1}\left(\boldsymbol{x}\right)-f_{\boldsymbol{\theta}_2}\left(\boldsymbol{x}\right)\right)d\mu\left(\boldsymbol{x}\right)$$

for any differentiable convex function $\varphi:(0,\infty)\to\mathbb{R}$ with $\varphi(0)=\lim_{t\to 0}\varphi(t)\in(-\infty,\infty)$. It is important to note that for $\varphi(t)=t\log t$, we get the Kullback-Leibler divergence,

$$d_{KL}\left(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\right)=\int\limits_{\mathcal{X}} f_{\boldsymbol{\theta}_1}(\boldsymbol{x})\log\frac{f_{\boldsymbol{\theta}_1}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_2}(\boldsymbol{x})}d\mu\left(\boldsymbol{x}\right) \tag{1.1}$$

and for $\varphi(t)=t^2$ and discrete probability distributions, the Euclidean distance, namely

$$E\left(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\right)=\sum_{i=1}^{M}\left(p_{\boldsymbol{\theta}_1}\left(x_i\right)-p_{\boldsymbol{\theta}_2}\left(x_i\right)\right)^2. \tag{1.2}$$

But the most important family, from the point of view of this Thesis, is the family obtained when $\varphi_\tau(t)=\frac{1}{\tau}t^{1+\tau}$ with $\tau\geq 0$. The corresponding family of divergences is called "density power divergences" (DPD), whose expression is given by

$$d_\tau(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\int_{\mathcal{X}}\left(\frac{1}{\tau}f_{\boldsymbol{\theta}_1}^{1+\tau}\left(\boldsymbol{x}\right)-\frac{1+\tau}{\tau}f_{\boldsymbol{\theta}_2}^{\tau}\left(\boldsymbol{x}\right)f_{\boldsymbol{\theta}_1}\left(\boldsymbol{x}\right)+f_{\boldsymbol{\theta}_2}^{\tau+1}\left(\boldsymbol{x}\right)\right)d\mu\left(\boldsymbol{x}\right). \tag{1.3}$$

This family of divergence measures was considered for the first time in Basu et al. (1998). They established that $d_\tau(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\geq 0$. The expression for $\tau=0$ is obtained as

$$\lim_{\tau\to 0}d_\tau(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=d_{KL}\left(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\right)$$

whose expression is given in (1.1). For $\tau=1$ we get, for discrete distributions, the Euclidean distance given in (1.2).

It is interesting to note that the DPD not only is a member of the Bregman's divergence measures but also a member of the family nof divergences measures considered in Jones et al. [2001],

$$\begin{aligned} d_{\tau,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= \frac{1}{\beta}\left(\int_{\mathcal{X}}\frac{1}{\tau}f_{\boldsymbol{\theta}_1}^{1+\tau}\left(\boldsymbol{x}\right)d\mu\left(\boldsymbol{x}\right)\right)^{\beta}-\frac{1+\tau}{\tau}\frac{1}{\beta}\left(\int_{\mathcal{X}}f_{\boldsymbol{\theta}_2}^{\tau}\left(\boldsymbol{x}\right)f_{\boldsymbol{\theta}_1}\left(\boldsymbol{x}\right)d\mu\left(\boldsymbol{x}\right)\right)^{\beta} \\ &\quad +\frac{1}{\beta}\left(\int_{\mathcal{X}}f_{\boldsymbol{\theta}_2}^{1+\tau}\left(\boldsymbol{x}\right)d\mu\left(\boldsymbol{x}\right)\right)^{\beta}, \end{aligned}$$

as, for $\beta=1$, we have

$$d_{\tau,\beta=1}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=d_\tau(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2),$$

i.e., the DPD. For $\beta=0$, we have

$$\begin{aligned} \lim_{\beta\to 0}d_{\tau,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= \log\left(\int_{\mathcal{X}}\frac{1}{\tau}f_{\boldsymbol{\theta}_1}^{1+\tau}\left(\boldsymbol{x}\right)d\mu\left(\boldsymbol{x}\right)\right)-\frac{1+\tau}{\tau}\log\left(\int_{\mathcal{X}}f_{\boldsymbol{\theta}_2}^{\tau}\left(\boldsymbol{x}\right)f_{\boldsymbol{\theta}_1}\left(\boldsymbol{x}\right)d\mu\left(\boldsymbol{x}\right)\right) \\ &\quad +\log\left(\int_{\mathcal{X}}f_{\boldsymbol{\theta}_2}^{1+\tau}\left(\boldsymbol{x}\right)d\mu\left(\boldsymbol{x}\right)\right). \end{aligned}$$

Jones et al. [2001] considered the Rényi Pseudodistance given by

$$\begin{aligned} R_\alpha\left(g,f_{\boldsymbol{\theta}}\right) &= \frac{1}{\alpha+1}\log\left(\int_{\mathcal{X}}f_{\boldsymbol{\theta}}^{\alpha+1}(x)dx\right)+\frac{1}{\alpha\left(\alpha+1\right)}\log\left(\int_{\mathcal{X}}g^{\alpha+1}(x)dx\right) \\ &\quad -\frac{1}{\alpha}\log\left(\int_{\mathcal{X}}f_{\boldsymbol{\theta}}^{\alpha}(x)g(x)dx\right). \end{aligned} \tag{1.4}$$

It can be seen that

$$\lim_{\beta\to 0}d_{\tau,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\left(\alpha+1\right)R_\alpha\left(g,f_{\boldsymbol{\theta}}\right).$$

1.2.2 Phi-divergence measures

The family of ϕ-divergence measures defined simultaneously by Csiszar [1963] and Ali and Silvey [1966] is defined by,

$$d_{\phi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{\mathcal{X}} f_{\boldsymbol{\theta}_2}(\boldsymbol{x}) \phi \left(\frac{f_{\boldsymbol{\theta}_1}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_2}(\boldsymbol{x})} \right) d\mu\left(\boldsymbol{x}\right), \ \phi \in \Phi^* \tag{1.5}$$

where Φ^* is the class of all convex functions $\phi\left(x\right)$, $x > 0$, such that at $x = 1, \phi\left(1\right) = 0$, and at $x = 0$, $0\phi\left(0/0\right) = 0$ and $0\phi\left(p/0\right) = p \lim_{u\to\infty} \phi\left(u\right)/u$. For every $\phi \in \Phi^*$, that is differentiable at $x = 1$, the function

$$\psi\left(x\right) \equiv \phi\left(x\right) - \phi'\left(1\right)\left(x-1\right),$$

also belongs to Φ^*. Then we have $d_{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = D_{\phi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, and ψ has the additional property that $\psi'\left(1\right) = 0$. The most important properties of the ϕ-divergence measures can be seen in Pardo [2005]. The Kullback-Leibler divergence measure is obtained for $\psi\left(x\right) = x\log x - x + 1$ or $\phi\left(x\right) = x \log x$. We can observe that $\psi\left(x\right) = \phi\left(x\right) - \phi'(1)(x-1)$. We shall denote by ϕ any function belonging to Φ or Φ^*.

From a statistical point of view, the most important family of ϕ-divergences is perhaps the family studied by Cressie and Read [1984]: the power-divergence family, given by

$$I_{\lambda}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) \equiv D_{\phi_{(\lambda)}}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) = \frac{1}{\lambda\left(\lambda+1\right)} \left(\int_{\mathcal{X}} \frac{f_{\boldsymbol{\theta}_1}^{\lambda+1}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_2}^{\lambda}(\boldsymbol{x})} d\mu(\boldsymbol{x}) - 1 \right) \tag{1.6}$$

for $-\infty < \lambda < \infty$. The power-divergence family is undefined for $\lambda = -1$ or $\lambda = 0$. However, if we define these cases by the continuous limits of $I_{\lambda}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)$ as $\lambda \to -1$ and $\lambda \to 0$, then $I_{\lambda}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)$ is continuous in λ. It is not difficult to establish that

$$\lim_{\lambda \to 0} I_{\lambda}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) = d_{KL}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)$$

and

$$\lim_{\lambda \to -1} I_{\lambda}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) = d_{KL}\left(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1\right).$$

We can observe that the power-divergence family is obtained from (1.5) with

$$\phi\left(x\right) = \begin{cases} \phi_{(\lambda)}\left(x\right) = & \frac{1}{\lambda(\lambda+1)}\left(x^{\lambda+1} - x - \lambda\left(x-1\right)\right); \ \lambda \neq 0, \lambda \neq -1, \\ \phi_{(0)}\left(x\right) = & \lim_{\lambda\to 0} \phi_{(\lambda)}\left(x\right) = x\log x - x + 1, \\ \phi_{(-1)}\left(x\right) = & \lim_{\lambda\to -1} \phi_{(\lambda)}\left(x\right) = -\log x + x - 1. \end{cases}$$

The power-divergence family was proposed independently by Liese and Vajda [1987] as a ϕ-divergence under the name I_a-divergence.

In this Thesis we shall use the DPD, $d_{\tau}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, given in (1.3) the RP, $R_{\alpha}\left(g, f_{\boldsymbol{\theta}}\right)$, given in (1.4), the ϕ-divergences measures given in (1.5) and the power-divergence family given in (1.6).

1.3 Minimum distance estimators

Suppose we have n independent and identically distributed (IID) observations $X_1, ..., X_n$ from a unidimensional random variable X with distribution function G and we model the data generating distribution by the parametric family $(\mathcal{X}, \beta_{\mathcal{X}}, P_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ with model distribution function $F_{\boldsymbol{\theta}}$ and density function $f_{\boldsymbol{\theta}}$. Our aim is to estimate the unknown parameter $\boldsymbol{\theta}$ for which the model distribution $F_{\boldsymbol{\theta}}$ is a "good" approximation of G in a suitable sense. In the likelihood approach, maximizing this closeness translates to maximizing the probability of observing the sample data; the estimate of $\boldsymbol{\theta}$ corresponds to that particular model distribution, under which the probability (or, likelihood) of the observed sample is the maximum. The resulting estimator is known as the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$. We are going to present a justification of the MLE in terms of divergence measures.

We denote by g the density function associated to the distribution function G. The Kullback-Leibler divergence measure between g and $f_{\boldsymbol{\theta}}$ is given by

$$d_{KL}(g, f_{\boldsymbol{\theta}}) = \int_{\mathcal{X}} g(x) \log \frac{g(x)}{f_{\boldsymbol{\theta}}(x)} dx = \int_{\mathcal{X}} g(x) \log g(x) dx - \int_{\mathcal{X}} g(x) \log f_{\boldsymbol{\theta}}(x) dx.$$

In order to minimize in $\boldsymbol{\theta}$, $d_{KL}(g, f_{\boldsymbol{\theta}})$, it will be sufficient to minimize

$$-\int_{\mathcal{X}} g(x) \log f_{\boldsymbol{\theta}}(x) dx = \int_{\mathcal{X}} \log f_{\boldsymbol{\theta}}(x) dG(x).$$

But $G(x)$ is unknown and we can consider as estimator of $G(x)$ the empirical distribution function

$$G_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(x_i),$$

where I_A is the indicator function of the set A, based on a random sample of size n, $X_1, ..., X_n$. Then we have to minimize

$$-\int_{\mathcal{X}} \log f_{\boldsymbol{\theta}}(x) dG_n(x) = -\frac{1}{n} \sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(X_i)$$

or equivalently to maximize

$$\frac{1}{n} \sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(X_i) = \frac{1}{n} \prod_{i=1}^{n} f_{\boldsymbol{\theta}}(X_i)$$

i.e., we get the MLE. Therefore the MLE has an interpretation in term of the Kullback-Leibler divergence. This result is the main idea for the development of minimum distance estimators.

1.3.1 Minimum DPD estimators

The minimum DPD estimators were introduced by Basu et al. [1998], by defining

$$\widehat{\boldsymbol{\theta}}_\tau = \arg \min_{\theta \in \Theta} d_\tau(g, f_{\boldsymbol{\theta}}),$$

i.e., we must minimize

$$\int_{\mathcal{X}} \left(\frac{1}{\tau} f_{\boldsymbol{\theta}}^{1+\tau}(x) - \frac{1+\tau}{\tau} f_{\boldsymbol{\theta}}^{\tau}(x) g(x) + g^{\tau+1}(x) \right) dx.$$

But the term $\int_{\mathcal{X}} g^{\tau+1}(x) dx$ has not any role in the minimization in $\boldsymbol{\theta}$ of $d_\tau(g, f_{\boldsymbol{\theta}})$. Therefore, we must minimize

$$\int_{\mathcal{X}} \frac{1}{\tau} f_{\boldsymbol{\theta}}^{1+\tau}(x) dx - \frac{1+\tau}{\tau} \int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{\tau}(x) dG(x).$$

In the same way that previously we can estimate G using the empirical distribution function based on a random sample of size n, $X_1, ..., X_n$, i.e. we must minimize, for $\tau > 0$,

$$\int_{\mathcal{X}} \frac{1}{\tau} f_{\boldsymbol{\theta}}^{1+\tau}(x) dx - \frac{1+\tau}{\tau} \frac{1}{n} \sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\tau}(X_i).$$

and the negative loglikelihood

$$-\frac{1}{n} \sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(X_i) u_{\boldsymbol{\theta}}(X_i)$$

for $\tau = 0$. Differentiating with respect to $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_\tau$ can be also be defined by the estimating equation,

$$\frac{1}{n}\sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\tau}(X_i)\, u_{\boldsymbol{\theta}}(X_i) - \int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{1+\tau}(x)\, u_{\boldsymbol{\theta}}(x)dx = \mathbf{0}_p, \tag{1.7}$$

for $\tau > 0$, where $\mathbf{0}_p$ is the null vector of dimension p, being $u_{\boldsymbol{\theta}}(x) = \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \theta}$.

Therefore the minimum DPD estimator is defined by

$$\widehat{\boldsymbol{\theta}}_\tau = \begin{cases} \arg\min_{\theta\in\Theta} \left(\int_{\mathcal{X}} \frac{1}{\tau} f_{\boldsymbol{\theta}}^{1+\tau}(x)\, dx - \frac{1+\tau}{\tau}\frac{1}{n}\sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\tau}(X_i) \right) & \tau > 0 \\ \text{MLE} & \tau = 0 \end{cases}. \tag{1.8}$$

Basu et al. [1998] established the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_\tau$, by

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(\mathbf{0}_p, \boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta})\boldsymbol{K}_\tau(\boldsymbol{\theta})\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta})),$$

being

$$\boldsymbol{J}_\tau(\boldsymbol{\theta}) = \int_{\mathcal{X}} u_{\boldsymbol{\theta}}(x)u_{\boldsymbol{\theta}}^T(x)f_{\boldsymbol{\theta}}^{1+\tau}(x)dx + \int_{\mathcal{X}} \{i_{\boldsymbol{\theta}}(x) - \tau u_{\boldsymbol{\theta}}(x)u_{\boldsymbol{\theta}}^T(x)\}\{g(x) - f_{\boldsymbol{\theta}}(x)\}f_{\boldsymbol{\theta}}^{\tau}(x)dx \tag{1.9}$$

and

$$\boldsymbol{K}_\tau(\boldsymbol{\theta}) = \int_{\mathcal{X}} u_{\boldsymbol{\theta}}(x)u_{\boldsymbol{\theta}}^T(x)f_{\boldsymbol{\theta}}^{2\tau}(x)g(x)dx - \boldsymbol{\xi}_\tau(\boldsymbol{\theta})\boldsymbol{\xi}_\tau^T(\boldsymbol{\theta}), \tag{1.10}$$

where $\boldsymbol{\xi}_\tau(\boldsymbol{\theta}) = \int_{\mathcal{X}} u_{\boldsymbol{\theta}}(x)f_{\boldsymbol{\theta}}^{\tau}(x)g(x)dx$, and $i_{\boldsymbol{\theta}}(x) = -\frac{\partial}{\partial\boldsymbol{\theta}}u_{\boldsymbol{\theta}}(x)$, the so called information function of the model. When the true distribution G belongs to the model so that $G = F_{\boldsymbol{\theta}}$ for some $\boldsymbol{\theta} \in \Theta$, the formula for $\boldsymbol{J}_\tau(\boldsymbol{\theta}), \boldsymbol{K}_\tau(\boldsymbol{\theta})$ and $\boldsymbol{\xi}_\tau(\boldsymbol{\theta})$ simplify to

$$\begin{aligned} \boldsymbol{J}_\tau(\boldsymbol{\theta}) &= \int_{\mathcal{X}} u_{\boldsymbol{\theta}}(x)u_{\boldsymbol{\theta}}^T(x)f_{\boldsymbol{\theta}}^{1+\tau}(x)dx, \\ \boldsymbol{K}_\tau(\boldsymbol{\theta}) &= \int_{\mathcal{X}} u_{\boldsymbol{\theta}}(x)u_{\boldsymbol{\theta}}^T(x)f_{\boldsymbol{\theta}}^{1+2\tau}(x)dx - \xi_\tau(\boldsymbol{\theta})\xi_\tau^T(\boldsymbol{\theta}), \\ \boldsymbol{\xi}_\tau(\boldsymbol{\theta}) &= \int_{\mathcal{X}} u_{\boldsymbol{\theta}}(x)f_{\boldsymbol{\theta}}^{1+\tau}(x)dx. \end{aligned}$$

They also established that the minimum density power divergence estimating equation (1.7) has a consistent sequence of roots $\widehat{\boldsymbol{\theta}}_\beta = \widehat{\boldsymbol{\theta}}_n$.

This result were extended in Ghosh et al. [2013] to the situation in which the observations are independent but not identically distributed. Let us assume that our observations $X_1, \ldots, X_n$ are independent but for each i, the density function of X_i is $g_i(x)$, $i = 1, \ldots, n$, with respect to some common dominating measure. We want to model g_i by the family $f_{i,\boldsymbol{\theta}}(x)$, $\theta \in \Theta$, $i = 1, \ldots, n$. Thus, while the distributions might be different, they all share the same parameter $\boldsymbol{\theta}$. In this situation, the model density is different for each X_i, and we need to calculate the divergence between the data and the model separately for each point, $d_1(\widehat{g}_1, f_{1,\boldsymbol{\theta}}), \ldots, d_n(\widehat{g}_n, f_{n,\boldsymbol{\theta}})$ and to define

$$d_\tau(\widehat{g}_i, f_{i,\boldsymbol{\theta}}^{1+\tau}) = \int_{\mathcal{X}} f_{i,\boldsymbol{\theta}}^{1+\tau}(x)dx - \left(1 + \frac{1}{\tau}\right)\int_{\mathcal{X}} f_{i,\boldsymbol{\theta}}^{\tau}(x)\widehat{g}_i(x)dx + K,$$

where K is a constant that does not depend on $\boldsymbol{\theta}$. But in case we only had one data point X_i to estimate g_i, the best possibility is to assume that g_i is the distribution which puts their entire mass on X_i. Then we have,

$$d_\tau(\widehat{g}_i, f_{i,\boldsymbol{\theta}}^{1+\tau}) = \int_{\mathcal{X}} f_{i,\boldsymbol{\theta}}^{1+\tau}(x) - \left(1 + \frac{1}{\tau}\right) f_{i,\boldsymbol{\theta}}^{\tau}(x)dx + K,$$

and

$$\widehat{\boldsymbol{\theta}}_\tau = \arg\min_{\boldsymbol{\theta}\in\Theta} \boldsymbol{H}_{n,\tau}(\boldsymbol{\theta}),$$

with

$$\boldsymbol{H}_{n,\tau}(\boldsymbol{\theta}) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(-\log f_{i,\boldsymbol{\theta}}(x_i)) & \tau = 0 \\ \frac{1}{n}\sum_{i=1}^{n}\left[\int_{\mathcal{X}} f_{i,\boldsymbol{\theta}}^{1+\tau}(x_i))dx - \left(1+\frac{1}{\tau}\right) f_{i,\boldsymbol{\theta}}^{\tau}(x_i))\right] & \tau > 0 \end{cases}.$$

In this case, we can see that the asymptotic distribution is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(\mathbf{0}_p, \boldsymbol{\Psi}_\tau^{-1}(\boldsymbol{\theta})\boldsymbol{\Omega}_\tau(\boldsymbol{\theta})\boldsymbol{\Psi}_\tau^{-1}(\boldsymbol{\theta})), \tag{1.11}$$

where

$$\begin{aligned}\boldsymbol{\Psi}_\tau(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\int u_{i,\boldsymbol{\theta}}(x)u_{i,\boldsymbol{\theta}}^T(x)f_{i,\boldsymbol{\theta}}^{1+\tau}(x)dx\right. \\ &\qquad \left. - \int \{i_{i,\boldsymbol{\theta}}(x) + \tau u_{i,\boldsymbol{\theta}}(x)u_{i,\boldsymbol{\theta}}^T(x)\}\{g_i(x) - f_{i,\boldsymbol{\theta}}(x)\}f_{i,\boldsymbol{\theta}}^{\tau}(x)dx\right], \\ \boldsymbol{\Omega}_\tau(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\int u_{i,\boldsymbol{\theta}}(x)u_{i,\boldsymbol{\theta}}^T(x)f_{i,\boldsymbol{\theta}}^{2\tau}(x)g_i(x)dx - \xi_{i,\tau}(\boldsymbol{\theta})\xi_{i,\tau}^T(\boldsymbol{\theta})\right], \\ \boldsymbol{\xi}_{i,\tau}(\boldsymbol{\theta}) &= \int u_{i,\boldsymbol{\theta}}(x)f_{i,\boldsymbol{\theta}}^{\tau}(x)g_i(x)dx.\end{aligned}$$

If we assume that the true distribution g_i belongs to the model, i.e, $g_i = f_{i,\boldsymbol{\theta}}(x)$ for some $\boldsymbol{\theta}$, the matrices $\boldsymbol{\Psi}_\tau(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_\tau(\boldsymbol{\theta})$ are given by

$$\boldsymbol{\Psi}_\tau(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\int u_{i,\boldsymbol{\theta}}(x)u_{i,\boldsymbol{\theta}}^T(x)f_{i,\boldsymbol{\theta}}^{1+\tau}(x)dx\right]$$

and

$$\begin{aligned}\boldsymbol{\Omega}_\tau(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\int u_{i,\boldsymbol{\theta}}(x)u_{i,\boldsymbol{\theta}}^T(x)f_{i,\boldsymbol{\theta}}^{2\tau+1}(x)dx - \xi_{i,\tau}(\boldsymbol{\theta})\xi_{i,\tau}^T(\boldsymbol{\theta})\right], \\ \boldsymbol{\xi}_{i,\tau}(\boldsymbol{\theta}) &= \int u_{i,\boldsymbol{\theta}}(x)f_{i,\boldsymbol{\theta}}^{\tau+1}(x)dx.\end{aligned}$$

In this Thesis, the observations associated to the methods for one-shot devices are, as we will see in the next chapters, independent but not identically distributed. Therefore, the result in (1.11) will be very important. This result was considered in Basu et al. [2018], in order to define Wald-type tests for simple and composite null hypotheses with independent but non identically distributed observations.

1.3.2 Minimum ϕ-divergence estimators

In the procedure given to obtain the minimum DPD estimator is very important that the term depending at the same time of $f_{\boldsymbol{\theta}}(x)$ and $g(x)$ will be linear in $g(x)$. In that case we can estimate

$g(x)dx$ by $dG_n(x)$ where $G_n(x)$ is the empirical distribution function associated to a random sample of size n, $X_1, ..., X_n$. We can see that in the case of the DPD the term is

$$\int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{\tau}(x)\, g(x)\, dx.$$

If in the term depending of $f_{\boldsymbol{\theta}}(x)$ and $g(x)$, $g(x)$ does not appear in a linear way it is not possible to estimate that term using the empirical distribution function. This is the case, in general, for the phi-divergence measures. In this case, we can define the minimum ϕ-divergence estimator (MϕE) by

$$\widehat{\boldsymbol{\theta}}_{\phi} = \arg\min_{\theta\in\Theta} d_{\phi}(f_{\boldsymbol{\theta}}, \widehat{g}),$$

where $\widehat{g}$ is a non-parametric estimator of the density function g. This situation is more complicated. But the MϕE has been used in discrete models because in this case the estimator is a BAN (Best asymptotically Normal) estimator. We are going to describe it because it will be used in some part of this Thesis.

Let $(\mathcal{X}, \beta_{\mathcal{X}},\ P_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ be the statistical space associated with the random variable X, where $\beta_{\mathcal{X}}$ is the σ-field of Borel subsets $A \subset \mathcal{X}$ and $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ is a family of probability distributions defined on the measurable space $(\mathcal{X},\ \beta_{\mathcal{X}})$ with Θ an open subset of $\mathbb{R}^{M_0}$, $M_0 \geq 1$. Let $\mathcal{P} = \{E_i\}_{i=1,...,M}$ be a partition of $\mathcal{X}$. The formula $\Pr_{\boldsymbol{\theta}}(E_i) = p_i(\boldsymbol{\theta})$, $i = 1, \ldots, M$, defines a discrete statistical model. Let $Y_1, \ldots, Y_n$ be a random sample from the population described by the random variable X, let $N_i = \sum_{j=1}^{n} I_{E_i}(Y_j)$ and $\widehat{p}_i = N_i/n$, $i = 1, \ldots, M$. Estimating $\boldsymbol{\theta}$ by MLE method, under the discrete statistical model, consists of maximizing for fixed $n_1, \ldots, n_M$,

$$\Pr{}_{\boldsymbol{\theta}}(N_1 = n_1, \ldots, N_M = n_M) = \frac{n!}{n_1! \ldots n_M!} p_1^{n_1}(\boldsymbol{\theta}) \times \ldots \times p_M^{n_M}(\boldsymbol{\theta}) \tag{1.12}$$

or, equivalently,

$$\log \Pr{}_{\boldsymbol{\theta}}(N_1 = n_1, \ldots, N_M = n_M) = -n d_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) + k \tag{1.13}$$

where $\widehat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_M)^T$, $\boldsymbol{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \ldots, p_M(\boldsymbol{\theta}))^T$ and k is independent of $\boldsymbol{\theta}$. Then, estimating $\boldsymbol{\theta}$ with the MLE of the discrete model is equivalent to minimizing the Kullback-Leibler divergence on $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{M_0}$. Since Kullback-Leibler divergence is a particular case of the ϕ-divergence measures, we can choose as estimator of $\boldsymbol{\theta}$ the value $\widehat{\boldsymbol{\theta}}_{\phi}$ verifying

$$d_{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}_{\phi})) = \inf_{\boldsymbol{\theta}\in\Theta\subseteq R^{M_0}} d_{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})),$$

where

$$d_{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = \sum_{i=1}^{M} \widehat{p}_i \phi\left(\frac{\widehat{p}_i}{p_i(\boldsymbol{\theta})}\right). \tag{1.14}$$

In general we can assume that there exists a function $\boldsymbol{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \ldots, p_M(\boldsymbol{\theta}))^T$ that maps each $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{M_0})^T$ into a point in

$$\triangle_M = \left\{\boldsymbol{p} = (p_1, ..., p_M)^T : p_i \geq 0, i = 1, ..., M, \sum_{i=1}^{M} p_i = 1\right\}.$$

As $\boldsymbol{\theta}$ ranges over the values of Θ, $\boldsymbol{p}(\boldsymbol{\theta})$ ranges over a subset T of Δ_M. When we assume that a given model is "correct", we are just assuming that there exists a value $\boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{p}(\boldsymbol{\theta}_0) = \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the true value of the multinomial probability, i.e., $\boldsymbol{\pi} \in T$.

Morales et al. [1995] established that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\phi} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_p, \boldsymbol{I}_F^{-1}(\boldsymbol{\theta}))$$

where $\boldsymbol{I}_F(\boldsymbol{\theta})$ is the Fisher information matrix defined by

$$\boldsymbol{I}_F(\boldsymbol{\theta}) = \left(\sum_{j=1}^{M} \frac{1}{p_j(\boldsymbol{\theta})} \frac{\partial p_j(\boldsymbol{\theta})}{\partial \theta_r} \frac{\partial p_j(\boldsymbol{\theta})}{\partial \theta_s} \right)_{\substack{r=1,\dots,M_0 \\ s=1,\dots,M_0}} .$$

1.4 One-shot devices

The reliability of a product, system, weapon, or piece of equipment can be defined as the ability of the device to perform as designed for, or, more simply, as the probability that the device does not fail when used. Engineers assess reliability by repeatedly testing the device and observing its failure rate. Certain products, called "one-shot" devices, make this approach challenging. One-shot devices can be used only once and after use the device is either destroyed or must be rebuilt. In this section, we introduce some basic concepts to understand this kind of devices.

1.4.1 Accelerated life tests

Most manufactured products are of high quality these days and so they will usually have long lifetimes. Consequently, if the products are tested under normal conditions, the failure times of the products will be quite large resulting in a large testing time. To reduce the experimental time and cost, therefore, accelerated life tests (ALTs) are commonly employed to evaluate the reliability of such products. An ALT shortens the life span of the products by increasing the levels of stress factors, such as temperature and humidity. After estimating the parameters from data collected under high-stress conditions, one usually extrapolates the life characteristics, such as mean lifetime and failure rates, from high stress conditions to normal operating conditions; see Meeter and Meeker [1994] and Meeker et al. [1998]. Some common stress factors that could be used for this purpose include air pressure, temperature, humidity and voltage which can be controlled easily in a laboratory setup.

1.4.2 Life-stress relationships

In ALTs, failure rate (the expected number of failures per unit time) is required to relate to stress factors such that measurements taken during the experiment can then be extrapolated back to the expected performance under normal operating conditions. A simple model, such a linear model, may be not enough to describe the relationship between the failure rate and the stress factors. The most common relationships are described in Meeker [1984], Wang and Kececioglu [2000] and Pascual [2007]. We consider, without loss of generality, the case of one risk, and we are going to clarify some of that relationships:

a) The Arrhenius Relationship for Temperature Acceleration

Let T be the applied temperature in degrees Celsius. Note that $T+273.15$ is the temperature on the Kelvin scale. Under the Arrhenius relationship, based on the Arrhenius Law, the failure rate λ_T relates to the temperature T in the form

$$\lambda_T = \exp\left(\gamma_0 + \gamma_1 \frac{11605}{T+273.15}\right),$$

where, γ_0 and γ_1 are derived from physical properties of the product being tested, and the test methods. For example, is the activation energy of the chemical reaction rate. The Arrhenius model is used to describe the failure time of lubricants, light-emitting diodes, insulating tapes, and bulb filaments (Pascual [2007]).

b) The Inverse Power Relationship for Voltage Acceleration

It is useful for describing the lifetime as a function of applied voltage. The parameter λ_V relates to applied voltage V as follows

$$\lambda_V = \frac{1}{\gamma_0 V^{\gamma_1}}.$$

c) Log-linear relationship

Commonly used in survival analysis, it is often used in practice due to its mathematical convenience. It shows the relative importance of stress factors in influencing the failure behavior, regardless of whether the model is correct or not. The parameter λ_x relates to a stress factor x in this case in the form

$$\lambda_x = \exp(\gamma_0 + \gamma_1 x).$$

Lifetime distribution models with the log-linear relationships with covariates will be the kind of relationship considered in the following chapters, where more than one stress factor will be also considered.

d) Other relationships

As pointed out by Meeker [1984], assumed life-time relationships will hold only over a limited range of stress levels. Therefore, there will typically a limit on the highest level of stress, denoted here by x_H. Although a higher limit will provide more precision for estimators, it can also cause serious bias, if the value is so high that the associated model is incorrect. The choice of the limit is usually made on the basis of previous results or the engineering (or biomedical) judgment.

d.1) Standardization Define the standardized stress as

$$\xi = \frac{x - x_D}{x_H - x_D},$$

such that $x_U = 0 \leq x$ and $x \leq x_H = 1$. x_U and x_H are the standardized use level, and upper bound, respectively. The standardized model is

$$\lambda_x = \beta_0 + \beta_1 \xi.$$

d.2) Quadratic model As a generalization of the log-linear relationship, it might be tempting to fit the quadratic model (given here in terms of ξ)

$$\lambda_x = \beta_0 + \beta_1 \xi + \beta_2 \xi^2.$$

However, there is general reluctance to use such a model if the desired inferences require much extrapolation (Little and Jebe [1975]).

1.4.3 Types of censoring

In any kind of data analysis, more information will result in a more complete analysis. In the case of lifetime data analysis, knowing the exact time of failure of the devices will be the most preferable situation. However, this is not possible in most of the practical cases, due to constraints on time and budget of the experiment. Incomplete data frequently arise in life-tests and are referred to as censored data. We can distinguish between left, right and interval censoring:

a) **Left censoring**

If the device under test fails before observable time, we define that to be left censored. Typical example includes failure of an alarm system to alert in an incident of fire. The left censoring is encountered on rare occasions in survival analysis because investigators are very particular in the selection of participants for the study.

b) **Right censoring**

On the other hand, if the study finished but the event of the interest was not observed, the device under test is said to be right censored. For instance, clothing or mobile phones are often replaced by new ones when they are still usable, and cars are sent to junkyard when they are still drivable. An example in survival analysis might be a clinical trial to study the defect of treatments on heart attack occurrence. The study ends after 10 years. Those patients who have had no heart attacks by the end of the last year are censored. Due to constraints on time and cost, right censoring is commonly encountered in life-testing experiments. Under this form of censoring, while some lifetimes will be completely observed, others will be known only to be beyond some times. The two main types of right censoring are as follows:

b.1) Type-I censoring The life-test is terminated at a pre-fixed time, resulting in a fixed censoring time and a random number of failures during the experimental period. Environmental data are almost always Type-I censored.

b.2) Type-II censoring The life-test is terminated as soon as a pre-fixed number of failures have been observed, resulting in a random censoring time and a fixed number of failures during the experimental period. Type-II censored samples are also known as failure-censored samples (Nelson [1980]).

The problem with Type-I censoring (random failures) is that, not guarantying a minimum number of failures may lead to ineffective inference. This problem is solved with Type-II censoring. However, the first censoring can be easier from managerial point of view since the duration of the test is known in advance.

c) **Interval censoring**

In some situations, test items are inspected for failure at many time points and one only knows that items failed in some intervals between two contiguous inspections. Such lifetime data are said to be interval censored and arise naturally when the test items are not constantly monitored. Any observation of a continuous random variable could be considered interval censored, because its value is reported to a few decimal places. This sort of fine-scale interval censoring is usually ignored and the values are treated as exactly observed.

Left and right censoring may cause floor and ceiling effects, while rounding data to fewer decimal places results in interval-censored data. It is important to distinguish between censoring and truncation. Censored values are those reported as less or greater than some value, or as an interval. On the other hand, truncated values are those that are not reported if exceeds some limit, but the truncated point is not recorded at all. Therefore, truncated data are less informative that censored data.

1.4.4 One shot device testing data

In this thesis, one-shot device testing data, which is an extreme case of interval censoring, is studied. Since one-shot devices can be used only once and are destroyed immediately after use, one can only know whether the failure time is either before or after a specific time. The lifetimes are either left- or right-censored, with the lifetime being less than the inspection time if the test outcome is a failure (resulting in left censoring) and the lifetime being more than the inspection time if the test outcome is a success (resulting in right censoring). Some examples of one-shot devices are nuclear weapons, space shuttles, automobile air bags, fuel injectors, disposable napkins, heat detectors, missiles (Olwell and Sorell [2001]) and fire extinguishers (Newby [2008]). In survival analysis, these data are called "current status data". For instance, in animal carcinogenicity experiments, one observes whether a tumor has occurred by the examination time for each subject.

Due to the advances in manufacturing design and technology, products have now become highly reliable with long lifetimes. This fact would pose a problem in the analysis of data if only few or no failures are observed. For this reason, accelerated life tests are often used by adjusting a controllable factor such as temperature in order to induce more failures in the experiment. The study of one-shot device from ALT data has been developed considerably recently, mainly motivated by the work of Fan et al. [2009]. In that paper, a Bayesian approach was presented to develop inference on the failure rate and reliability of devices. They found the normal prior to be the best one when the failure observations are rare, that is, when the devices are highly reliable. Balakrishnan and Ling [2012a] developed an expectation-maximization (EM) algorithm for the determination of the maximum likelihood estimator (MLE) of model parameters under exponential lifetime distribution for devices with a single stress factor. Balakrishnan and Ling [2012b] further extended their work to a model with multiple stress factors. Balakrishnan and Ling [2013] developed more general inferential results for devices with Weibull lifetimes under non-constant shape parameters, while Balakrishnan and Ling [2014a] provided inferential work for devices with gamma lifetimes. In Balakrishnan et al. [2015a,b] the problem of one-shot devices under competing risks was considered for the first time. All these results are recorded in the thesis of Ling [2012] and So [2016].

1.5 Scope of the Thesis

Most of the above results are based on MLE, which is well-known to be efficient, but also non-robust. Therefore, testing procedures based on MLE face serious robustness problems. The main scope of this Thesis is to develop robust estimators and test statistics (based on the DPD measures) for one-shot device testing data. All the theoretical results will be supported by simulation studies and illustrated by numerical examples.

The Thesis proceeds as follows. In Chapter 2, we assume the one-shot devices with lifetimes having exponential distribution with a single-stress relationship. In Chapter 3, the exponential distribution with multiple-stress relationship is considered, generalizing the results in Chapter 2. Next, in Chapter 4 and Chapter 5, we consider the situation when lifetimes follow, respectively, a Gamma and a Weibull distribution with non-constant shape parameters. In Chapter 6, similar procedures are applied to other distribution functions, such as Lindley and lognormal. Chapter 7 develop robust inference for one-shot device testing under the proportional hazards assumption and, in Chapter 8, we consider the competing risk model (assuming different possible causes of failure) with exponential lifetimes. Chapter 9 summarizes the main results of previous chapters and gives some ideas about future work. The Appendixes A and B briefly present some other results, which have also been obtained by the candidate during her Ph.D. studies.

Chapter 2

Robust inference for one-shot device testing under exponential distribution with a simple stress factor

2.1 Introduction

Let us consider the problem of one-shot device testing along with an accelerating factor, in which the failure time of the devices is assumed to follow an exponential distribution. In this context, Rodrigues et al. [1993] presented two approaches based on the likelihood ratio statistics for comparing exponential accelerated life models. Fan et al. [2009] considered a single stress factor to the accelerated life test plan for one-shot devices, and analyzed the data by using a Bayesian approach in which the model parameters in the prior information were assumed to be close to the true values. In contrast, Balakrishnan and Ling [2012a] developed an EM algorithm for a single stress model, and made a comparative study with the mentioned Bayesian approach, showing that the EM method is more appropriate for moderately and lowly reliable products. Finally, Chimitova and Balakrishnan [2015] made a comparison of several goodness-of-fit tests for one-shot device testing data.

In this chapter we develop robust estimators and statistics for one-shot device testing under the exponential distribution with a simple stress factor. In Section 2.2, we present a description of the one-shot device model as well as the MLE of the model parameters. Section 2.3 develops the weighted minimum DPD estimator as a natural extension of the MLE, as well as its asymptotic distribution. In Section 2.4, Z-type test statistics are introduced for testing some hypotheses about the parameters of the one-shot device model. The Influence Function of proposed estimators and test statistics is developed in Section 2.5. In Section 2.6, an extensive simulation study is presented in order to empirically illustrate the robustness of the weighted minimum DPD estimators, as well as the Z-type test introduced earlier. A data-driven choice procedure of the optimal tuning parameter given a data set is provided in also provided in Section 2.6. Finally, some numerical examples are presented in Section 2.7, with one of them relating to a reliability situation and the other two are from real applications to tumorigenicity experiments.

The results of this Chapter have been published in the form of a paper (Balakrishnan et al. [2019b]).

2.2 Model description and MLE

Consider a reliability testing experiment in which the devices are stratified into I testing conditions and, in the i-th testing condition K_i units or devices are tested under some stress factor (say, for example, temperature) x_i, $i = 1, \dots, I$. In the i-th test group, the number of failures, n_i, is collected. This setting is summarized in Table 2.2.1.

Table 2.2.1: Data on one-shot devices testing at a simple stress level and collected at different inspection times

Condition	Inspection Time	Devices	Failures	Temperature
1	IT_1	K_1	n_1	x_1
2	IT_2	K_2	n_2	x_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	IT_I	K_I	n_I	x_I

Let us denote the random variable for the failure time under condition i as T_{ik}, for $i = 1, \ldots, I$ and $k = 1, \ldots, K_i$, respectively. We shall assume here, that the true lifetime T_{ik} follows an exponential distribution with unknown failure rate, $\lambda_i(\boldsymbol{\theta})$, related to the stress factor x_i in loglinear form as

$$\lambda_i(\boldsymbol{\theta}) = \theta_0 \exp(\theta_1 x_i),$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$ is the model parameter vector, $\boldsymbol{\theta} \in \Theta = \mathbb{R}^+ \times \mathbb{R}$. Therefore, the corresponding density function and distribution function of the failure time under condition i are, respectively,

$$f(t; x_i, \boldsymbol{\theta}) = \lambda_i(\boldsymbol{\theta}) \exp\{-\lambda_i(\boldsymbol{\theta})t\} = \theta_0 \exp(\theta_1 x_i) \exp\{-\theta_0 \exp(\theta_1 x_i)t\}$$

and

$$F(t; x_i, \boldsymbol{\theta}) = 1 - \exp\{-\lambda_i(\boldsymbol{\theta})t\} = 1 - exp\{-\theta_0 \exp(\theta_1 x_i)\}. \tag{2.1}$$

Let us denote by $R(t; x_i, \boldsymbol{\theta}) = 1 - F(t; x_i, \boldsymbol{\theta})$ the reliability function, the probability that a unit lasts lifetime t. Assuming independent observations, the likelihood function, based on the observed data as in Table 2.2.1, is given by

$$\mathcal{L}(n_1, \ldots, n_I; \boldsymbol{\theta}) = \prod_{i=1}^{I} F^{n_i}(IT_i; x_i, \boldsymbol{\theta})\, R^{K_i - n_i}(IT_i; x_i, \boldsymbol{\theta}). \tag{2.2}$$

Definition 2.1 *The MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_0, \widehat{\theta}_1)^T$, is given by*

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \ \log \mathcal{L}(n_1, \ldots, n_I; \boldsymbol{\theta}), \tag{2.3}$$

where $\mathcal{L}(n_1, \ldots, n_I; \boldsymbol{\theta})$ was given in (2.2).

We now introduce empirical and theoretical probability vectors, respectively,

$$\widehat{\boldsymbol{p}}_i = (\widehat{p}_{i1}, \widehat{p}_{i2})^T, \quad i = 1, \ldots, I, \tag{2.4}$$

and

$$\boldsymbol{\pi}_i(\boldsymbol{\theta}) = (\pi_{i1}(\boldsymbol{\theta}), \pi_{i2}(\boldsymbol{\theta}))^T, \quad i = 1, \ldots, I, \tag{2.5}$$

with $\widehat{p}_{i1} = \frac{n_i}{K_i}$, $\widehat{p}_{i2} = 1 - \frac{n_i}{K_i}$, $\pi_{i1}(\boldsymbol{\theta}) = F(IT_i; x_i, \boldsymbol{\theta})$ and $\pi_{i2}(\boldsymbol{\theta}) = R(IT_i; x_i, \boldsymbol{\theta})$.

The Kullback-Leibler divergence measure between $\widehat{\boldsymbol{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\theta})$ is given by

$$d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \widehat{p}_{i1} \log\left(\frac{\widehat{p}_{i1}}{\pi_{i1}(\boldsymbol{\theta})}\right) + \widehat{p}_{i2} \log\left(\frac{\widehat{p}_{i2}}{\pi_{i2}(\boldsymbol{\theta})}\right)$$

and similarly the weighted Kullback-Leibler divergence measure of all the units, where $K = \sum_{i=1}^{I} K_i$ is the total number of devices, is given by

$$\sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \frac{1}{K} \sum_{i=1}^{I} K_i \left[\widehat{p}_{i1} \log\left(\frac{\widehat{p}_{i1}}{\pi_{i1}(\boldsymbol{\theta})}\right) + \widehat{p}_{i2} \log\left(\frac{\widehat{p}_{i2}}{\pi_{i2}(\boldsymbol{\theta})}\right)\right]. \tag{2.6}$$

For more details about the Kullback-Leibler divergence measure, see Pardo [2005].

Theorem 2.2 *The likelihood function $\mathcal{L}(n_1, \ldots, n_I; \boldsymbol{\theta})$, given in (2.2), is related to the weighted Kullback-Leibler divergence measure through*

$$\sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = c - \frac{1}{K} \log \mathcal{L}(n_1, \ldots, n_I; \boldsymbol{\theta}),$$

with c being a constant not dependent on $\boldsymbol{\theta}$.

Proof. We have,

$$\begin{aligned}
\sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \sum_{i=1}^{I} \frac{n_i}{K} \log \left(\frac{\frac{n_i}{K_i}}{F(IT_i; x_i, \boldsymbol{\theta})} \right) + \frac{K_i - n_i}{K} \log \left(\frac{\frac{K_i - n_i}{K_i}}{1 - F(IT_i; x_i, \boldsymbol{\theta})} \right) \\
&= c - \frac{1}{K} \sum_{i=1}^{I} \left\{ n_i \log \left(F(IT_i; x_i, \boldsymbol{\theta}) \right) + (K_i - n_i) \log \left(R(IT_i; x_i, \boldsymbol{\theta}) \right) \right\} \\
&= c - \frac{1}{K} \log \left(\prod_{i=1}^{I} F^{n_i}(IT_i; x_i, \boldsymbol{\theta}) R^{K_i - n_i}(IT_i; x_i, \boldsymbol{\theta}) \right) \\
&= c - \frac{1}{K} \log \mathcal{L}(n_1, \ldots, n_I; \boldsymbol{\theta}),
\end{aligned}$$

where $c = \sum_{i=1}^{I} \frac{n_i}{K_i} \log \left(\frac{n_i}{K_i} \right) + \frac{K_i - n_i}{K_i} \log \left(\frac{K_i - n_i}{K_i} \right)$ does not depend on the parameter $\boldsymbol{\theta}$. ■

Based on Theorem 2.2, we have the following alternative definition for the MLE of $\boldsymbol{\theta}$.

Definition 2.3 *The MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, can be also defined as*

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})). \tag{2.7}$$

2.3 Weighted minimum DPD estimator

Based on expression (2.7), we could think of defining an estimator by minimizing any (weighted) divergence measure between the empirical and theoretical probability vectors. As explained in the Introduction, there are many different divergence measures (or distances) known in the literature, and the natural question is whether all of them are valid to define estimators with good properties. Initially, the answer is yes, but we must think in terms of efficiency as well as robustness of the defined estimators. From an asymptotic point of view, it is well-known that the MLE is a BAN (Best Asymptotically Normal) estimator, but at the same time we know that the MLE has a very poor behavior, in general, with regard to robustness. It is well-known that a gain in robustness leads to a loss in efficiency. Therefore, the distances (divergence measures) that we must use are those which result in estimators having good properties in terms of robustness with low loss of efficiency. The DPD measure introduced by Basu et al. [1998] has the required properties and has been studied for many different statistical problems until now.

Given the probability vectors $\widehat{\boldsymbol{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\theta})$ defined in (2.4) and (2.5), respectively, the DPD between the two probability vectors, with tuning parameter $\beta \geq 0$, is given by

$$\begin{aligned}
d_\beta(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \left(\pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta+1}(\boldsymbol{\theta}) \right) - \frac{\beta+1}{\beta} \left(\widehat{p}_{i1} \pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2} \pi_{i2}^{\beta}(\boldsymbol{\theta}) \right) \\
&\quad + \frac{1}{\beta} \left(\widehat{p}_{i1}^{\beta+1} + \widehat{p}_{i2}^{\beta+1} \right), \quad \text{if } \beta > 0,
\end{aligned} \tag{2.8}$$

and $d_{\beta=0}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \lim_{\beta \to 0^+} d_\beta(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta}))$, if $\beta = 0$.

We observe that in (2.8), the term $\frac{1}{\beta}\left(\widehat{p}_{i1}^{\beta+1}+\widehat{p}_{i2}^{\beta+1}\right)$ has no role in the minimization with respect to $\boldsymbol{\theta}$. Therefore, we can consider the equivalent measure

$$d_{\beta}^{*}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \left(\pi_{i1}^{\beta+1}(\boldsymbol{a}) + \pi_{i2}^{\beta+1}(\boldsymbol{\theta})\right) - \frac{\beta+1}{\beta}\left(\widehat{p}_{i1}\pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2}\pi_{i2}^{\beta}(\boldsymbol{\theta})\right). \tag{2.9}$$

Definition 2.4 *Based on (2.7) and (2.9), we can define the weighted minimum DPD estimator for $\boldsymbol{\theta}$ as*

$$\widehat{\boldsymbol{\theta}}_{\beta} = \arg\min_{\boldsymbol{\theta}\in\Theta}\sum_{i=1}^{I}\frac{K_i}{K}d_{\beta}^{*}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})), \quad \text{for } \beta > 0,$$

and, in particular, for $\beta = 0$, we have the MLE.

Theorem 2.5 *The weighted minimum DPD estimator of $\boldsymbol{\theta}$ with tuning parameter $\beta \geq 0$, $\widehat{\boldsymbol{\theta}}_{\beta}$, can be obtained as the solution of the following system of equations:*

$$\sum_{i=1}^{I}(K_i\ F(IT_i; x_i, \boldsymbol{\theta}) - n_i)\, f(IT_i; x_i, \boldsymbol{\theta})\left[F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\right] IT_i = 0, \tag{2.10}$$

$$\sum_{i=1}^{I}(K_i\ F(IT_i; x_i, \boldsymbol{\theta}) - n_i)\, f(IT_i; x_i, \boldsymbol{\theta})\left[F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\right] IT_i x_i = 0. \tag{2.11}$$

Proof. We have

$$\frac{\partial F(IT_i; x_i, \boldsymbol{\theta})}{\partial\theta_0} = \exp\left\{-\theta_0\exp\left(\theta_1 x_i\right) IT_i\right\}\exp\left\{\theta_1 x_i\right\} IT_i = f(IT_i; x_i, \boldsymbol{\theta})\frac{IT_i}{\theta_0} \tag{2.12}$$

and

$$\frac{\partial F(IT_i; x_i, \boldsymbol{\theta})}{\partial\theta_1} = \exp\left\{-\theta_0\exp\left(\theta_1 x_i\right) IT_i\right\}\exp\left\{\theta_1 x_i\right\} IT_i\theta_0 x_i = f(IT_i; x_i, \boldsymbol{\theta}) IT_i x_i. \tag{2.13}$$

We denote

$$d_{\beta}^{*}\left(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})\right) = \mathcal{T}_{i1,\beta}(\boldsymbol{\theta}) + \mathcal{T}_{i2,\beta}(\boldsymbol{\theta}),$$

where $\mathcal{T}_{i1,\beta}(\boldsymbol{\theta})$ and $\mathcal{T}_{i2,\beta}(\boldsymbol{\theta})$ are as follows, for $\beta > 0$:

$$\mathcal{T}_{i1,\beta}(\boldsymbol{\theta}) = \left\{F^{\beta+1}(IT_i; x_i, \boldsymbol{\theta}) - (1+\tfrac{1}{\beta})F^{\beta}(IT_i; x_i, \boldsymbol{\theta})\frac{n_i}{K_i}\right\},$$

$$\mathcal{T}_{i2,\beta}(\boldsymbol{\theta}) = \left\{R^{\beta+1}(IT_i; x_i, \boldsymbol{\theta}) - (1+\tfrac{1}{\beta})R^{\beta}(IT_i; x_i, \boldsymbol{\theta})\frac{K_i - n_i}{K_i}\right\}.$$

Based on (2.12), we have

$$\frac{\partial\mathcal{T}_{i1,\beta}(\boldsymbol{\theta})}{\partial\theta_0} = (\beta+1)\left(F(IT_i; x_i, \boldsymbol{\theta}) - \frac{n_i}{K_i}\right) f(IT_i; x_i, \boldsymbol{\theta})F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\frac{IT_i}{\theta_0}$$

and

$$\frac{\partial\mathcal{T}_{i2,\beta}(\boldsymbol{\theta})}{\partial\theta_0} = (\beta+1)\left(F(IT_i; x_i, \boldsymbol{\theta}) - \frac{n_i}{K_i}\right) f(IT_i; x_i, \boldsymbol{\theta})R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\frac{IT_i}{\theta_0}.$$

On the other hand, by (2.13), we have

$$\frac{\partial\mathcal{T}_{i1,\beta}(\boldsymbol{\theta})}{\partial\theta_1} = (\beta+1)\left(F(IT_i; x_i, \boldsymbol{\theta}) - \frac{n_i}{K_i}\right) f(IT_i; x_i, \boldsymbol{\theta})F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})IT_i x_i$$

and

$$\frac{\partial\mathcal{T}_{i2,\beta}(\boldsymbol{\theta})}{\partial\theta_1} = (\beta+1)\left(F(IT_i; x_i, \boldsymbol{\theta}) - \frac{n_i}{K_i}\right) f(IT_i; x_i, \boldsymbol{\theta})R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})IT_i x_i.$$

Finally, the system of equations is given by

$$\sum_{i=1}^{I} \frac{K_i}{K}\left(\frac{\partial \mathcal{T}_{i1,\beta}(\boldsymbol{\theta})}{\partial \theta_0} + \frac{\partial \mathcal{T}_{i2,\beta}(\boldsymbol{\theta})}{\partial \theta_0}\right) = 0,$$
$$\sum_{i=1}^{I} \frac{K_i}{K}\left(\frac{\partial \mathcal{T}_{i1,\beta}(\boldsymbol{\theta})}{\partial \theta_1} + \frac{\partial \mathcal{T}_{i2,\beta}(\boldsymbol{\theta})}{\partial \theta_1}\right) = 0.$$

If we consider $\beta = 0$, we get the system needed to solve to get the MLE. Hence, the previous system of equations is valid not only for tuning parameters $\beta > 0$, but also for $\beta = 0$. ■

Theorem 2.6 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter $\boldsymbol{\theta}$. Then, the asymptotic distribution of the weighted minimum DPD estimator $\widehat{\boldsymbol{\theta}}_\beta$ is given by*

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \xrightarrow[K\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K}\begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta})\left[F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\right], \tag{2.14}$$

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K}\begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta}) \tag{2.15}$$
$$\times F(IT_i; x_i, \boldsymbol{\theta}) R(IT_i; x_i, \boldsymbol{\theta})\left[F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\right]^2.$$

Proof. From Ghosh et al. [2013], it is known that

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \xrightarrow[K\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_2, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{j=1}^{2} \frac{K_i}{K} \boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta})\pi_{ij}^{\beta+1}(\boldsymbol{\theta}),$$
$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{j=1}^{2} \frac{K_i}{K} \boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta})\pi_{ij}^{2\beta+1}(\boldsymbol{\theta}) - \sum_{i=1}^{I} \frac{K_i}{K}\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})\boldsymbol{\xi}_{i,\beta}^T(\boldsymbol{\theta}),$$

with

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta}) = \sum_{j=1}^{2} \boldsymbol{u}_{ij}(\boldsymbol{\theta})\pi_{ij}^{\beta+1}(\boldsymbol{\theta}),$$
$$\boldsymbol{u}_{ij}(\boldsymbol{\theta}) = \frac{\partial \log \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\pi_{ij}(\boldsymbol{\theta})}\frac{\partial \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (-1)^{j+1}\frac{f(IT_i; x_i, \boldsymbol{\theta})IT_i}{\pi_{ij}(\boldsymbol{\theta})}\begin{pmatrix} \frac{1}{\theta_0} \\ x_i \end{pmatrix}.$$

Since $\boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta}) = \frac{f^2(IT_i;x_i,\boldsymbol{\theta})IT_i^2}{\pi_{ij}^2(\boldsymbol{\theta})}\begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix}$, we have

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K}\begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta})\sum_{j=1}^{2}\pi_{ij}^{\beta-1}(\boldsymbol{\theta})$$
$$= \sum_{i=1}^{I} \frac{K_i}{K}\begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta})\left[F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta})\right].$$

In a similar manner

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})\boldsymbol{\xi}_{i,\beta}^T(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta}) \left(\sum_{j=1}^{2} (-1)^{j+1} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \right)^2$$

and

$$\boldsymbol{K}_{\beta}(\boldsymbol{\theta}) = \sum_{i=1}^{I} \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta}) \left[\sum_{j=1}^{2} \pi_{ij}^{2\beta-1}(\boldsymbol{\theta}) - \left(\sum_{j=1}^{2} (-1)^{j+1} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \right)^2 \right].$$

Since

$$\sum_{j=1}^{2} \pi_{ij}^{2\beta-1}(\boldsymbol{\theta}) - \left(\sum_{j=1}^{2} (-1)^{j+1} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \right)^2 = \pi_{i1}(\boldsymbol{\theta})\pi_{i2}(\boldsymbol{\theta}) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right)^2,$$

it holds

$$\begin{aligned} \boldsymbol{K}_{\beta}(\boldsymbol{\theta}) &= \sum_{i=1}^{I} \frac{K_i}{K} \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta}) \pi_{i1}(\boldsymbol{\theta})\pi_{i2}(\boldsymbol{\theta}) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right)^2 \\ &= \sum_{i=1}^{I} \frac{K_i}{K} \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix} IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta}) \\ &\quad \times F(IT_i; x_i, \boldsymbol{\theta}) R(IT_i; x_i, \boldsymbol{\theta}) \left[F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) \right]^2. \end{aligned}$$

■

2.4 Robust Z-type tests

We are now interested in testing the null hypothesis of a linear combination of $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$, H_0: $m_0\theta_0 + m_1\theta_1 = d$, or equivalently

$$H_0\colon\ \boldsymbol{m}^T\boldsymbol{\theta} = d, \tag{2.16}$$

where $\boldsymbol{m}^T = (m_0, m_1)$. In this setting, it is important to know the asymptotic distribution of the weighted minimum DPD estimator of $\boldsymbol{\theta}$. In particular, in case we wish to test if the different temperatures do not affect the model of the one-shot devices, we need to consider $\boldsymbol{m}^T = (m_0, m_1) = (0, 1)$ and $d = 0$. In the following definition, we present Z-type test statistics based on $\widehat{\boldsymbol{\theta}}_{\beta}$.

Definition 2.7 *Let $\widehat{\boldsymbol{\theta}}_{\beta} = (\widehat{\theta}_{0,\beta}, \widehat{\theta}_{1,\beta})^T$ be the weighted minimum DPD estimator of $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$. Then, the family of Z-type test statistics for testing (2.16) is given by*

$$Z_K(\widehat{\boldsymbol{\theta}}_{\beta}) = \sqrt{\frac{K}{\boldsymbol{m}^T \boldsymbol{J}_{\beta}^{-1}(\widehat{\boldsymbol{\theta}}_{\beta}) \boldsymbol{K}_{\beta}(\widehat{\boldsymbol{\theta}}_{\beta}) \boldsymbol{J}_{\beta}^{-1}(\widehat{\boldsymbol{\theta}}_{\beta}) \boldsymbol{m}}} \left(\boldsymbol{m}^T \widehat{\boldsymbol{\theta}}_{\beta} - d \right). \tag{2.17}$$

In the following theorem, the asymptotic distribution of $Z_K(\widehat{\boldsymbol{\theta}}_{\beta})$ is presented.

Theorem 2.8 *The asymptotic distribution of Z-type test statistics, $Z_K(\widehat{\boldsymbol{\theta}}_{\beta})$, defined in (2.17), is a standard normal.*

Proof. Let $\boldsymbol{\theta}^0$ be the true value of parameter $\boldsymbol{\theta}$. It is clear that under (2.16), we have

$$\boldsymbol{m}^T \widehat{\boldsymbol{\theta}}_{\beta} - d = \boldsymbol{m}^T (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^0)$$

and we know that

$$\sqrt{K}(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^0) \xrightarrow[K\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{J}_{\beta}^{-1}(\boldsymbol{\theta}^0) \boldsymbol{K}_{\beta}(\boldsymbol{\theta}^0) \boldsymbol{J}_{\beta}^{-1}(\boldsymbol{\theta}^0) \right),$$

from which it follows that

$$\sqrt{K}\left(\boldsymbol{m}^T\widehat{\boldsymbol{\theta}}_\beta - d\right) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, \boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{m}\right).$$

Dividing the left hand side by

$$\sqrt{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}},$$

since $\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}$ is a consistent estimator of $\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{m}$, the desired result is obtained by following the Slutsky's theorem. ■

Based on Theorem 2.8, the null hypothesis in (2.16) will be rejected, with significance level α, if

$$\left|Z_K(\widehat{\boldsymbol{\theta}}_\beta)\right| > z_{\frac{\alpha}{2}}, \tag{2.18}$$

where $z_{\frac{\alpha}{2}}$ is a right hand side quantile of order $\frac{\alpha}{2}$ of a normal distribution. Now, we shall present a result providing an asymptotic approximation, to the power function, for the test statistic defined in (2.18).

Theorem 2.9 *Let $\boldsymbol{\theta}^* \in \Theta$ be the true value of the parameter $\boldsymbol{\theta}$ so that*

$$\widehat{\boldsymbol{\theta}}_\beta \underset{K\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \boldsymbol{\theta}^* \in \Theta$$

and $\boldsymbol{m}^T\boldsymbol{\theta}^* \neq d$. Then, the approximate power function of the test statistic in (2.18) at $\boldsymbol{\theta}^*$ is as given below, where $\Phi(\cdot)$ is the standard normal distribution function,

$$\pi\left(\boldsymbol{\theta}^*\right) \simeq 2\left[1 - \Phi\left(z_{\frac{\alpha}{2}} - \sqrt{\frac{K}{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\boldsymbol{K}_\beta(\boldsymbol{\theta}^*)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\boldsymbol{m}}}(\boldsymbol{m}^T\boldsymbol{\theta}^* - d)\right)\right]. \tag{2.19}$$

Proof. The power function of $Z_K(\widehat{\boldsymbol{\theta}}_\beta)$ at $\boldsymbol{\theta}^*$ can be obtained as follows:

$$\begin{aligned}
\pi\left(\boldsymbol{\theta}^*\right) &= \Pr\left(\left|Z_K(\widehat{\boldsymbol{\theta}}_\beta)\right| > z_{\frac{\alpha}{2}} | \boldsymbol{\theta} = \boldsymbol{\theta}^*\right) \qquad (2.20)\\
&= 2\Pr\left(Z_K(\widehat{\boldsymbol{\theta}}_\beta) > z_{\frac{\alpha}{2}} | \boldsymbol{\theta} = \boldsymbol{\theta}^*\right)\\
&= 2\Pr\left(\sqrt{\frac{K}{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}}}(\boldsymbol{m}^T\widehat{\boldsymbol{\theta}}_\beta - d) > z_{\frac{\alpha}{2}} | \boldsymbol{\theta} = \boldsymbol{\theta}^*\right)\\
&= 2\Pr\left(\sqrt{\frac{K}{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}}}\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*) > \right.\\
&\qquad \left. z_{\frac{\alpha}{2}} - \sqrt{\frac{K}{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}}}(\boldsymbol{m}^T\boldsymbol{\theta}^* - d)\right).
\end{aligned}$$

Finally, since $\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}$ is a consistent estimator of

$$\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\boldsymbol{K}_\beta(\boldsymbol{\theta}^*)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\boldsymbol{m}$$

and

$$\boldsymbol{m}^T\sqrt{K}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\bar{\boldsymbol{K}}_\beta(\boldsymbol{\theta}^*)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\boldsymbol{m}),$$

the desired result follows by following the Slutsky's theorem. ■

Remark 2.10 *Based on the above results, it is possible to provide an explicit expression for the number of devices as*

$$\widehat{K}(\beta,\alpha) = \left[\frac{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{m}}{\boldsymbol{m}^T\widehat{\boldsymbol{\theta}}_\beta - d}\left(z_{\frac{\alpha}{2}} - \Phi^{-1}(1 - \tfrac{\pi^*}{2})\right)^2\right] + 1,$$

necessary to get a fixed power π^* for a specific significance level α. Here, $[\cdot]$ denotes the largest integer value less than or equal to $\cdot$.

2.5 Robustness of the weighted minimum DPD estimators and Z-type tests

An important concept in robustness theory is the influence function (Hampel et al. [1986]). For any estimator defined in terms of an statistical functional $\boldsymbol{U}(F)$ from the true distribution F, its influence function (IF) is defined as

$$IF(t, \boldsymbol{U}, F) = \lim_{\varepsilon \downarrow 0} \frac{\boldsymbol{U}(F_\varepsilon) - \boldsymbol{U}(F)}{\varepsilon} = \left. \frac{\partial \boldsymbol{U}(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon = 0^+}, \tag{2.21}$$

where $F_\varepsilon = (1-\varepsilon)F + \varepsilon \Delta_t$, with ε being the contamination proportion and Δ_t being the degenerate distribution at the contamination point t. Thus, the (first-order) IF, as a function of t, measures the standardized asymptotic bias (in its first-order approximation) caused by the infinitesimal contamination at the point t. The maximum of this IF over t indicates the extent of bias due to contamination and so smaller its value, the more robust the estimator is.

2.5.1 Robustness of the weighted minimum DPD estimators

Let us denote by G_i the true distribution function of a Bernoulli random variable with an unknown probability of success, for the i-th group of K_i observations, having mass function g_i. Similarly, by $F_{i,\boldsymbol{\theta}}$ the distribution function of Bernoulli random variable having a probability of success equal to $\pi_{i1}(\boldsymbol{\theta})$, with probability mass function $f_i(\cdot, \boldsymbol{\theta})$ $(i = 1, ..., I)$, which are related to the model. In vector notation, we consider $\boldsymbol{G} = (G_1 \otimes \boldsymbol{1}_{K_1}^T, \ldots, G_I \otimes \boldsymbol{1}_{K_I}^T)^T$ and $\boldsymbol{F_\theta} = (F_{1,\boldsymbol{\theta}} \otimes \boldsymbol{1}_{K_1}^T, \ldots, F_{I,\boldsymbol{\theta}} \otimes \boldsymbol{1}_{K_I}^T)^T$.

For any estimator defined in terms of a statistical functional $\boldsymbol{U}(\boldsymbol{G})$ in the set-up of data from the true distribution function $\boldsymbol{G}$, its IF in accordance with (2.21) is defined as

$$IF(\boldsymbol{t}, \boldsymbol{U}, \boldsymbol{G}) = \lim_{\varepsilon \downarrow 0} \frac{\boldsymbol{U}(\boldsymbol{G}_{\varepsilon,\boldsymbol{t}}) - \boldsymbol{U}(\boldsymbol{G})}{\varepsilon} = \left. \frac{\partial \boldsymbol{U}(\boldsymbol{G}_{\varepsilon,\boldsymbol{t}})}{\partial \varepsilon} \right|_{\varepsilon = 0^+},$$

where $\boldsymbol{G}_{\varepsilon,\boldsymbol{t}} = (1-\varepsilon)\boldsymbol{G} + \varepsilon \Delta_{\boldsymbol{t}}$, with ε being the contamination proportion and $\Delta_{\boldsymbol{t}}$ being the distribution function of the degenerate random variable at the contamination point

$$\boldsymbol{t} = (t_{11}, ..., t_{1K_1}, ..., t_{I1}, ..., t_{IK_I})^T \in \mathbb{R}^{IK}.$$

We first need to define the statistical functional $\boldsymbol{U}_\beta(\boldsymbol{G})$ corresponding to the weighted minimum DPD estimator as the minimizer of the weighted sum of DPDs between the true and model densities. This is defined as the minimizer of

$$H_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y \in \{0,1\}} \left[f_i^{\beta+1}(y, \boldsymbol{\theta}) - \frac{\beta+1}{\beta} f_i^\beta(y, \boldsymbol{\theta}) g_i(y) \right] \right\}, \tag{2.22}$$

where $g_i(y)$ is the probability mass function associated to G_i and

$$f_i(y, \boldsymbol{\theta}) = y\pi_{i1}(\boldsymbol{\theta}) + (1-y)\pi_{i2}(\boldsymbol{\theta}), \quad y \in \{0,1\}.$$

If we choose $g_i(y) \equiv f_i(y, \boldsymbol{\theta})$, expression (2.22) is minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, implying the Fisher consistence of the minimum DPD estimator functional $\boldsymbol{U}_\beta(\boldsymbol{G})$ in our model.

Under appropriate differentiability conditions as the solution of the estimating equations

$$\frac{\partial H_\beta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y \in \{0,1\}} \left[f_i^\beta(y, \boldsymbol{\theta}) \frac{\partial f_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - f_i^{\beta-1}(y, \boldsymbol{\theta}) \frac{\partial f_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} g_i(y) \right] \right\} = \boldsymbol{0}, \tag{2.23}$$

In order to get the IF of the minimum DPD estimator at $F_{\boldsymbol{\theta}}$ with respect to the k-th element of the i_0-th group of observations we replace $\boldsymbol{\theta}$ in (2.23) by

$$\boldsymbol{\theta}_{\varepsilon}^{i_0}=\boldsymbol{U}_{\beta}(G_1\otimes\mathbf{1}_{K_1}^T,\ldots,G_{i_0-1}\otimes\mathbf{1}_{K_{i_0-1}}^T,G_{i_0,\varepsilon}\otimes\mathbf{1}_{K_{i_0}}^T,G_{i_0+1},\ldots,G_I\otimes\mathbf{1}_{K_I}^T),$$

where $G_{i_0,\varepsilon}$ is the distribution function associated to the probability mass function

$$g_{i_0,\varepsilon,k}(y)=(1-\varepsilon)f_i(y,\boldsymbol{\theta}^0)+\varepsilon\Delta_{t_{i_0,k}}(y),$$

where $\Delta_{t_{i_0,k}}(y)=y\Delta_{t_{i_0,k}}^{(1)}+(1-y)\Delta_{t_{i_0,k}}^{(2)}$, with $\Delta_{t_{i_0,k}}^{(1)}$ being the degenerating function at point (i_0,k), $\Delta_{t_{i_0,k}}^{(2)}=(1-\Delta_{t_{i_0,k}}^{(1)})$, and

$$g_i(y)=\begin{cases} f_i(y,\boldsymbol{\theta}^0) & \text{if } i\neq i_0;\\ g_{i_0,\varepsilon,k}(y) & \text{if } i=i_0.\end{cases}$$

Then, we have

$$\begin{aligned}\left.\frac{\partial H_{\beta}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}} =&\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}f_i^{\beta}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\right\}\\ &-\sum_{i\neq i_0}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}f_i^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}f_i(y,\boldsymbol{\theta}^0)\right\}\\ &-\frac{K_{i_0}}{K}\left\{\sum_{y\in\{0,1\}}f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\left[(1-\varepsilon)f_{i_0}(y,\boldsymbol{\theta}^0)+\varepsilon\Delta_{t_{i_0,k}}(y)\right]\right\}\end{aligned}\tag{2.24}$$

Now, we are going to get the derivative of (2.24) with respect to ε.

$$\begin{aligned}&\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[\beta f_i^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\frac{\partial\boldsymbol{\theta}_{\varepsilon}^{i_0}}{\partial\varepsilon}\right.\right.\\ &\qquad\left.\left.+f_i^{\beta}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\frac{\partial\boldsymbol{\theta}_{\varepsilon}^{i_0}}{\partial\varepsilon}\right]\right\}\\ &-\sum_{i\neq i_0}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[(\beta-1)f_i^{\beta-2}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\frac{\partial\boldsymbol{\theta}_{\varepsilon}^{i_0}}{\partial\varepsilon}f_i(y,\boldsymbol{\theta}^0)\right.\right.\\ &\qquad\left.\left.+f_i^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\frac{\partial\boldsymbol{\theta}_{\varepsilon}^{i_0}}{\partial\varepsilon}f_i(y,\boldsymbol{\theta}^0)\right]\right\}\\ &-\frac{K_{i_0}}{K}\left\{\sum_{y\in\{0,1\}}\left[(\beta-1)f_{i_0}^{\beta-2}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\frac{\partial\boldsymbol{\theta}_{\varepsilon}^{i_0}}{\partial\varepsilon}\left[(1-\varepsilon)f_i(y,\boldsymbol{\theta}^0)+\varepsilon\Delta_{t_{i_0,k}}(y)\right]\right.\right.\\ &\qquad-f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\frac{\partial\boldsymbol{\theta}_{\varepsilon}^{i_0}}{\partial\varepsilon}\left[(1-\varepsilon)f_i(y,\boldsymbol{\theta}^0)+\varepsilon\Delta_{t_{i_0,k}}(y)\right]\\ &\qquad\left.\left.-f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}^{i_0})\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}^{i_0}}\left[-f_i(y,\boldsymbol{\theta}^0)+\Delta_{t_{i_0,k}}(y)\right]\right]\right\}=\mathbf{0}\end{aligned}$$

Now, we evaluate the previous expression in $\varepsilon=0$, and we have

$$
\begin{aligned}
& IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[\beta f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right.\right.\\
&\qquad\left.\left.+f_i^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right\}\\
& -IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})\left[\sum_{i\neq i_0}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[(\beta-1)f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right.\right.\right.\\
&\qquad\left.\left.+f_i^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right\}\\
& \left.-\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}\left[(\beta-1)f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}-f_i^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial^2 f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right]\\
& -\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\Delta_{t_{i_0,k}}(y)+\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}=\mathbf{0}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
& IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[\beta f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right.\right.\\
&\qquad\left.\left.+f_i^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right\}\\
& -IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})\left[\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[(\beta-1)f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right.\right.\right.\\
&\qquad\left.\left.\left.+f_i^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right\}\right.\\
& -\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\Delta_{t_{i_0,k}}(y)+\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}=\mathbf{0}.
\end{aligned}
$$

Simplifying,

$$
\begin{aligned}
& IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right\}\\
& -\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\Delta_{t_{i_0,k}}(y)+\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}=\mathbf{0}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0}) &= \frac{\Pi_{1,i_0,k}}{\Pi_{2,k}} \qquad (2.25)\\
&= \frac{\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\Delta_{t_{i_0,k}}(y)-\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}}f_{i_0}^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}}{\sum_{i=1}^{I}\frac{K_i}{K}\left\{\sum_{y\in\{0,1\}}\left[f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right\}}.
\end{aligned}
$$

Let us now develop $\Pi_{1,i_0,k}$ and $\Pi_{2,i,k}$ in (2.25) trying to simplify the form of the IF.

$$
\begin{aligned}
&\sum_{y\in\{0,1\}} f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\Delta_{t_{i_0,k}}(y)-\frac{K_{i_0}}{K}\sum_{y\in\{0,1\}} f_{i_0}^{\beta}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_{i_0}(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\\
&=\sum_{y\in\{0,1\}}\left[f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}^0)\left(\Delta_{t_{i_0,k}}(y)-f_{i_0}(y,\boldsymbol{\theta})\right)\right]\\
&=\pi_{i_01}^{\beta-1}(\boldsymbol{\theta}^0)\left.\frac{\partial\pi_{i_01}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left(\Delta_{t_{i_0,k}}^{(1)}-\pi_{i_01}(\boldsymbol{\theta})\right)+\pi_{i_02}^{\beta-1}(\boldsymbol{\theta}^0)\left.\frac{\partial\pi_{i_02}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left(\Delta_{t_{i_0,k}}^{(2)}-\pi_{i_02}(\boldsymbol{\theta})\right)\\
&=\pi_{i_01}^{\beta-1}(\boldsymbol{\theta}^0)\left.\frac{\partial\pi_{i_01}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left(\Delta_{t_{i_0,k}}^{(1)}-\pi_{i_01}(\boldsymbol{\theta})\right)-\pi_{i_02}^{\beta-1}(\boldsymbol{\theta}^0)\left.\frac{\partial\pi_{i_01}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left(\pi_{i_01}(\boldsymbol{\theta})-\Delta_{t_{i_0,k}}^{(1)}\right)\\
&=\left.\frac{\partial\pi_{i_01}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left(\pi_{i_01}^{\beta-1}(\boldsymbol{\theta}^0)+\pi_{i_02}^{\beta-1}(\boldsymbol{\theta}^0)\right)\left(\pi_{i_01}(\boldsymbol{\theta})-\Delta_{t_{i_0,k}}^{(1)}\right)
\end{aligned}
$$

Then

$$
\Pi_{1,i_0,k}=\frac{K_{i_0}}{K}\left.\frac{\partial\pi_{i_01}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left(\pi_{i_01}^{\beta-1}(\boldsymbol{\theta}^0)+\pi_{i_02}^{\beta-1}(\boldsymbol{\theta}^0)\right)\left(\pi_{i_01}(\boldsymbol{\theta})-\Delta_{t_{i_0,k}}^{(1)}\right). \tag{2.26}
$$

On the other hand

$$
\begin{aligned}
&\sum_{y\in\{0,1\}}\left[f_i^{\beta-1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\\
&=\sum_{y\in\{0,1\}}\left[f_i^{\beta+1}(y,\boldsymbol{\theta}^0)\left.\frac{\partial\log f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\left.\frac{\partial\log f_i(y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\\
&=\pi_{i1}^{\beta+1}(\boldsymbol{\theta}^0)\boldsymbol{u}_{i1}(\boldsymbol{\theta}^0)\boldsymbol{u}_{i1}^T(\boldsymbol{\theta}^0)+\pi_{i2}^{\beta+1}(\boldsymbol{\theta}^0)\boldsymbol{u}_{i2}(\boldsymbol{\theta}^0)\boldsymbol{u}_{i2}^T(\boldsymbol{\theta}^0)\\
&=\sum_{j=1}^{2}\pi_{ij}^{\beta+1}(\boldsymbol{\theta}^0)\boldsymbol{u}_{ij}(\boldsymbol{\theta}^0)\boldsymbol{u}_{ij}^T(\boldsymbol{\theta}^0).
\end{aligned}
$$

Therefore,

$$
\Pi_{2,k}=\sum_{i=1}^{K}\frac{K_i}{K}\sum_{j=1}^{2}\pi_{ij}^{\beta+1}(\boldsymbol{\theta}^0)\boldsymbol{u}_{ij}(\boldsymbol{\theta}^0)\boldsymbol{u}_{ij}^T(\boldsymbol{\theta}^0)=J_\beta(\boldsymbol{\theta}^0) \tag{2.27}
$$

Then, taking unto account (2.26) and (2.27), we get

$$
\begin{aligned}
IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})=&\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\frac{K_{i_0}}{K}\left.\frac{\partial\pi_{i_01}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\\
&\times\left(\pi_{i_01}^{\beta-1}(\boldsymbol{\theta}^0)+\pi_{i_02}^{\beta-1}(\boldsymbol{\theta}^0)\right)\left(\pi_{i_01}(\boldsymbol{\theta}^0)-\Delta_{t_{i_0,k}}^{(1)}\right)
\end{aligned} \tag{2.28}
$$

Proposition 2.11 *Let us consider the one-shot device testing under the exponential distribution with a simple stress factor defined in (2.1). The IF with respect to the $k-th$ observation of the i_0-th group is given by*

$$
\begin{aligned}
IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})=&\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\frac{K_{i_0}}{K}\left(F^{\beta-1}(IT_{i_0};x_{i_0},\boldsymbol{\theta}^0)+R^{\beta-1}(IT_{i_0};x_{i_0},\boldsymbol{\theta}^0)\right)\\
&\times\left(F(IT_{i_0};x_{i_0},\boldsymbol{\theta}^0)-\Delta_{t_{i_0,k}}^{(1)}\right)f(IT_{i_0};x_{i_0},\boldsymbol{\theta}^0)IT_{i_0}\boldsymbol{\nu}_{i_0},
\end{aligned} \tag{2.29}
$$

where $\Delta_{t_{i_0,k}}^{(1)}$ is the degenerating function at point $t_{i_0,k}$ and $\boldsymbol{\nu}_{i_0}=(1/\theta_0,x_{i_0})^T$.

Proof. Straightforward from (2.28) and taking into account that $\left.\frac{\partial \pi_{i_0 1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \boldsymbol{\nu}_{i_0}$. ∎

In order to get the IF of the minimum DPD estimator at $F_{\boldsymbol{\theta}}$ with respect to all the observations, we replace the parameter $\boldsymbol{\theta}$ in (2.23) by

$$\boldsymbol{\theta}_{\varepsilon}^{i_0} = \boldsymbol{U}_{\beta}(G_{1,\varepsilon} \otimes \mathbf{1}_{K_1}^T, \ldots, G_{i_0-1,\varepsilon} \otimes \mathbf{1}_{K_{i_0-1}}^T, G_{i_0,\varepsilon} \otimes \mathbf{1}_{K_{i_0}}^T, G_{i_0+1,\varepsilon}, \ldots, G_{I,\varepsilon} \otimes \mathbf{1}_{K_I}^T),$$

and the probability mass function $g_i(y)$ by

$$g_{i,\varepsilon}(y) = (1-\varepsilon) f_i(y, \boldsymbol{\theta}^0) + \varepsilon \Delta_{t_i}(y),$$

where $\Delta_{t_i}(y) = \sum_{k=1}^{K_i} \Delta_{t_{i,k}}(y)$, and we get

$$\begin{aligned}\left.\frac{\partial H_{\beta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} =& \sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y\in\{0,1\}} f_i^{\beta}(y, \boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \right\} \\ & - \sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y\in\{0,1\}} f_i^{\beta-1}(y, \boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} f_i(y, \boldsymbol{\theta}^0) \right\}\end{aligned} \tag{2.30}$$

Differentiating with respect to ε, we have

$$\begin{aligned}&\sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y\in\{0,1\}} \left[\beta f_i^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \frac{\partial \boldsymbol{\theta}_{\varepsilon}}{\partial \varepsilon} + f_i^{\beta}(y,\boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \frac{\partial \boldsymbol{\theta}_{\varepsilon}}{\partial \varepsilon} \right] \right\} \\ &\quad - \sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y\in\{0,1\}} \left[(\beta-1) f_i^{\beta-2}(y,\boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \frac{\partial \boldsymbol{\theta}_{\varepsilon}}{\partial \varepsilon} g_{i,\varepsilon,k}(y) \right.\right. \\ &\left.\left. - f_{i_0}^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial^2 f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \frac{\partial \boldsymbol{\theta}_{\varepsilon}}{\partial \varepsilon} g_{i,\varepsilon,k}(y) - f_i^{\beta-1}(y,\boldsymbol{\theta}_{\varepsilon}) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\varepsilon}} \left[-f_i(y,\boldsymbol{\theta}^0) + \Delta_{t_{i,k}}(y) \right] \right] \right\} = \mathbf{0}\end{aligned}$$

Finally,

$$\begin{aligned}IF(\boldsymbol{t}, \boldsymbol{U}_{\beta}, F_{\boldsymbol{\theta}^0}) =& \frac{\Pi_{1,k}}{\Pi_{2,k}} \\ =& \frac{\sum_{i=1}^{I} \frac{K_i}{K} \sum_{y\in\{0,1\}} f_i^{\beta-1}(y,\boldsymbol{\theta}^0) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \Delta_{t_{i,k}}(y) - \sum_{i=1}^{I} \frac{K_i}{K} \sum_{y\in\{0,1\}} f_i^{\beta}(y,\boldsymbol{\theta}^0) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}}{\sum_{i=1}^{I} \frac{K_i}{K} \left\{ \sum_{y\in\{0,1\}} \left[f_i^{\beta-1}(y,\boldsymbol{\theta}^0) \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \left.\frac{\partial f_i(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] \right\}}.\end{aligned} \tag{2.31}$$

Proposition 2.12 *Let us consider the one-shot device testing under the exponential distribution with a simple stress factor defined in (2.1). The IF with respect to all the observations is given by*

$$\begin{aligned}IF(\boldsymbol{t}, \boldsymbol{U}_{\beta}, F_{\boldsymbol{\theta}^0}) =& \boldsymbol{J}_{\beta}^{-1}(\boldsymbol{\theta}^0) \sum_{i=1}^{I} \frac{K_i}{K} \left[\left(F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}^0) + R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}^0) \right) \right. \\ & \left. \times \left(F(IT_i; x_i, \boldsymbol{\theta}^0) - \Delta_{t_i}^{(1)} \right) f(IT_i; x_i, \boldsymbol{\theta}^0) IT_i \boldsymbol{\nu}_i \right],\end{aligned} \tag{2.32}$$

where $\Delta_{t_i}^{(1)} = \sum_{k=1}^{K_i} \Delta_{t_{i,k}}^{(1)}$ *and* $\boldsymbol{\nu}_i = (1/\theta_0, x_i)^T$.

Proof. Straightforward from (2.31) and taking into account that $\left.\frac{\partial \pi_{i_1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \boldsymbol{\nu}_i$. ∎

Remark 2.13 *Let*

$$\begin{aligned}
h_1(IT,x,\boldsymbol{\theta}) &= \left(F^{\beta-1}(IT;x,\boldsymbol{\theta}) + R^{\beta-1}(IT;x,\boldsymbol{\theta})\right) f(IT;x,\boldsymbol{\theta}) IT\frac{1}{\theta_0} \\
&= \Big[\exp\left\{\theta_1 x - \theta_0 \exp\left(\theta_1 x\right) IT\right\} \left[1-\exp\left\{-\theta_0\exp\left(\theta_1 x\right) IT\right\}\right]^{\beta-1} \\
&\quad + \exp\left\{\theta_1 x - \beta\theta_0\exp\left(\theta_1 x\right) IT\right\}\theta_0 IT\Big]\frac{1}{\theta_0} \\
h_2(IT,x,\boldsymbol{\theta}) &= \left(F^{\beta-1}(IT;x,\boldsymbol{\theta}) + R^{\beta-1}(IT;x,\boldsymbol{\theta})\right) f(IT;x,\boldsymbol{\theta}) IT\frac{1}{\theta_0} \\
&= \Big[\exp\left\{\theta_1 x - \theta_0 \exp\left(\theta_1 x\right) IT\right\} \left[1-\exp\left\{-\theta_0\exp\left(\theta_1 x\right) IT\right\}\right]^{\beta-1} \\
&\quad + \exp\left\{\theta_1 x - \beta\theta_0\exp\left(\theta_1 x\right) IT\right\}\theta_0 IT\Big] x
\end{aligned}$$

be the factors of the influence function of $\boldsymbol{\theta}$ given in (2.29) and (2.32). Based on this, might be commented on conditions for boundedness of the influence functions presented in this Chapter, either with respect to an observation other with respect to all the observations, that they are bounded on $t_{i_0,k}$ or $\boldsymbol{t}$, however if $\beta = 0$ the norm of the bidimensional influence functions can be very large on (x, IT), in comparison with $\beta > 0$, since

$$\begin{aligned}
\lim_{\substack{x\to+\infty \\ (\theta_1<0)}} h_1(IT,x,\boldsymbol{\theta}) = \lim_{\substack{x\to+\infty \\ (\theta_1>0)}} h_2(IT,x,\boldsymbol{\theta}) &= \lim_{t\to+\infty} h_1(IT,x,\boldsymbol{\theta}) \\
&= \lim_{IT\to+\infty} h_2(IT,x,\boldsymbol{\theta}) \begin{cases} =\infty, & \text{if } \beta = 0 \\ <\infty, & \text{if } \beta > 0 \end{cases}.
\end{aligned}$$

This implies that the proposed weighted minimum DPD estimators with $\beta > 0$ are robust against leverage points, but the classical MLE is clearly non-robust. Same happens for large ITs too, but in accelerated processes inspection time tends not to be large.

In addition, it is interesting to note that

$$F(IT;x,\boldsymbol{\theta}) \underset{x\to+\infty}{\longrightarrow} \begin{cases} 1, & \text{if } \theta_1 > 0 \\ 0, & \text{if } \theta_1 < 0 \end{cases},$$

which matches with the degenerated distributions for cells, and similarly

$$\lambda(\boldsymbol{\theta}) \underset{x\to+\infty}{\longrightarrow} \begin{cases} +\infty, & \text{if } \theta_1 > 0 \\ 0, & \text{if } \theta_1 < 0 \end{cases}$$

approaches both boundaries of $\lambda(\boldsymbol{\theta}) \in (0, +\infty)$.

2.5.2 Robustness of the Z-type tests

Next, we study the robustness of the proposed Z-type test statistics. The IF of a testing procedure, as introduced by Ronchetti and Rousseeuw [1979] for IID data, is also defined as in the case of estimation but with the statistical functional corresponding to the test statistics and it is studied under the null hypothesis. This concept has been extended to the non- homogeneous data, by Aerts and Haesbroeck [2017] and Ghosh and Basu [2018]. In our context, the functional associated with the Z-type test, evaluated at $\boldsymbol{U}_\beta(\boldsymbol{G})$ is given by

$$Z_K(\boldsymbol{U}_\beta(\boldsymbol{G})) = \sqrt{\frac{K}{\boldsymbol{m}^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{U}_\beta(\boldsymbol{G}))\boldsymbol{K}_\beta(\boldsymbol{U}_\beta(\boldsymbol{G}))\boldsymbol{J}_\beta^{-1}(\boldsymbol{U}_\beta(\boldsymbol{G}))\boldsymbol{m}}}\left(\boldsymbol{m}^T\boldsymbol{U}_\beta(\boldsymbol{G}) - d\right). \tag{2.33}$$

The influence function with respect to the $k-$th observation of the i_0-th group of observations, of the functional associated with the Z -type test statistics for testing the composite null hypothesis in (2.18), is then given by

$$IF(t_{i_0,k}, Z_K, F_{\boldsymbol{\theta}^0}) = \left.\frac{\partial Z_K(F_{\boldsymbol{\theta}_\varepsilon^{i_0}})}{\partial \varepsilon}\right|_{\varepsilon=0^+}.$$

But,

$$\left.\frac{\partial Z_K(F_{\boldsymbol{\theta}_\varepsilon^{i_0}})}{\partial \varepsilon}\right|_{\varepsilon=0^+} = \Phi_K(\boldsymbol{\theta}^0)\boldsymbol{m}^T \left.\frac{\partial \boldsymbol{U}_\beta(F_{\boldsymbol{\theta}_\varepsilon^{i_0}})}{\partial \varepsilon}\right|_{\varepsilon=0^+},$$

where $\left.\frac{\partial \boldsymbol{U}_\beta(F_{\boldsymbol{\theta}_\varepsilon^{i_0}})}{\partial \varepsilon}\right|_{\varepsilon=0^+}$ is the IF of the estimator and

$$\Phi_K(\boldsymbol{\theta}^0) = \sqrt{\frac{K}{\boldsymbol{m}^T \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{m}}}. \tag{2.34}$$

Therefore,

Proposition 2.14 *Let us consider the one-shot device testing under the exponential distribution with a simple stress factor defined in (2.1). The IF of the functional associated with the Z -type test statistics for testing the composite null hypothesis in (2.16), with respect to the $k-$th observation of the i_0-th group is given by*

$$IF(t_{i_0,k}, Z_K, F_{\boldsymbol{\theta}^0}) = \Phi_K(\boldsymbol{\theta}^0)\boldsymbol{m}^T IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}),$$

where $IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})$ is given in (2.29) and $\Phi_K(\boldsymbol{\theta}^0)$ is given in (2.34).

Similarly, for all the indices,

Proposition 2.15 *Let us consider the one-shot device testing under the exponential distribution with a simple stress factor defined in (2.1). The IF of the functional associated with the Z -type test statistics for testing the composite null hypothesis in (2.16), with respect to all the observations is given by*

$$IF(\boldsymbol{t}, Z_K, F_{\boldsymbol{\theta}^0}) = \Phi_K(\boldsymbol{\theta}^0)\boldsymbol{m}^T IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}),$$

where $IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})$ is given in (2.32) and $\Phi_K(\boldsymbol{\theta}^0)$ is given in (2.34).

From these results, same conclusions are derived about the boundedness of the influence function associated with the Z -type test statistics presented on this Chapter.

2.6 Simulation study

In this section, a simulation study is carried out to examine the behavior of the weighted minimum DPD estimators of the parameters of the one-shot device model, studied in this chapter, as well as the corresponding Z-type tests, based on weighted minimum DPD estimators. We pay special attention to the robustness issue here. It is interesting to note, in this context, the following. For each fixed time, IT_i, under a fixed temperature, x_i, K_i devices are tested. In particular, a balanced data with equal sample size for each group is considered.

As it happens for product binomial sampling model, we must consider "outlying cells" rather than "outlying observations". A cell which does not follow the one-shot device model will be called an outlying cell or outlier. The strong outliers may lead to reject a model fitting even if the rest of the cells fit the model properly. In other words, even though the cells seem to fit reasonably well the model, the outlying cells contribute to an increase in the values of the residuals as well as the divergence measure between the data and the fitted values according to the one-shot device model considered. Therefore, it is very important to have robust estimators as well as robust test statistics in order to avoid the undesirable effects of outliers in the data. The main purpose of this simulation study is to empirically illustrate that inside the family of weighted minimum DPD estimators developed in this chapter, some estimators may have better robust properties than the MLE, and the Z-type tests constructed from them can be at the same time more robustness than the classical Z-type test constructed through the MLEs.

2.6.1 Weighted minimum density power divergence estimators

The simulation study is carried out to compare the behavior of some weighted minimum DPD estimators with respect to the MLEs of the parameters in the one-shot device model under the exponential distribution with a simple stress factor. In order to evaluate the performance of the proposed weighted minimum DPD estimators, as well as the MLEs, the root of the mean square errors (RMSEs) are considered. A model in which $I = 9$, different conditions are obtained from the combination of the temperatures $x \in \{35, 45, 55\}$, being the inspection times $IT \in \{10, 20, 30\}$ and $K_i = 20\ \forall i = 1, \ldots, I$, as in Table 2.6.1, and the simulation experiment proposed by Balakrishnan and Ling [2012b]. This model has been examined under three choices of $(\theta_0, \theta_1) = (0.005, 0.05)$, $(\theta_0, \theta_1) = (0.004, 0.05)$ and $(\theta_0, \theta_1) = (0.003, 0.05)$ for low-moderate, moderate and moderate-high reliability, respectively.

Table 2.6.1: Exponential distribution at a simple stress level: simulation scheme

i	x_i	IT_i	K_i
1	35	10	20
2	45	20	20
3	55	30	20
4	35	10	20
5	45	20	20
6	55	30	20
7	35	10	20
8	45	20	20
9	55	30	20

To evaluate the robustness of the weighted minimum DPD estimators, we have studied the behavior of this model under the consideration of an outlying cell for $i = 1$, with $10,000$ replications and estimators corresponding to the tuning parameter $\beta \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$. The reduction of each one of the parameters of the outlying cell, denoted by $\tilde{\theta}_0$ or $\tilde{\theta}_1$ ($\theta_0 \geq \tilde{\theta}_0$ or $\theta_1 \geq \tilde{\theta}_1$) increases the mean of its lifetime distribution function given in (2.1). The results obtained by decreasing parameter θ_0 and by decreasing parameter θ_1 are shown in Figure 2.6.1. In all the cases, we can see how the MLEs and the weighted minimum DPD estimators with small values of tuning parameter β present the smallest RMSEs for weak outliers, i.e., when $\tilde{\theta}_0$ is close to θ_0 ($1 - \tilde{\theta}_0/\theta_0$ is close to 0) or $\tilde{\theta}_1$ is close to θ_1 ($1 - \tilde{\theta}_1/\theta_1$ is close to 0). On the other hand, large values of tuning parameter β make the weighted minimum DPD estimators to present the smallest RMSEs, for medium and strong outliers, i.e., when $\tilde{\theta}_0$ is not close to θ_0 ($1 - \tilde{\theta}_0/\theta_0$ is not close to 0) or $\tilde{\theta}_1$ is not close to θ_1 ($1 - \tilde{\theta}_1/\theta_1$ is not close to 0). Therefore, the MLE of (θ_0, θ_1) is very efficient when there are no outliers, but highly non-robust when there are outliers. On the other hand, the weighted minimum DPD estimators with moderate values of the tuning parameter β exhibit a little loss of efficiency without outliers, but at the same time possess a considerable improvement in robustness in the presence of outliers. Actually, these values of the tuning parameter β are the

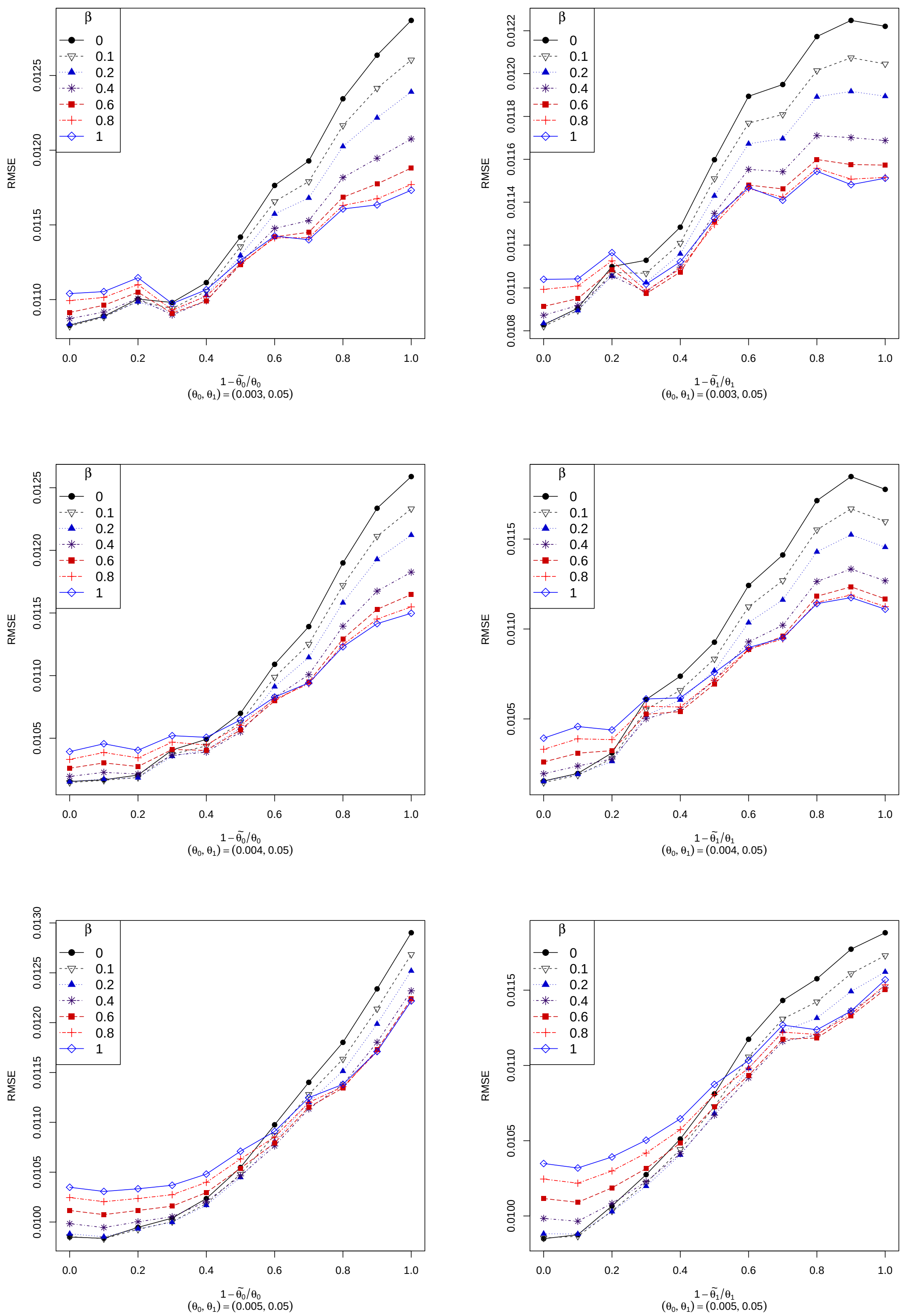

Figure 2.6.1: Exponential distribution at a simple stress level: RMSEs of $\boldsymbol{\theta}$ estimates

most appropriate ones for the estimators of the parameters in the one-shot device model according to robustness theory: To improve in a considerable way the robustness of the estimators, a small amount of efficiency needs to be compromised.

2.6.2 Z-type tests

Let us consider the same Simulation Scheme defined in Section 2.6.1 (Table 2.6.1). We are interested in testing the null hypothesis $H_0 : \theta_1 = 0.05$ against the alternative $H_1 : \theta_1 \neq 0.05$, through the Z-type test statistics based on weighted minimum DPD estimators. Under the null hypothesis, we consider as true parameters $(\theta_0, \theta_1) = (0.004, 0.05)$, while under the alternative we consider as true parameters $(\theta_0, \theta_1) = (0.004, 0.02)$. In Figure 2.6.2, we present the empirical significance level (measured as the proportions of test statistics exceeding in absolute value the standard normal quantile critical value) based on 10,000 replications. The empirical power (obtained in a similar manner) is also presented in the right hand side of Figure 2.6.2. Notice that, in all the cases, the observed levels are quite close to the nominal level of 0.05. The empirical power is similar for the different values of the tuning parameters β, a bit lower for large values of β, and closer to one as the sample size K increases.

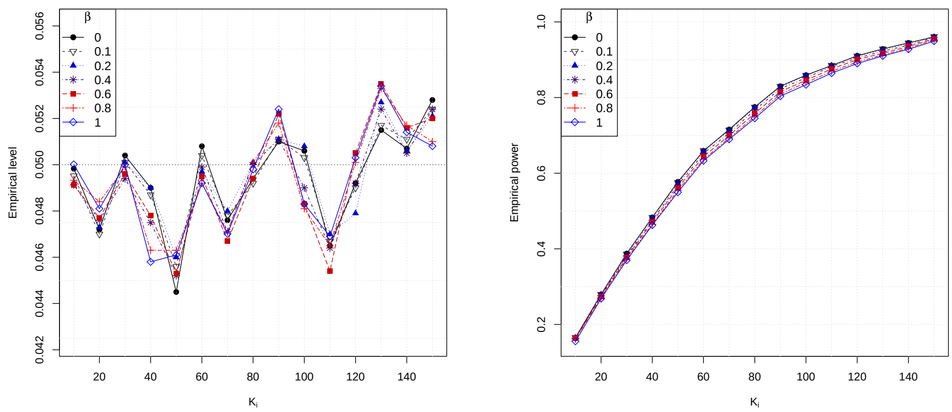

Figure 2.6.2: Exponential distribution at a simple stress level: simulated levels (left) and powers (right) with no outliers in the data.

To evaluate the robustness of the level and the power of the Z-type tests based on weighted minimum DPD estimators with an outlier placed on the first-row cell, we perform the simulation for the same test and the same true values for the null and alternative hypotheses, in two different scenarios depending on the way the outlying cell is considered. In the first scenario, we keep θ_1 the same and modify the true value of θ_0 to be $\tilde{\theta}_0 \leq \theta_0$, and in the second one, we keep θ_0 the same and modify the true value of θ_1 to be $\tilde{\theta}_1 \leq \theta_1$. Both cases have been analyzed for different values of K and decreasing $\tilde{\theta}_0$ in the first scenario (increasing $1 - \tilde{\theta}_0/\theta_0$) or decreasing $\tilde{\theta}_1$ in the second scenario (increasing $1 - \tilde{\theta}_1/\theta_1$).

The results for the first scenario are presented in Figure 2.6.3. The empirical level for the one-shot device model with K_i from 10 to 150, true value $(\theta_0, \theta_1) = (0.004, 0.05)$ and $\tilde{\theta}_0 = 0.001$ for the outlying cell is presented on the top left panel. Similarly, the empirical power for the one-shot device model with K_i from 10 to 150, true parameter $(\theta_0, \theta_1) = (0.004, 0.02)$ and $\tilde{\theta}_0 = 0.001$ for the outlying cell is presented on the right top panel. In addition, the empirical level for the one-shot device model with $1 - \tilde{\theta}_0/\theta_0$ from 0 to 1 for the outlying cell and true value $(\theta_0, \theta_1) = (0.004, 0.05)$ and $K_i = 20$ is presented on the bottom left panel. Similarly, the empirical power for the one-shot device model with $1 - \tilde{\theta}_0/\theta_0$ from 0 to 1 for the outlying cell and true value and true parameter $(\theta_0, \theta_1) = (0.004, 0.02)$ is presented on the bottom right panel.

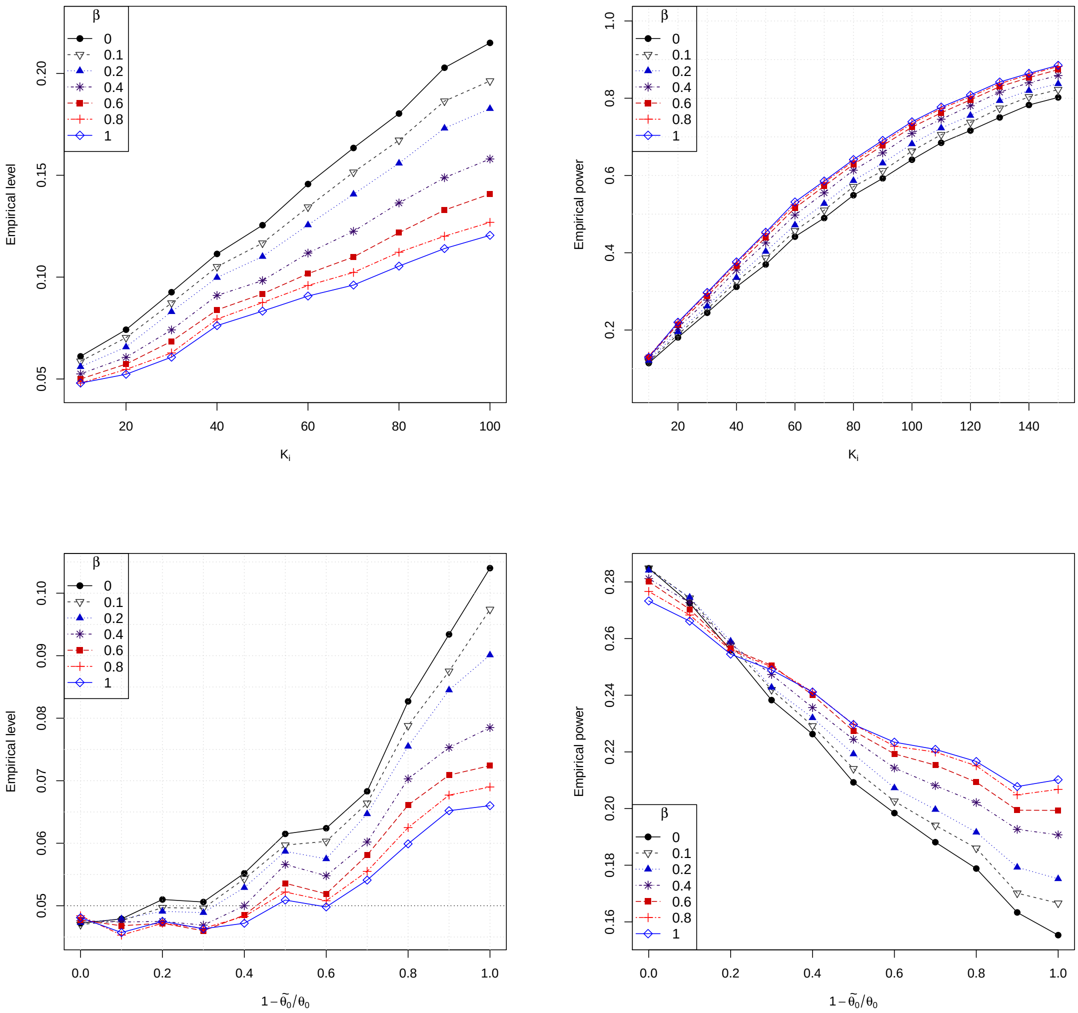

Figure 2.6.3: Exponential distribution at a simple stress level: simulated levels (left) and powers (right) with an θ_0-contaminated outlying cell in the data.

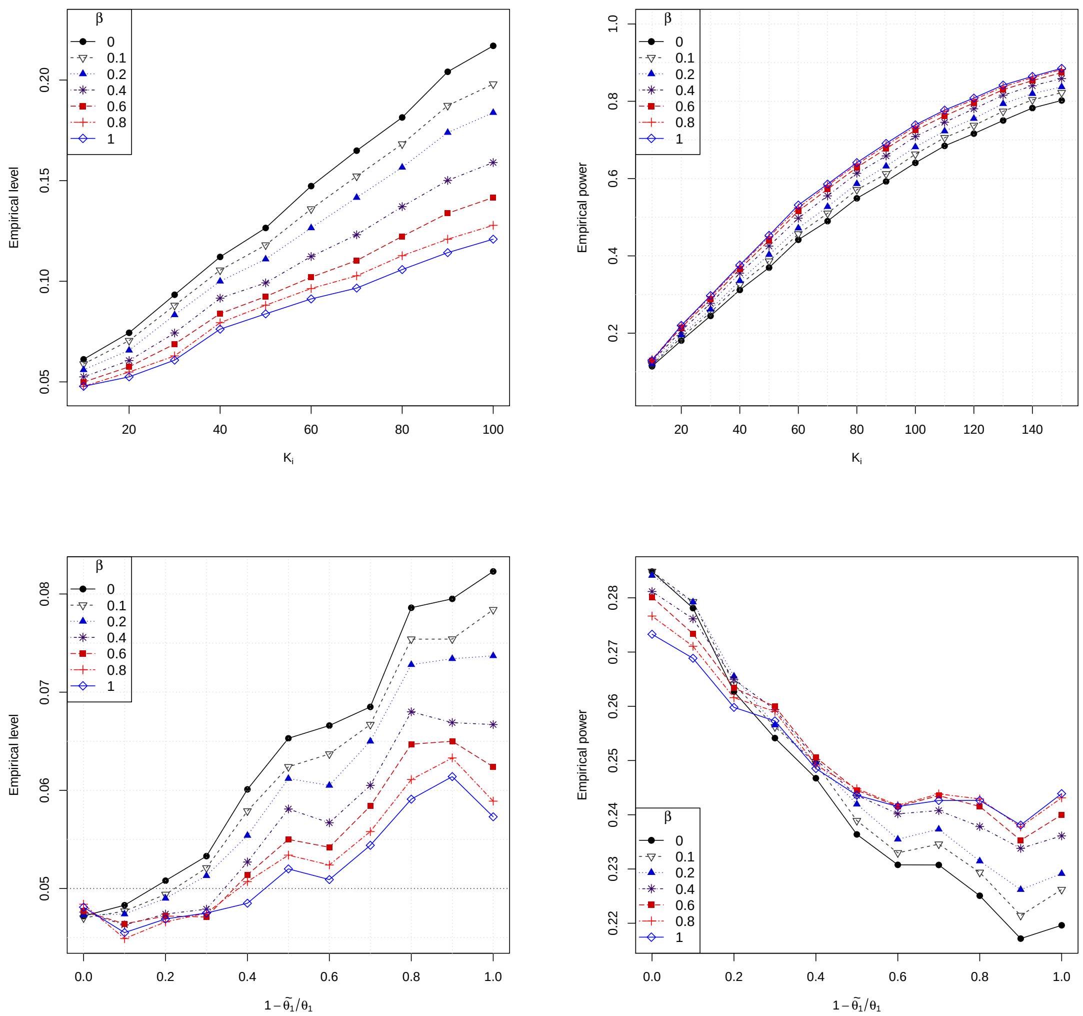

Figure 2.6.4: Exponential distribution at a simple stress level: simulated levels (left) and powers (right) with an θ_1-contaminated outlying cell in the data.

Notice that the outlying cell represents 1/9 of the total observations in the last plots. For large values of K_i (very large sample sizes), there is a large inflation in the empirical level and shrinkage of the empirical power, but for the Z-type test statistic based on the weighted minimum DPD estimators with large values of the tuning parameter β, the effect of the outlying cell is weaker in comparison to those of smaller values of β, including the MLEs ($\beta = 0$). If $\tilde{\theta}_0$ is separated from θ_0 ($1 - \tilde{\theta}_0/\theta_0$ increases from 0 to 1), the empirical level of the Z-type test statistics based on the weighted minimum DPD estimators is not stable around the nominal level, but being closer as the tuning parameter β becomes larger. If $\tilde{\theta}_0$ is separated from θ_0 ($1 - \tilde{\theta}_0/\theta_0$ increases from 0 to 1), the empirical power of the Z-type test statistics based on the weighted minimum DPD estimators decreases, but being more slowly as the tuning parameter β becomes larger.

Figure 2.6.4 presents the results for the second scenario, in which $\tilde{\theta}_1 = 0.01$ for the outlying cell on the top left panel and $\tilde{\theta}_1 = -0.01$ for the outlying cell on the top right panel. Even though the outliers are, in the current scenario, slightly more pronounced with respect to the previous scenario, in general terms, we arrive at the same conclusions as in the previous scenario.

2.6.3 Choice of tuning parameter

Throughout this section, we have noted that the robustness of the proposed weighted minimum DPD estimator seems to increase with increasing β; but, their pure data efficiency decrease slightly. From the results of our simulation study, a moderately large value of β is expected to provide the best trade-off for possibly contaminated data. Although a possible ad-hoc choice of β may work quite well in practice, when working with real data, a data-driven choice of β would be better and convenient.

A useful procedure of the data-based selection of β for the weighted minimum DPD estimator was proposed by Warwick and Jones [2005]. It consists of minimizing the estimated mean squared error, an approach that requires pilot estimation of model parameters. We can adopt a similar approach to obtain a suitable data-driven β in our model. In this approach, we minimize an estimate of the asymptotic MSE of the weighted minimum DPD estimator $\widehat{\boldsymbol{\theta}}_\beta$ given by

$$\widehat{MSE}(\beta) = (\boldsymbol{\theta}_\beta - \boldsymbol{\theta}_P)^T(\boldsymbol{\theta}_\beta - \boldsymbol{\theta}_P) + \frac{1}{K}\text{trace}\left\{\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_\beta)\boldsymbol{K}_\beta(\boldsymbol{\theta}_\beta)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_\beta)\right\},$$

where $\boldsymbol{\theta}_P$ is a pilot estimator, whose choice will be empirically discussed, as the overall procedure depends on this choice. If we take $\boldsymbol{\theta}_P = \widehat{\boldsymbol{\theta}}_\beta$, the approach coincides with that of Hong and Kim [2001], but it does not take into account the model misspecification.

However, as pointed out by Basu et al. [2017], when dealing with the robustness issue, the estimation of the variance component should not assume the model to be true. So, following the general formulation of Ghosh and Basu [2015], we have the following result:

Proposition 2.16 *Let us consider the one-shot device model under the exponential distribution with a single stress factor with distribution function (2.1). The model robust estimates of $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ defined in (2.14) and (2.15), respectively, can be obtained as*

$$\begin{aligned}\widehat{\boldsymbol{J}}_\beta(\boldsymbol{\theta}) =& (\beta+1)\boldsymbol{J}_\beta(\boldsymbol{\theta}) + \sum_{i=1}^{I}\frac{K_i}{K}\left[F(IT_i;x_i,\boldsymbol{\theta}) - \frac{n_i}{K_i}\right] \\ &\times\left\{\boldsymbol{C}_1^{(i)}(\boldsymbol{\theta})\left[F^{\beta-1}(IT_i;x_i,\boldsymbol{\theta}) + R^{\beta-1}(IT_i;x_i,\boldsymbol{\theta})\right] - \boldsymbol{C}_2^{(i)}(\boldsymbol{\theta})\left[F^{\beta-2}(IT_i;x_i,\boldsymbol{\theta}) - R^{\beta-2}(IT_i;x_i,\boldsymbol{\theta})\right]\right\} \\ &-\beta\sum_{i=1}^{I}\frac{K_i-n_i}{K_i}\boldsymbol{\Delta}^{(i)}(\boldsymbol{\theta})IT_i^2 f^2(IT_i;x_i,\boldsymbol{\theta})\left[\frac{n_i}{K_i}F^{\beta-2}(IT_i;x_i,\boldsymbol{\theta}) + \frac{K_i-n_i}{K_i}R^{\beta-2}(IT_i;x_i,\boldsymbol{\theta})\right],\end{aligned} \tag{2.35}$$

$$\widehat{\boldsymbol{K}}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Delta}^{(i)}(\boldsymbol{\theta})IT_i^2 f^2(IT_i;x_i,\boldsymbol{\theta}) \tag{2.36}$$

$$
\times \left\{ \left[\frac{n_i}{K_i} F^{2\beta-2}(IT_i; x_i, \boldsymbol{\theta}) + \frac{K_i - n_i}{K_i} R^{\beta-2}(IT_i; x_i, \boldsymbol{\theta}) \right] \right.
$$
$$
\left. - \left[\frac{n_i}{K_i} F^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) - \frac{K_i - n_i}{K_i} R^{\beta-1}(IT_i; x_i, \boldsymbol{\theta}) \right]^2 \right\},
$$

where

$$
\boldsymbol{\Delta}^{(i)}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{x_i}{\theta_0} \\ \frac{x_i}{\theta_0} & x_i^2 \end{pmatrix},
$$
$$
\boldsymbol{C}_1^{(i)}(\boldsymbol{\theta}) = \begin{pmatrix} -\frac{1}{\theta_0^2} & 0 \\ 0 & 0 \end{pmatrix} IT_i^2 f(IT_i; x_i, \boldsymbol{\theta}) + \boldsymbol{\Delta}^{(i)}(\boldsymbol{\theta}) IT_i f(IT_i; x_i, \boldsymbol{\theta}) F(IT_i; x_i, \boldsymbol{\theta}), \tag{2.37}
$$
$$
\boldsymbol{C}_2^{(i)}(\boldsymbol{\theta}) = \boldsymbol{\Delta}^{(i)}(\boldsymbol{\theta}) IT_i^2 f^2(IT_i; x_i, \boldsymbol{\theta}). \tag{2.38}
$$

Proof. Following Ghosh and Basu [2015], we have

$$
\widehat{\boldsymbol{J}}_\beta(\boldsymbol{\theta}) = (\beta+1)\boldsymbol{J}_\beta(\boldsymbol{\theta}) + \sum_{i=1}^{I}\sum_{j=1}^{2} \frac{K_i}{K} \left\{ \frac{\partial \boldsymbol{u}_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \pi_{ij}^{\beta+1}(\boldsymbol{\theta}) - \frac{n_{ij}}{K_i} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \left[\beta \boldsymbol{u}_{ij}(\boldsymbol{\theta}) \boldsymbol{u}_{ij}^T(\boldsymbol{\theta}) - \frac{\partial \boldsymbol{u}_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \right\}
$$

and

$$
\widehat{\boldsymbol{K}}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{j=1}^{2} \frac{K_i}{K} \left\{ \boldsymbol{u}_{ij}(\boldsymbol{\theta}) \boldsymbol{u}_{ij}^T(\boldsymbol{\theta}) \frac{n_{ij}}{K_i} \pi_{ij}^{2\beta}(\boldsymbol{\theta}) \right\} - \sum_{i=1}^{I}\sum_{j=1}^{2} \frac{K_i}{K} \boldsymbol{\xi}_{ij,\beta}^{*}(\boldsymbol{\theta}) \boldsymbol{\xi}_{ij,\beta}^{*T}(\boldsymbol{\theta})
$$

with $\boldsymbol{\xi}_{ij,\beta}^{*}(\boldsymbol{\theta}) = \frac{n_{ij}}{K_i} \boldsymbol{u}_{ij}(\boldsymbol{\theta}) \pi_{ij}^{\beta}(\boldsymbol{\theta})$, $n_{i1} = n_i$ and $n_{i2} = K_i - n_i$.

The required result follows taking into account that

$$
\frac{\partial \boldsymbol{u}_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\pi_{ij}(\boldsymbol{\theta})} \boldsymbol{C}_1^{(i)}(\boldsymbol{\theta}) + \frac{1}{\pi_{ij}^2(\boldsymbol{\theta})} \boldsymbol{C}_2^{(i)}(\boldsymbol{\theta}),
$$

where $\boldsymbol{C}_1^{(i)}(\boldsymbol{\theta})$ and $\boldsymbol{C}_2^{(i)}(\boldsymbol{\theta})$ are as given in (2.37) and (2.38), respectively. ■

Let us reconsider the previous simulation study with $(\theta_0, \theta_1) = (0.004, 0.05)$, but now we perform the selection of β following the above proposal for each iteration with different possible pilot estimators. Let us consider as potential pilot parameters $\beta_P = \{0, 0.3, 0.6, 0.9\}$. The selection of β is done through a grid search of $[0, 1]$ with spacing 0.01 and 10,000 samples. In Figure 2.6.5, we show the simulated true RMSEs for this scenario and the average optimal values of β for this same scenario. We can observe how the use of pilot estimators leads us to different optimal values of β, but, in general cases, optimal values of β are higher when a higher degree of contamination is considered, as expected. It seems that the best trade-off between the efficiency in pure data and the robustness under contaminated data is provided by the pilot choice $\beta_P = 0.4$ and so we suggest it as our pilot estimator. This method, summarized in Algorithm 2.6.5, will be applied in the following section, in which three real data examples are presented.

2.7 Real data examples

In this section, we present some numerical examples to illustrate the inferential results developed in the preceding sections. The first one is an application to the reliability example considered by Balakrishnan and Ling [2012a] which motivated the simulation scheme, and the other two are real applications to tumorigenicity experiments considered earlier by other authors.

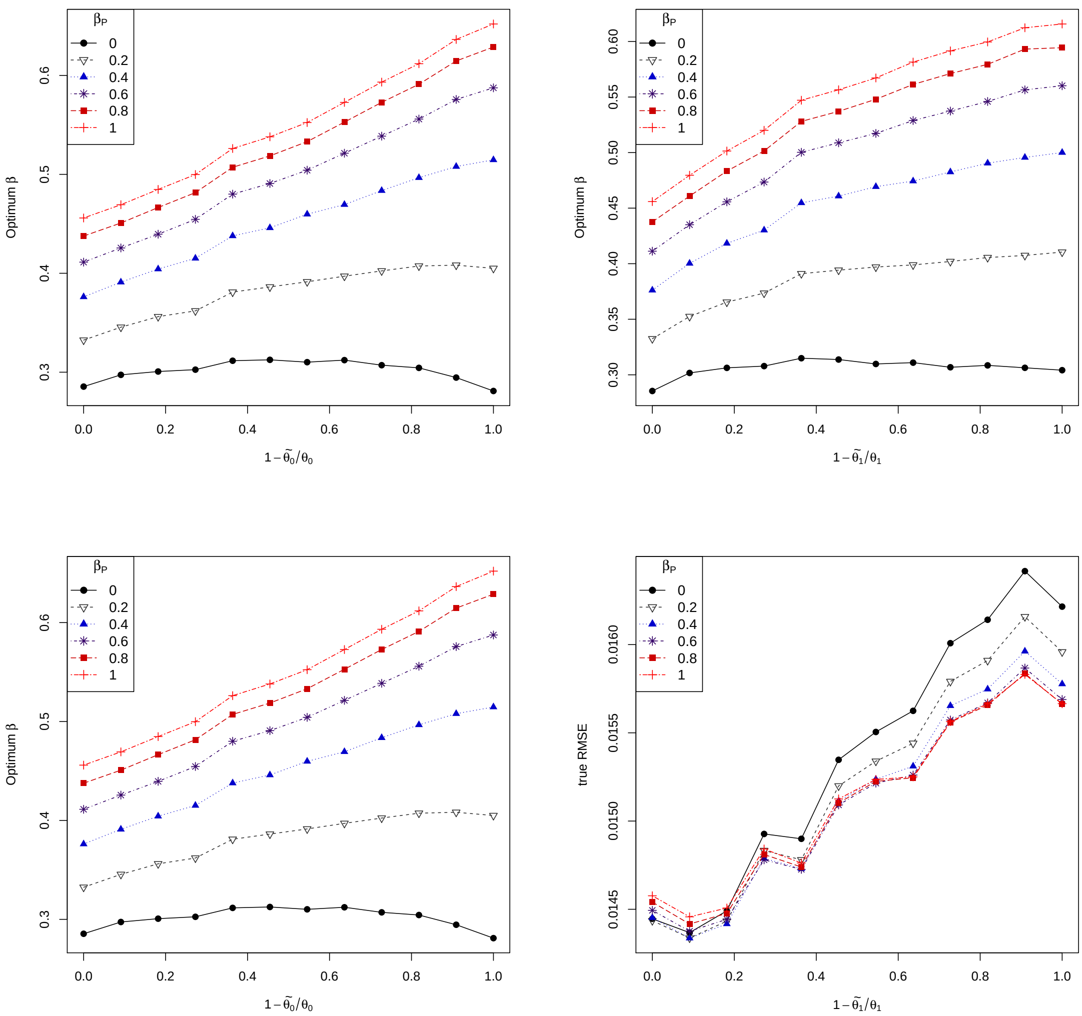

Figure 2.6.5: Exponential distribution at a simple stress level: average optimal values of β for different values of the pilot estimators and their corresponding RMSEs.

Algorithm 1 Algorithm for the data-driven selection of β

Goal: Optimal fitting of the model given any data set

Initialization: $\boldsymbol{\theta}_P = \widehat{\boldsymbol{\theta}}_{0.4}$ (empirical suggestion)

1: **for** each β in a grid of $[0, 1]$ **do**

2: Compute the estimated squared bias, $B_\beta = (\widehat{\boldsymbol{\theta}}_\beta - \widehat{\boldsymbol{\theta}}_{0.4})^T(\widehat{\boldsymbol{\theta}}_\beta - \widehat{\boldsymbol{\theta}}_{0.4})$.

3: Compute the total estimated variance, $V_\beta = \frac{1}{K}\text{trace}\left[\widehat{\boldsymbol{J}}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\widehat{\boldsymbol{K}}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\widehat{\boldsymbol{J}}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\right]$.

4: Compute the total estimated MSE, $\widehat{MSE}_\beta = B_\beta + V_\beta$.

5: **end for**

6: **return** $\beta_{opt} =$ arg min $\widehat{MSE}_\beta$.

7: **compute** $\widehat{\boldsymbol{\theta}}_{\beta_{opt}}$ as your final estimate with optimally chosen tuning parameter.

2.7.1 Reliability experiment (Balakrishnan and Ling, 2012)

In Balakrishnan and Ling [2012a], an example is presented, in which 90 devices were tested at temperatures $x_i \in \{35, 45, 55\}$, each with 10 units being detonated at times $IT_i \in \{10, 20, 30\}$, respectively. In this example, we have $I = 9$, and $K_i = 10$, $i = 1, \ldots, I$. The number of failures observed is summarized in Table 2.7.1. In this one-shot device testing experiment, there were in all 48 failures out a total of 90 tested devices.

Table 2.7.1: Reliability experiment.

i	IT_i	K_i	n_i	x_i
1	10	10	3	35
2	20	10	3	35
3	30	10	7	35
4	10	10	1	45
5	20	10	5	45
6	30	10	7	45
7	10	10	6	55
8	20	10	7	55
9	30	10	9	55

The weighted minimum DPD estimators of the parameters of the one-shot device model are considered. As tuning parameters, $\beta \in \{0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9,\ 1,\ 2,\ 3,\ 4\}$ are taken. The estimates of the reliability function at mission times (time points at which we are interested in the reliability of the unit) $t \in \{10, 20, 30\}$, namely $R(10, S_0, \widehat{\boldsymbol{\theta}}_\beta)$, $R(20; x_0, \widehat{\boldsymbol{\theta}}_\beta)$, $R(30; x_0, \widehat{\boldsymbol{\theta}}_\beta)$, respectively, are also computed, as well as the expected mean of the lifetime, namely,

$$E_\beta(T|x_0) = \frac{1}{\lambda_{x_0}(\widehat{\boldsymbol{\theta}}_\beta)} = \frac{1}{\widehat{\theta}_{0,\beta}e^{\widehat{\theta}_{1,\beta}x_0}},$$

under the normal operating temperature $x_0 = 25$.

Table 2.7.2 shows that the mean lifetime obtained by the MLE ($\beta = 0$) is greater than that obtained from the alternative weighted minimum DPD estimators. However, results for all considered choices of β seem to be quite similar. We now apply Algorithm 1 to the data. Optimal β results to be $\beta_{opt} = 0.62$ and the corresponding optimal parameters, $\widehat{\theta}_{0,opt} = 0.0049$ and $\widehat{\theta}_{1,opt} = 0.04696$.

2.7.2 ED01 Data

In 1974, the National Center for Toxicological Research made an experiment on 24000 female mice randomized to a control group or one of seven dose levels of a known carcinogen, called 2-Acetylaminofluorene (2-AAF). Table 1 in Lindsey and Ryan [1993] shows the results obtained when

Table 2.7.2: Reliability experiment: estimates of the model parameters, the reliability function at times $t \in \{10, 20, 30\}$, and mean lifetime at normal temperature of $25°C$

β	$\widehat{\theta}_{0,\beta}$	$\widehat{\theta}_{1,\beta}$	$R(10; 25, \widehat{\boldsymbol{\theta}}_\beta)$	$R(20; 25, \widehat{\boldsymbol{\theta}}_\beta)$	$R(30; 25, \widehat{\boldsymbol{\theta}}_\beta)$	$E_\beta(T\|25)$
0	0.00487	0.04732	0.85300	0.72761	0.62065	62.89490
0.1	0.00489	0.04722	0.85288	0.72741	0.62039	62.83953
0.2	0.00490	0.04714	0.85277	0.72722	0.62016	62.79031
0.3	0.00491	0.04706	0.85268	0.72706	0.61995	62.74654
0.4	0.00492	0.04700	0.85260	0.72693	0.61978	62.70965
0.5	0.00493	0.04695	0.85253	0.72681	0.61963	62.67944
0.6	0.00494	0.04690	0.85247	0.72671	0.61950	62.65188
0.7	0.00495	0.04687	0.85246	0.72669	0.61947	62.64457
0.8	0.00495	0.04683	0.85236	0.72651	0.61925	62.59732
0.9	0.00496	0.04681	0.85233	0.72646	0.61918	62.58398
1	0.00496	0.04681	0.85239	0.72656	0.61931	62.61131
2	0.00496	0.04679	0.85231	0.72644	0.61915	62.57739
3	0.00494	0.04687	0.85255	0.72684	0.61966	62.68584
4	0.00491	0.04700	0.85292	0.72748	0.62048	62.85869

the highest dose level (150 parts per million) was administered. The original study considered four different outcomes: Number of animals dying tumour free (DNT) and with tumour (DWT), and sacrificed without tumour (SNT) and with tumour (SWT), summarized over three time intervals at 12, 18 and 33 months. A total of 3355 mice were involved in the experiment. We make an analysis combining SNT and SWT as the sacrificed group ($r = 0$); and denoting the cause of DNT as natural death ($r = 1$), and the cause of DWT as death due to cancer ($r = 2$). This modified data are presented in Table 2.7.3. Here, $x = 0$ refers to the control group and $x = 1$ is the test group.

Table 2.7.3: ED01 experiment: number of mice sacrificed ($r = 0$) and died without tumour ($r = 1$) and with tumour ($r = 2$)

i	IT_i	K_i	x_i	$n_{i,r=0}$	$n_{i,r=1}$	$n_{i,r=2}$
1	12	145	0	115	22	8
2	12	175	1	110	49	16
3	18	830	0	780	42	8
4	18	620	1	540	54	26
5	33	960	0	675	200	85
6	33	625	1	510	64	51

The weighted minimum DPD estimators of the model parameters and the corresponding estimates of mean lifetimes are presented in Table 2.7.4. Here, we distinguish between sacrifice or nature death ($r = 0, 1$) and death due to cancer ($r = 2$). Note that, in the model under consideration in this chapter, only one possible failure cause is considered, so both estimations are computed separately.

Figure 2.7.1 shows the total estimated mean lifetimes for the control group ($x_0 = 0$) and the test group ($x_0 = 1$), computed as

$$E_\beta(T|x_0) = \frac{1}{\lambda^*_{x_0}(\widehat{\boldsymbol{\theta}}_\beta) + \lambda^{**}_{x_0}(\widehat{\boldsymbol{\theta}}_\beta)} = \frac{1}{\widehat{\theta}^*_{0,\beta} e^{\widehat{\theta}^*_{1,\beta} x_0} + \widehat{\theta}^{**}_{0,\beta} e^{\widehat{\theta}^{**}_{1,\beta} x_0}},$$

the estimators with $\beta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, show a reduction when the carcinogenic drug is administered, but the other ones, $\beta \in \{0, 0.8, 0.9, 1\}$, do not show this behavior. Thus, in

Table 2.7.4: ED01 experiment: estimates of the model parameters and expected lifetimes

	Sacrificed/ death without tumor				Death with tumor			
β	$\widehat{\theta}^*_{0,\beta}$	$\widehat{\theta}^*_{1,\beta}$	$E^*_\beta(T\|0)$	$E^*_\beta(T\|1)$	$\widehat{\theta}^{**}_{0,\beta}$	$\widehat{\theta}^{**}_{1,\beta}$	$E^{**}_\beta(T\|0)$	$E^{**}_\beta(T\|1)$
0	0.00594	-0.12980	168.333	191.665	0.00216	0.27620	463.425	351.582
0.1	0.00702	0.09355	142.352	129.639	0.00250	0.32870	399.794	287.795
0.2	0.00698	0.06495	143.302	134.290	0.00250	0.31173	400.433	293.189
0.3	0.00703	0.00999	142.253	140.840	0.00249	0.29613	401.393	298.513
0.4	0.00690	0.00998	145.019	143.578	0.00249	0.27957	401.602	303.655
0.5	0.00677	0.00998	147.662	146.195	0.00249	0.26421	401.839	308.537
0.6	0.00666	0.00998	150.085	148.594	0.00283	0.00997	353.925	350.414
0.7	0.00682	-0.06678	146.635	156.763	0.00249	0.23702	401.985	317.157
0.8	0.00680	-0.08753	147.020	160.468	0.00279	0.00997	358.642	355.083
0.9	0.00679	-0.10530	147.321	163.680	0.00278	0.00997	360.357	356.781
1	0.00678	-0.11980	147.546	166.324	0.00277	0.00995	361.607	358.028

this case, we observe that the first weighted minimum DPD estimators give a more meaningful result in the context of the laboratory experiment than, in particular, the MLE ($\beta = 0$).

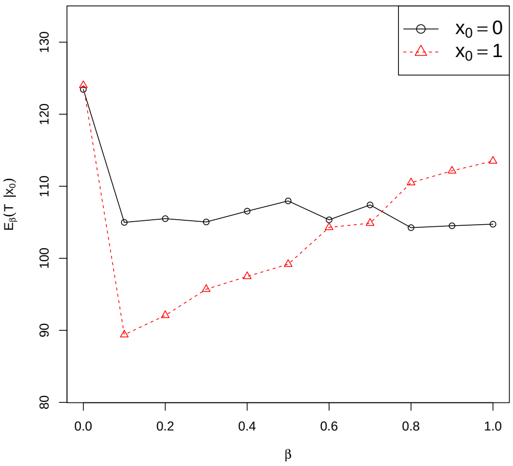

Figure 2.7.1: ED01 experiment: estimated mean lifetimes, for different values of the tuning parameter.

Let us apply the ad-hoc procedure for the choice of the optimal tuning parameter presented in Algorithm 1 to the ED01 data (2.7.5). Is important to notice that this ad-hoc choice of β does not depend on the results of data analysis and expert knowledge. In this sense, we see that, once the optimal values of the parameters are obtained, expected lifetimes in control group are seen to be higher than in the group to which the carcinogen is applied, which is a result that is consistent in the context of the experiment studied.

2.7.3 Benzidine Dihydrochloride data

The benzidine dihydrochloride experiment was also conducted at the National Center for Toxicological Research to examine the incidence in mice of liver tumours induced by the drug, and studied by Lindsey and Ryan [1993]. The inspection times used on the mice were 9.37, 14.07 and 18.7 months. In Table 2.7.6, the numbers of mice sacrificed ($r = 0$), died without tumour ($r = 1$) and died with tumour ($r = 2$), are shown, for two different doses of drug: 60 parts per million

Table 2.7.5: ED01 experiment: optimal choice of the estimators

Sacrificed/ death without tumor			Death with tumor				
β^*_{opt}	$\widehat{\theta}^*_{0,opt}$	$\widehat{\theta}^*_{1,opt}$	β^{**}_{opt}	$\widehat{\theta}^{**}_{0,opt}$	$\widehat{\theta}^{**}_{1,opt}$	$E_{\beta,opt}(T\|0)$	$E_{\beta,opt}(T\|1)$
0.24	0.00711	0.00998	0	0.00216	0.27620	107.827	99.681

($x = 1$) and 400 parts per million ($x = 2$). As in the previous example, we consider as "failures" the mice died due to cancer.

Table 2.7.6: Benzidine Dihydrochloride experiment: number of mice sacrificed ($r = 0$) and died without tumour ($r = 1$) and with tumour ($r = 2$)

i	IT_i	K_i	x_i	$n_{i,r=0}$	$n_{i,r=1}$	$n_{i,r=2}$
1	9.37	72	1	70	2	0
2	9.37	25	2	22	3	0
3	14.07	49	1	48	1	0
4	14.07	35	2	14	4	17
5	18.7	46	1	35	4	7
6	18.7	11	2	1	1	9

Table 2.7.7 shows the weighted minimum DPD estimators of the model parameters and the corresponding estimates of mean lifetimes. Although some differences are observed in the results for different values of the tuning parameter, in all the cases, the mean lifetime shows a reduction when the carcinogenic drug is administered. The optimal values are computed and presented in Table 2.7.8.

Table 2.7.7: Benzidine Dihydrochloride data: estimates of model parameters and expected lifetimes

	Death without tumor				Death with tumor			
β	$\widehat{\theta}^*_{0,\beta}$	$\widehat{\theta}^*_{1,\beta}$	$E^*_\beta(T\|1)$	$E^*_\beta(T\|2)$	$\widehat{\theta}^{**}_{0,\beta}$	$\widehat{\theta}^{**}_{1,\beta}$	$E_\beta(T\|1)$	$E_\beta(T\|2)$
0	0.00114	1.03606	309.9401	109.9828	0.00029	2.41598	154.1152	21.91462
0.1	0.00137	0.88718	301.3767	124.1119	0.00034	2.43535	139.4039	19.19976
0.2	0.00141	0.86736	298.6438	125.4474	0.00036	2.41128	136.2217	19.05452
0.3	0.00144	0.85118	297.0690	126.8120	0.00037	2.39392	134.5945	19.08241
0.4	0.00148	0.83279	294.5364	128.0470	0.00039	2.37048	132.5912	19.16100
0.5	0.00151	0.81685	292.6733	129.3098	0.00040	2.35318	130.3661	19.05474
0.6	0.00154	0.80124	290.7204	130.4671	0.00042	2.33504	128.7609	19.09932
0.7	0.00157	0.78793	289.2526	131.5478	0.00044	2.31249	126.9138	19.13369
0.8	0.00160	0.77577	287.6854	132.4363	0.00045	2.29739	125.3146	19.09985
0.9	0.00163	0.76485	286.3007	133.2454	0.00046	2.29251	124.5972	19.09096
1	0.00165	0.75493	284.8778	133.9056	0.00047	2.27947	123.1959	19.05346

Table 2.7.8: Benzidine Dihydrochloride experiment: optimal choice of the estimators

Sacrificed/ death without tumor			Death with tumor				
β^*_{opt}	$\widehat{\theta}^*_{0,opt}$	$\widehat{\theta}^*_{1,opt}$	β^{**}_{opt}	$\widehat{\theta}^{**}_{0,opt}$	$\widehat{\theta}^{**}_{1,opt}$	$E_{\beta,opt}(T\|1)$	$E_{\beta,opt}(T\|2)$
0.30	0.00143	0.85118	0	0.00029	2.41598	150.859	22.60977

Chapter 3

Robust inference for one-shot device testing under exponential distribution with multiple stress factors

3.1 Introduction

As pointed out in Section 1.4.1, to assess the reliability of one-shot devices, ALTs are frequently preformed to reduce the time to failure so that enough life data can be obtained in a reasonable period of time. Although this can be done through a single stress factor, aging is usually induced in devices by various accelerating factors such as temperature, pressure, humidity, and voltage simultaneously. Even though a single-stress test at a very high stress level may attain aging within limited time, a multiple-stress accelerated life test would enable us to achieve the same without requiring any of the stress factors to be set at very high levels. If maintaining a stress factor at high stress level for testing purposes is expensive, one could introduce several stress factors set at slightly elevated stress levels, causing more devices to fail than would under a single-stress test. For this reason, a multiple-stress model becomes better suited for the prediction of lifetimes of products, subjected to electrical, thermal, and mechanical stresses; see, for example, Srinivas and Ramu [1992] and Bartnikas and Morin [2004]. In Balakrishnan and Ling [2012b], an EM algorithm for developing inference is developed, based on one-shot device testing data under the exponential distribution when there are multiple stress factors.

Section 3.1.1 generalizes the model presented in Chapter 2 to the case of multiple stresses, and defines the weighted minimum DPD estimators. In particular, in this chapter, the failure time of the devices is assumed to follow an exponential distribution, following the notation in Section 3.1.2. The rest of the chapter is organized as follows: in Section 3.2, the estimating equations for the weighted minimum DPD estimators and their asymptotic distribution is developed, and their robustness is studied through the influence function study. In Section 3.3, robust Wald-type tests are provided for testing linear hypotheses. Finally, an extensive simulation study is carried out and some numerical examples are presented in Section 3.4 and Section 3.5, respectively.

The results of this Chapter have been published in the form of a paper (Balakrishnan et al. [2020c]).

3.1.1 One-shot device Inference with multiple stresses

Let us suppose now, that the devices are stratified into I testing conditions and that in the i-th testing condition K_i units are tested with J types of stress factors being maintained at certain levels, and the conditions of those units are then observed at pre-specified inspection times IT_i, for $i = 1, \ldots, I$. In the i-th test group, the number of failures, n_i, is collected. The data thus observed can be summarized as in Table 3.1.1.

Table 3.1.1: Data on one-shot devices testing at multple stress levels and collected at different inspection times

Condition	Inspection Time	Devices	Failures	Covariates		
				Stress 1	$\cdots$	Stress J
1	IT_1	K_1	n_1	x_{11}	$\cdots$	x_{1J}
2	IT_2	K_2	n_2	x_{21}	$\cdots$	x_{2J}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
I	IT_I	K_I	n_I	x_{I1}	$\cdots$	x_{IJ}

In this setting, we consider that the density and distribution functions are given, respectively, by $f(t;\boldsymbol{x}_i,\boldsymbol{\theta})$ and $F(t;\boldsymbol{x}_i,\boldsymbol{\theta})$, where $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{iJ})^T$ is the vector of stresses associated to the test condition i $(i = 1, \ldots, I)$, and $\boldsymbol{\theta} \in \Theta = \mathbb{R}^S$ is the model parameter vector (S will depend on the distribution associated to the model). The reliability function is denoted by $R(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = 1 - F(t;\boldsymbol{x}_i,\boldsymbol{\theta})$.

Assuming independent observations, the likelihood function based on the observed data, presented in Table 3.1.1, is given by

$$\mathcal{L}(n_1,\ldots,n_I;\boldsymbol{\theta}) \propto \prod_{i=1}^{I} F^{n_i}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}) R^{K_i-n_i}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}), \tag{3.1}$$

and the corresponding MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, will be obtained by maximizing in $\boldsymbol{\theta}$ the equation (3.1) or, equivalently, its logarithm. This is, $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \log \mathcal{L}(n_1,\ldots,n_I;\boldsymbol{\theta})$.

The empirical and theoretical probability vectors are given, respectively, by

$$\widehat{\boldsymbol{p}}_i = (\widehat{p}_{i1}, \widehat{p}_{i2})^T, \quad i = 1,\ldots,I, \tag{3.2}$$

and

$$\boldsymbol{\pi}_i(\boldsymbol{\theta}) = (\pi_{i1}(\boldsymbol{\theta}), \pi_{i2}(\boldsymbol{\theta}))^T, \quad i = 1,\ldots,I, \tag{3.3}$$

with $\widehat{p}_{i1} = \frac{n_i}{K_i}$, $\widehat{p}_{i2} = 1 - \frac{n_i}{K_i}$, $\pi_{i1}(\boldsymbol{\theta}) = F(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})$ and $\pi_{i2}(\boldsymbol{\theta}) = R(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})$. Let us consider the weighted Kullback divergence given in equation (2.6) between probability vectors (3.2) and (3.3). Following Theorem 2.2, it is straightforward that the MLE can be obtained as its minimization

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})), \tag{3.4}$$

where $K = \sum_{i=1}^{I} K_i$.

Based on this idea, we can now define the weighted minimum DPD estimators for the one-shot device model with multiple stresses

Definition 3.1 *Let us consider the framework in Table 3.1.1, we can define the weighted minimum DPD estimator for $\boldsymbol{\theta}$ as*

$$\widehat{\boldsymbol{\theta}}_\beta = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{I} \frac{K_i}{K} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})), \quad \text{for } \beta > 0,$$

where $d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta}))$ is given in (2.9), and $\widehat{\boldsymbol{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\theta}))$ are given in (3.2) and (3.3), respectively. For $\beta = 0$, we have the MLE, $\widehat{\boldsymbol{\theta}}$ defined in (3.4).

3.1.2 The Exponential Distribution

As a generalization of the model presented in the previous chapter, we shall assume here, that the true lifetime follows an exponential distribution with unknown failure rate $\lambda_i(\boldsymbol{\theta})$, related to the stress factor $\boldsymbol{x}_i$ in loglinear form as

$$\lambda_i(\boldsymbol{\theta}) = \exp(\boldsymbol{x}_i^T \boldsymbol{\theta}),$$

where $\boldsymbol{x}_i = (x_{i0}, x_{i1}, \ldots, x_{iJ})^T$, and $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_J)^T$. Thus, here $\Theta = \mathbb{R}^{J+1}$. The corresponding density function and distribution function are, respectively,

$$f(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \lambda_i(\boldsymbol{\theta}) \exp\{-\lambda_i(\boldsymbol{\theta})t\} = \exp(\boldsymbol{x}_i^T \boldsymbol{\theta}) \exp\{-\exp(\boldsymbol{x}_i^T \boldsymbol{\theta})t\} \tag{3.5}$$

and

$$F(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - \exp\{-\lambda_i(\boldsymbol{\theta})t\} = 1 - \exp\{-t \exp(\boldsymbol{x}_i^T \boldsymbol{\theta})\}. \tag{3.6}$$

On the other hand, the reliability at time t and the mean lifetime under normal operating conditions $\boldsymbol{x}_i$ are given by

$$R(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - F(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \exp\left(-t \exp\left(\boldsymbol{x}_i^T \boldsymbol{\theta}\right)\right) \tag{3.7}$$

and

$$E[T_i] = \frac{1}{\lambda_i} = \exp\left(-\boldsymbol{x}_i^T \boldsymbol{\theta}\right).$$

3.2 Weighted minimum DPD estimator

In this section, we will first obtain the estimating equations for the unknown parameter $\boldsymbol{\theta}$ and the asymptotic distribution of the weighted minimum DPD estimators.

3.2.1 Estimation and asymptotic distribution

Theorem 3.2 *For $\beta \geq 0$, the estimating equations are given by*

$$\sum_{i=1}^{I} (K_i F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) - n_i) f(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i \boldsymbol{x}_i \left(F^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + R^{\beta-1}(t; \boldsymbol{x}_i, \boldsymbol{\theta})\right) = \mathbf{0}_{J+1},$$

where $f(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta})$, $F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta})$ and $R(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta})$ are given, respectively, by (3.5), (3.6) and (3.7) and $\mathbf{0}_{J+1}$ is the null column vector of dimension $J+1$.

Proof. The estimating equations are given by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{I} \frac{K_i}{K} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \sum_{i=1}^{I} \frac{K_i}{K} \frac{\partial}{\partial \boldsymbol{\theta}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \mathbf{0}_{J+1},$$

with

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \left(\frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i2}^{\beta+1}(\boldsymbol{\theta})\right) - \frac{\beta+1}{\beta} \left(\widehat{p}_{i1} \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2} \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i2}^{\beta}(\boldsymbol{\theta})\right) \\
&= (\beta+1) \left(\pi_{i1}^{\beta}(\boldsymbol{\theta}) - \pi_{i2}^{\beta}(\boldsymbol{\theta}) - \widehat{p}_{i1} \pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \widehat{p}_{i2} \pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) \\
&= (\beta+1) \left((\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}) \pi_{i1}^{\beta-1}(\boldsymbol{\theta}) - (\pi_{i2}(\boldsymbol{\theta}) - \widehat{p}_{i2}) \pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) \\
&= (\beta+1) \left((\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}) \pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + (\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}) \pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) \\
&= (\beta+1) (\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}).
\end{aligned} \tag{3.8}$$

Taking into account that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) = f(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i \boldsymbol{x}_i,$$

we obtain

$$\frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{I} \frac{K_i}{K} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \frac{\beta+1}{K} \sum_{i=1}^{I} (K_i \pi_{i1}(\boldsymbol{\theta}) - n_i) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right) f(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i \boldsymbol{x}_i,$$

and then the required result follows. ■

Theorem 3.3 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter $\boldsymbol{\theta}$. The asymptotic distribution of the weighted minimum DPD estimator, $\widehat{\boldsymbol{\theta}}_\beta$, is given by*

$$\sqrt{K} \left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0 \right) \xrightarrow[K \to \infty]{\mathcal{L}} \mathcal{N} \left(\boldsymbol{0}_{J+1}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \boldsymbol{K}_\beta(\boldsymbol{\theta}^0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i}^{I} \frac{K_i}{K} \boldsymbol{x}_i \boldsymbol{x}_i^T f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2 \left(F^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) \right), \tag{3.9}$$

$$\begin{aligned} \boldsymbol{K}_\beta(\boldsymbol{\theta}) = & \sum_{i=1}^{I} \frac{K_i}{K} \boldsymbol{x}_i \boldsymbol{x}_i^T f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2 F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) R(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) \\ & \times \left(F^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + R^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) \right)^2. \end{aligned} \tag{3.10}$$

Proof. From Ghosh et al. [2013], it is known that

$$\sqrt{K} \left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0 \right) \xrightarrow[K \to \infty]{\mathcal{L}} \mathcal{N} \left(\boldsymbol{0}_{J+1}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \boldsymbol{K}_\beta(\boldsymbol{\theta}^0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \sum_{j=1}^{2} \frac{K_i}{K} \boldsymbol{u}_{ij}(\boldsymbol{\theta}) \boldsymbol{u}_{ij}^T(\boldsymbol{\theta}) \pi_{ij}^{\beta+1}(\boldsymbol{\theta}),$$

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \sum_{j=1}^{2} \frac{K_i}{K} \boldsymbol{u}_{ij}(\boldsymbol{\theta}) \boldsymbol{u}_{ij}^T(\boldsymbol{\theta}) \pi_{ij}^{2\beta+1}(\boldsymbol{\theta}) - \sum_{i=1}^{I} \frac{K_i}{K} \boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta}) \boldsymbol{\xi}_{i,\beta}^T(\boldsymbol{\theta}),$$

with

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta}) = \sum_{j=1}^{2} \boldsymbol{u}_{ij}(\boldsymbol{\theta}) \pi_{ij}^{\beta+1}(\boldsymbol{\theta}),$$

$$\boldsymbol{u}_{ij}(\boldsymbol{\theta}) = \frac{\partial \log \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\pi_{ij}(\boldsymbol{\theta})} \frac{\partial \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (-1)^{j+1} \frac{f(IT_i; \boldsymbol{\theta}, \boldsymbol{x}_i) IT_i}{\pi_{ij}(\boldsymbol{\theta})} \boldsymbol{x}_i, .$$

Because $\boldsymbol{u}_{ij}(\boldsymbol{\theta}) \boldsymbol{u}_{ij}^T(\boldsymbol{\theta}) = \boldsymbol{x}_i \boldsymbol{x}_i^T \frac{f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2}{\pi_{ij}^2(\boldsymbol{\theta})}$, we have

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} \boldsymbol{x}_i \boldsymbol{x}_i^T f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2 \sum_{j=1}^{2} \pi_{ij}^{\beta-1}(\boldsymbol{\theta}).$$

In a similar manner

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta}) \boldsymbol{\xi}_{i,\beta}^T(\boldsymbol{\theta}) = \boldsymbol{x}_i \boldsymbol{x}_i^T f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2 \left(\sum_{j=1}^{2} (-1)^{j+1} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \right)^2$$

and

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} \boldsymbol{x}_i \boldsymbol{x}_i^T f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2 \left[\sum_{j=1}^{2} \pi_{ij}^{2\beta-1}(\boldsymbol{\theta}) - \left(\sum_{j=1}^{2} (-1)^{j+1} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \right)^2 \right].$$

Since

$$\sum_{j=1}^{2} \pi_{ij}^{2\beta-1}(\boldsymbol{\theta}) - \left(\sum_{j=1}^{2} (-1)^{j+1} \pi_{ij}^{\beta}(\boldsymbol{\theta}) \right)^2 = \pi_{i1}(\boldsymbol{\theta})\pi_{i2}(\boldsymbol{\theta}) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right)^2,$$

it holds

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} \boldsymbol{x}_i \boldsymbol{x}_i^T f^2(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) IT_i^2 \pi_{i1}(\boldsymbol{\theta})\pi_{i2}(\boldsymbol{\theta}) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right)^2.$$

■

3.2.2 Study of the Influence Function

In Section 2.5, the IF of the weighted minimum DPD estimators under the exponential distribution with one stress factor were computed. Same procedures are followed to obtain the IF for the case of multiple stress factors:

Proposition 3.4 *Let us consider the one-shot device testing under the exponential distribution with multiple stress factors. The IF with respect to the $k-$th observation of the i_0-th group is given by*

$$\begin{aligned} IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}) =& \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \frac{K_{i_0}}{K} f(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0) IT_{i_0} \boldsymbol{x}_{i_0} \\ & \times \left(F^{\beta-1}(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0) + R^{\beta-1}(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0) \right) \left(F(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0) - \Delta_{t_{i_0}}^{(1)} \right), \end{aligned} \tag{3.11}$$

where $\Delta_{t_{i_0},k}^{(1)}$ is the degenerating function at point (t_{i_0}, k).

Proposition 3.5 *Let us consider the one-shot device testing under the exponential distribution with multiple stress factors. The IF with respect to all the observations is given by*

$$\begin{aligned} IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}) =& \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \sum_{i=1}^{I} \frac{K_i}{K} f(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}^0) IT_i \boldsymbol{x}_i \\ & \times \left(F^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}^0) + R^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}^0) \right) \left(F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}^0) - \Delta_{t_i}^{(1)} \right), \end{aligned} \tag{3.12}$$

where $\Delta_{t_i}^{(1)} = \sum_{k=1}^{K_i} \Delta_{t_{i,k}}^{(1)}$.

3.3 Wald-type tests

In this section, we develop robust Wald-type tests, presenting also some results in relation with their power function. IF of the proposed Wald-type tests are finally computed.

3.3.1 Definition and study of the level

Let us consider the function $\boldsymbol{m} : \mathbb{R}^{J+1} \longrightarrow \mathbb{R}^r$, where $r \leq J+1$. Then, $\boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{0}_r$ represents a composite null hypothesis. We assume that the $(J+1) \times r$ matrix

$$\boldsymbol{M}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{m}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

exists and is continuous in $\boldsymbol{\theta}$ with rank $\boldsymbol{M}(\boldsymbol{\theta}) = r$. For testing

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \text{ against } H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0, \tag{3.13}$$

where $\boldsymbol{\Theta}_0 = \left\{\boldsymbol{\theta} \in \mathbb{R}^{J+1} : \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{0}_r\right\}$, we can consider the following Wald-type test statistics

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) = K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta), \tag{3.14}$$

where $\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta) = \boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)$ and $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are as in (3.9) and (3.10), respectively.

In the following theorem, we present the asymptotic distribution of $W_K(\widehat{\boldsymbol{\theta}}_\beta)$.

Theorem 3.6 *The asymptotic null distribution of the proposed Wald-type test statistics, given in Equation (3.14), is a chi-squared (χ^2) distribution with r degrees of freedom. This is,*

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \chi^2_r.$$

Proof. Let $\boldsymbol{\theta}^0 \in \Theta_0$ be the true value of parameter $\boldsymbol{\theta}$. It is then clear that

$$\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) = \boldsymbol{m}(\boldsymbol{\theta}^0) + \boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0) + o_p\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right\|\right) = \boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0) + o_p\left(K^{-1/2}\right).$$

But, $\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\boldsymbol{0}_{J+1}, \boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\right)$. Therefore, we have

$$\sqrt{K}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\boldsymbol{0}_r, \boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)$$

and taking into account that $rank(\boldsymbol{M}\left(\boldsymbol{\theta}^0\right)) = r$, we obtain

$$K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \chi^2_r.$$

But, $\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}$ is a consistent estimator of $\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}$ and, therefore,

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \chi^2_r.$$

■

Based on Theorem 3.6, we will reject the null hypothesis in (3.13) if

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) > \chi^2_{r,\alpha}, \tag{3.15}$$

where $\chi^2_{r,\alpha}$ is the upper percentage point of order α of χ^2_r distribution.

3.3.2 Some results relating to the power function

In many cases, the power function of this testing procedure cannot be derived explicitly. In the following theorem, we present a useful asymptotic result for approximating the power function of the Wald-type test statistics given in (3.14). We shall assume that $\boldsymbol{\theta}^* \notin \boldsymbol{\Theta}_0$ is the true value of the parameter such that

$$\widehat{\boldsymbol{\theta}}_\beta \underset{K\to\infty}{\xrightarrow{P}} \boldsymbol{\theta}^*, \tag{3.16}$$

and we denote $\ell_\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \boldsymbol{m}^T(\boldsymbol{\theta}_1)\left(\boldsymbol{M}^T(\boldsymbol{\theta}_2)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)\boldsymbol{M}(\boldsymbol{\theta}_2)\right)^{-1}\boldsymbol{m}(\boldsymbol{\theta}_1)$. We then have the following result.

Theorem 3.7 *We have*

$$\sqrt{K}\left(\ell_\beta\left(\widehat{\boldsymbol{\theta}}_\beta,\boldsymbol{\theta}^*\right)-\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\mathcal{N}(0,\sigma^2_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)),$$

where

$$\sigma^2_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)=\left.\frac{\partial\ell_\beta\left(\boldsymbol{\theta},\boldsymbol{\theta}^*\right)}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\boldsymbol{\Sigma}_\beta\left(\boldsymbol{\theta}^*\right)\left.\frac{\partial\ell_\beta\left(\boldsymbol{\theta},\boldsymbol{\theta}^*\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}.$$

Proof. Under the assumption that

$$\widehat{\boldsymbol{\theta}}_\beta\underset{K\to\infty}{\xrightarrow{P}}\boldsymbol{\theta}^*,$$

the asymptotic distribution of $\ell_\beta\left(\widehat{\boldsymbol{\theta}}_1,\widehat{\boldsymbol{\theta}}_2\right)$ coincides with the asymptotic distribution of $\ell_\beta\left(\widehat{\boldsymbol{\theta}}_1,\boldsymbol{\theta}^*\right)$. A first-order Taylor expansion of $\ell_\beta\left(\widehat{\boldsymbol{\theta}}_\beta,\boldsymbol{\theta}\right)$ at $\widehat{\boldsymbol{\theta}}_\beta$, around $\boldsymbol{\theta}^*$, yields

$$\left(\ell_\beta\left(\widehat{\boldsymbol{\theta}}_\beta,\boldsymbol{\theta}^*\right)-\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)=\left.\frac{\partial\ell_\beta\left(\boldsymbol{\theta},\boldsymbol{\theta}^*\right)}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\left(\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}^*\right)+o_p(K^{-1/2}).$$

Now, the result readily follows since

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}^*\right)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\mathcal{N}\left(\mathbf{0}_{J+1},\boldsymbol{\Sigma}_\beta\left(\boldsymbol{\theta}^*\right)\right).$$

■

Remark 3.8 *Using Theorem 3.7, we can give an approximation for the power function of the Wald-type test statistics in $\boldsymbol{\theta}^*$, satisfying (3.16), as follows:*

$$\begin{aligned}\pi_{W,K}\left(\boldsymbol{\theta}^*\right)&=\Pr\left(W_K\left(\widehat{\boldsymbol{\theta}}_\beta\right)>\chi^2_{r,\alpha}\right)\\&=\Pr\left(K\left(\ell_\beta\left(\widehat{\boldsymbol{\theta}}_\beta,\boldsymbol{\theta}^*\right)-\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)>\chi^2_{r,\alpha}-K\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)\\&=\Pr\left(\frac{\sqrt{K}\left(\ell_\beta\left(\widehat{\boldsymbol{\theta}}_\beta,\boldsymbol{\theta}^*\right)-\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)}{\sigma_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)}>\frac{1}{\sigma_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)}\left(\frac{\chi^2_{r,\alpha}}{\sqrt{K}}-\sqrt{K}\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)\right)\\&=1-\Phi_K\left(\frac{1}{\sigma_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)}\left(\frac{\chi^2_{r,\alpha}}{\sqrt{K}}-\sqrt{K}\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)\right)\end{aligned}$$

for a sequence of distributions functions $\Phi_K(x)$ tending uniformly to the standard normal distribution $\Phi(x)$. It is clear that

$$\lim_{K\to\infty}\pi_{W,K}\left(\boldsymbol{\theta}^*\right)=1,$$

*i.e., the Wald-type test statistics are consistent in the sense of Fraser (**?**).*

The above approximation of the power function of the Wald-type test statistics can be used to obtain the sample size K necessary in order to achieve a pre-fixed power $\pi_{W,K}\left(\boldsymbol{\theta}^*\right)=\pi_0$, say. To do so, it is necessary to solve the equation

$$\pi_0=1-\Phi_K\left(\frac{1}{\sigma_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)}\left(\frac{\chi^2_{r,\alpha}}{\sqrt{K}}-\sqrt{K}\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\right)\right).$$

The solution, in K, of the above equation yields $\widehat{K}_\beta=\left[\widehat{K}^*_\beta\right]+1$, where

$$\widehat{K}^*_\beta=\frac{\widehat{A}_\beta+\widehat{B}_\beta+\sqrt{\widehat{A}_\beta(\widehat{A}_\beta+2\widehat{B}_\beta)}}{2\ell^2_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)},$$

with $\widehat{A}_\beta=\sigma^2_{W_K,\beta}\left(\boldsymbol{\theta}^*\right)\left(\Phi^{-1}\left(1-\pi_0\right)\right)^2$ and $\widehat{B}_\beta=2\ell_\beta\left(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*\right)\chi^2_{r,\alpha}$.

We may also find an approximation of the power of $W_K\left(\widehat{\boldsymbol{\theta}}_\beta\right)$ at an alternative hypotheses close to the null hypothesis. Let $\boldsymbol{\theta}^K \in \Theta - \Theta_0$ be a given alternative and let $\boldsymbol{\theta}^0$ be the element in Θ_0 closest to $\boldsymbol{\theta}^K$ in the sense of Euclidean distance. A first possibility to introduce contiguous alternative hypotheses is to consider a fixed $\boldsymbol{d} \in \mathbb{R}^p$ and to permit $\boldsymbol{\theta}^K$ to move towards $\boldsymbol{\theta}^0$ as K increases in such a way that

$$H_{1,K} : \boldsymbol{\theta}^K = \boldsymbol{\theta}^0 + K^{-1/2}\boldsymbol{d}. \tag{3.17}$$

A second approach is to relax the condition $\boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{0}_r$ defining Θ_0. Let $\boldsymbol{d}^* \in \mathbb{R}^r$ and consider the sequence $\left\{\boldsymbol{\theta}^K\right\}$ of parameters moving towards $\boldsymbol{\theta}^0$ such that

$$H_{1,K}^* : \boldsymbol{m}(\boldsymbol{\theta}^K) = K^{-1/2}\boldsymbol{d}^*. \tag{3.18}$$

Note that a Taylor series expansion of $\boldsymbol{m}(\boldsymbol{\theta}^K)$ around $\boldsymbol{\theta}^0$ yields

$$\boldsymbol{m}(\boldsymbol{\theta}^K) = \boldsymbol{m}(\boldsymbol{\theta}^0) + \boldsymbol{M}^T(\boldsymbol{\theta}^0)\left(\boldsymbol{\theta}^K - \boldsymbol{\theta}^0\right) + o\left(\left\|\boldsymbol{\theta}^K - \boldsymbol{\theta}^0\right\|\right). \tag{3.19}$$

Upon substituting $\boldsymbol{\theta}^K = \boldsymbol{\theta}^0 + K^{-1/2}\boldsymbol{d}$ in (3.19) and taking into account that $\boldsymbol{m}(\boldsymbol{\theta}^0) = \boldsymbol{0}_r$, we get

$$\boldsymbol{m}(\boldsymbol{\theta}^K) = K^{-1/2}\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{d} + o\left(\left\|\boldsymbol{\theta}^K - \boldsymbol{\theta}^0\right\|\right),$$

so that the equivalence in the limit is obtained for $\boldsymbol{d}^* = \boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{d}$.

Theorem 3.9 *We have the following results under both versions of the alternative hypothesis:*

i) $W_K\left(\widehat{\boldsymbol{\theta}}_\beta\right) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \chi^2_r\left(\boldsymbol{d}^T\boldsymbol{M}(\boldsymbol{\theta}^0)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{d}\right)$ *under* $H_{1,K}$ *in (3.17);*

ii) $W_K\left(\widehat{\boldsymbol{\theta}}_\beta\right) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \chi^2_r\left(\boldsymbol{d}^{*T}\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{d}^*\right)$ *under* $H_{1,K}^*$ *in (3.18).*

Proof. A Taylor series expansion of $\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta)$ around $\boldsymbol{\theta}^K$ yields

$$\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) = \boldsymbol{m}(\boldsymbol{\theta}^n) + \boldsymbol{M}^T(\boldsymbol{\theta}^K)(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^K) + o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^K\right\|\right).$$

We have

$$\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) = K^{-1/2}\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{d} + \boldsymbol{M}^T(\boldsymbol{\theta}^K)(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^K) + o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^K\right\|\right) + o\left(\left\|\boldsymbol{\theta}^K - \boldsymbol{\theta}^0\right\|\right).$$

As

$$\sqrt{K}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^K) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(\boldsymbol{0}_{J+1}, \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0))$$

and $\sqrt{K}\left(o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^K\right\|\right) + o\left(\left\|\boldsymbol{\theta}^K - \boldsymbol{\theta}^0\right\|\right)\right) = o_p(1)$, we have

$$\sqrt{K}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{d}, \boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right).$$

From the relationship $\boldsymbol{d}^* = \boldsymbol{M}(\boldsymbol{\theta}^0)^T\boldsymbol{d}$, if $\boldsymbol{m}(\boldsymbol{\theta}^n) = K^{-1/2}\boldsymbol{d}^*$, we can observe that

$$\sqrt{K}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\boldsymbol{d}^*, \boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right).$$

In the present case, the quadratic form is $W_K\left(\widehat{\boldsymbol{\theta}}_\beta\right) = \boldsymbol{Z}^T\boldsymbol{Z}$ with

$$\boldsymbol{Z} = \sqrt{K}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1/2}$$

and

$$\boldsymbol{Z} \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1/2}\boldsymbol{M}(\boldsymbol{\theta}^0)^T\boldsymbol{d}, \boldsymbol{I}_r\right),$$

where $\boldsymbol{I}_r$ is the identity matrix of order r. Hence, the application of the result is immediate and the noncentrality parameter is given by

$$\boldsymbol{d}^T\boldsymbol{M}(\boldsymbol{\theta}^0)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{M}(\boldsymbol{\theta}^0)^T\boldsymbol{d} = \boldsymbol{d}^{*T}\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{d}^*.$$

■

Remark 3.10 *If we consider* $\boldsymbol{d} = \sqrt{K}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)$, *with* $\boldsymbol{\theta}^*$ *satisfying (3.16), we have*

$$\boldsymbol{\theta}^K = \boldsymbol{\theta}^0 + K^{-1/2}K^{1/2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0) = \boldsymbol{\theta}^*$$

and therefore we can use the asymptotic result in relation to $H_{1,K}$ *in order to get an approximation of the power function in* $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

3.3.3 Study of the Influence Function

Next, we study the robustness of the proposed Wald-type test statistics. In our context, the functional associated with the Wald-type tests, evaluated at $\boldsymbol{U}_\beta(\boldsymbol{G})$ is given by

$$W_K(\boldsymbol{U}_\beta(\boldsymbol{G})) = K\boldsymbol{m}^T(\boldsymbol{U}_\beta(\boldsymbol{G}))\left(\boldsymbol{M}^T(\boldsymbol{U}_\beta(\boldsymbol{G}))\boldsymbol{\Sigma}(\boldsymbol{U}_\beta(\boldsymbol{G}))\boldsymbol{M}(\boldsymbol{U}_\beta(\boldsymbol{G}))\right)^{-1}\boldsymbol{m}(\boldsymbol{U}_\beta(\boldsymbol{G})).$$

The IF with respect to the $k-$th observation of the i_0-th group of observations, of the functional associated with the Wald-type test statistics for testing the composite null hypothesis in (3.13), is then given by

$$IF(t_{i_0,k}, W_K, F_{\boldsymbol{\theta}^0}) = \left.\frac{\partial W_K(F_{\boldsymbol{\theta}_\varepsilon^{i_0}})}{\partial \varepsilon}\right|_{\varepsilon=0^+} = 0.$$

It, therefore, becomes necessary to consider the second-order IF, as presented in the following result.

Theorem 3.11 *The second-order IF of the functional associated with the Wald-type test statistics, with respect to the $k-$th observation of the i_0-th group of observations, is given by*

$$\begin{aligned} IF_2(t_{i_0,k}, W_K, F_{\boldsymbol{\theta}^0}) &= \left.\frac{\partial^2 W_K(F_{\boldsymbol{\theta}_\varepsilon^{i_0}})}{\partial \varepsilon^2}\right|_{\varepsilon=0^+} \\ &= 2\ IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})\boldsymbol{m}^T(\boldsymbol{\theta}^0)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\boldsymbol{\theta}^0)IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}), \end{aligned}$$

where $IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})$ *is given in (3.11).*

Similarly, for all the indices,

Theorem 3.12 *The second-order IF of the functional associated with the Wald-type test statistics, with respect to all the observations, is given by*

$$\begin{aligned} IF_2(\boldsymbol{t}, W_K, F_{\boldsymbol{\theta}^0}) &= \left.\frac{\partial^2 W_K(F_{\boldsymbol{\theta}_\varepsilon})}{\partial \varepsilon^2}\right|_{\varepsilon=0^+} \\ &= 2\ IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})\boldsymbol{m}^T(\boldsymbol{\theta}^0)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\boldsymbol{\theta}^0)IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}), \end{aligned}$$

where $IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})$ *is given in (3.12).*

Note that the second-order influence functions of the proposed Wald-type tests are quadratic functions of the corresponding IFs of the weighted minimum DPD estimator for any type of contamination.

3.4 Simulation Study

In this section, Monte Carlo simulations of size 2,000 were carried out to examine the behavior of the weighted minimum DPD estimators of the model parameters under the exponential lifetimes assumption.

Based on the simulation experiment proposed by Balakrishnan and Ling [2012b], we considered the devices to have exponential lifetimes subjected to two types of stress factors at two different conditions each, the first one at levels 55 and 70 and the second one at levels 85 and 100, and tested at three different inspection times $IT = \{2, 5, 8\}$. Thus, we can consider a table, such as in Table 3.1.1, with $I = 12$ rows corresponding to each of the 12 testing conditions. To evaluate the robustness of the weighted minimum DPD estimators, we have studied the behavior of this model under the consideration of an outlying cell (for example, the last one) in this table.

3.4.1 Weighted minimum DPD estimators

We carried out a simulation study to compare the behavior of some weighted minimum DPD estimators with respect to the MLEs of the parameters in the one-shot device model under the exponential distribution with multiple stresses. In order to evaluate the performance of the proposed weighted minimum DPD estimators, as well as the MLEs, we consider the RMSEs. The model has been examined under $(\theta_0, \theta_1, \theta_2) = (-6.5, 0.03, 0.03)$, different samples sizes $K_i \in [40, 200]$, $i = 1, \ldots, 12$, and different degrees of contamination. The estimates have been computed with values of the tuning parameter $\beta \in \{0, 0.2, 0.4, 0.6, 0.8\}$.

In the top of Figure 3.4.1, efficiency of weighted minimum DPD estimators is measured under different samples sizes K_i with pure data (left) and contaminated data (right) where the observations in the $i = 12$ testing condition have been generated under $(\theta_0, \theta_1, \tilde{\theta}_2) = (-6.5, 0.03, 0.025)$. Same experiment is carried out by contaminating the last two testing conditions (top left of Figure 3.4.4). The efficiency is then measured for the last-cell-contaminated data, generated under $(\theta_0, \tilde{\theta}_1, \tilde{\theta}_2) = (-6.5, 0.025, 0.025)$ (top right of Figure 3.4.4). In the case of pure data, the MLE (at $\beta = 0$) presents the most efficient behavior having the least RMSE for each sample size, while weighted minimum DPD estimators with larger β have slightly larger RMSEs. For the contaminated data, the behavior of the weighted minimum DPD estimators is almost the opposite; the best behavior (least RMSE) is obtained for larger values of β. In both cases, as expected, the RMSEs decrease as the sample size increases.

The efficiency is also studied for different degrees of contamination of the parameters θ_1 (left) and θ_2 (right), as displayed in the top of Figure 3.4.2. Here, $K_i = 100$ and the degree of contamination is given by $4(1 - \frac{\tilde{\theta}_j}{\theta_j}) \in [0, 1]$ with $j \in \{1, 2\}$. In both cases, we can see how the MLEs and the weighted minimum DPD estimators with small values of tuning parameter β present the smallest RMSEs for weak outliers, i.e., when the degree of contamination is close to 0 ($\tilde{\theta}_j$ is close to θ_j). On the other hand, large values of tuning parameter β result in the weighted minimum DPD estimators having the smallest RMSEs, for medium and strong outliers, i.e., when the degree of contamination away from 0 ($\tilde{\theta}_j$ is not close to θ_j).

In view of the results achieved, we note that the MLE is very efficient when there are no outliers, but highly non-robust when outliers are present in the data. On the other hand, the weighted minimum DPD estimators with moderate values of the tuning parameter β exhibit a little loss of efficiency when there are no outliers, but at the same time a considerable improvement in robustness is achieved when there are outliers in the data. Actually, these values of the tuning parameter β are the most appropriate ones for the estimators of the parameters in the model following the robustness theory: To improve in a considerable way the robustness of the estimators, a small amount of efficiency needs to be compromised.

3.4.2 Wald-type tests

Let us now empirically evaluate the robustness of the weighted minimum DPD estimator based Wald-type tests for the model. The simulation is performed with the same model as in Table 3.1.1, where $(\theta_0, \theta_1, \theta_2) = (-6.5, 0.03, 0.03)$. We first study the observed level (measured as the proportion of test statistics exceeding the corresponding chi-square critical value) of the test under the true null hypothesis $H_0 : \theta_2 = 0.03$ against the alternative $H_1 : \theta_2 \neq 0.03$. In the middle of Figure 3.4.1, these levels are plotted for different values of the samples sizes, for pure data (left) and for contaminated data ($\tilde{\theta}_2 = 0.025$, right). Same experiment is carried out by contaminating the last two testing conditions (middle left of Figure 3.4.4). The empirical levels are then measured for the last-cell-contaminated data, generated under $(\theta_0, \tilde{\theta}_1, \tilde{\theta}_2) = (-6.5, 0.025, 0.025)$ (middle right of Figure 3.4.4). In the middle of Figure 3.4.2, the degree of contamination for both θ_1 and θ_2 is changed with a fixed value of $K_i = 100$. Notice that when the pure data are considered, all the observed levels are quite close to the nominal level of 0.05. In the case of contaminated data, the level of the classical Wald test (at $\beta = 0$) as well as the proposed Wald-type tests with small β break down, while the weighted minimum DPD estimator based Wald-type tests for moderate and large values of β provide greater stability in their levels.

To investigate the power robustness of these tests (obtained in a similar manner), we change the true data generating parameter value to be $\theta_2 = 0.035$ and the resulting empirical powers are plotted in the bottom of Figures 3.4.1 and 3.4.2 and in the bottom left of Figure 3.4.4) (when the last two cells are contaminated). The empirical powers are then measured for the last-cell-contaminated data, generated under $(\theta_0, \tilde{\theta}_1, \tilde{\theta}_2) = (-6.5, 0.035, 0.025)$ (bottom right of Figure 3.4.4). Again, the classical Wald test (at $\beta = 0$) presents the best behavior under the pure data, while the Wald-type tests with larger $\beta > 0$ lead to better stability in the case of contaminated samples. Same tests are also evaluated with a higher/lower value of reliability ($\theta_0 = -6$) obtaining the same conclusions as detailed above (see Figure 3.4.3).

These results show the poor behavior in terms of robustness of the Wald test based on the MLEs of the parameters of one-shot devices under the exponential model with multiple stresses. Additionally, the robustness properties of the Wald-type test statistics based on the weighted minimum DPD estimators with large values of the tuning parameter β are often better as they maintain both level and power in a stable manner.

3.5 Real data examples

In this Section, two numerical examples are presented to illustrate the model and the estimators developed in the preceding sections.

3.5.1 Mice Tumor Toxicological data

As mentioned earlier, current status data with covariates, which generally occur in the area of survival analysis, can be seen as one-shot device testing data with stress factors and we therefore apply here the methods developed in the preceding sections to a real data from a study in toxicology. These data, originally reported by Kodell and Nelson [1980] (Table 1) and recently analyzed by Balakrishnan and Ling [2013], are taken from the National Center for Toxicological Research and consisted of 1816 mice, of which 553 had tumors, involving the strain of offspring (F1 or F2), gender (females or males), and concentration of benzidine dihydrochloride (60 ppm, 120 ppm, 200 ppm or 400 ppm) as the stress factors. The F1 strain consisted of offspring from matings of BALB/c males to C57BL/6 females, while the F2 strain consisted of offspring from non-brother-sister matings of the Fl progeny. For each testing condition, the numbers of mice tested and the numbers of mice that developed tumors were all recorded. Note that we consider mice with tumors as those that died of tumors, sacrificed with tumors, and died of competing risks with liver tumors.

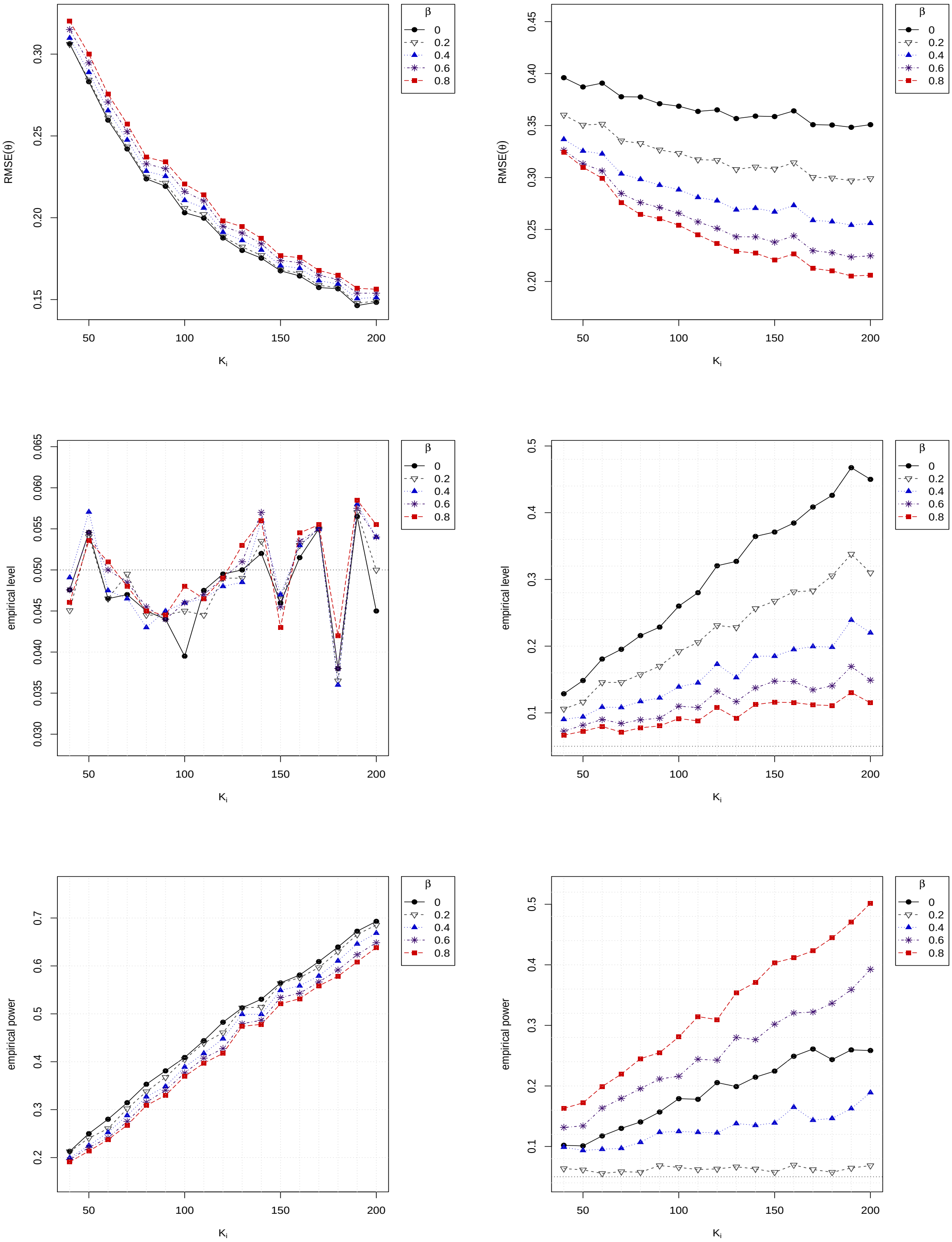

Figure 3.4.1: Exponential distribution at multiple stress levels: RMSEs (top panel) of the weighted minimum DPD estimators of $\boldsymbol{\theta}$, the simulated levels (middle panel) and powers (bottom panel) of the Wald-type tests under the pure data (left) and under the contaminated data (right).

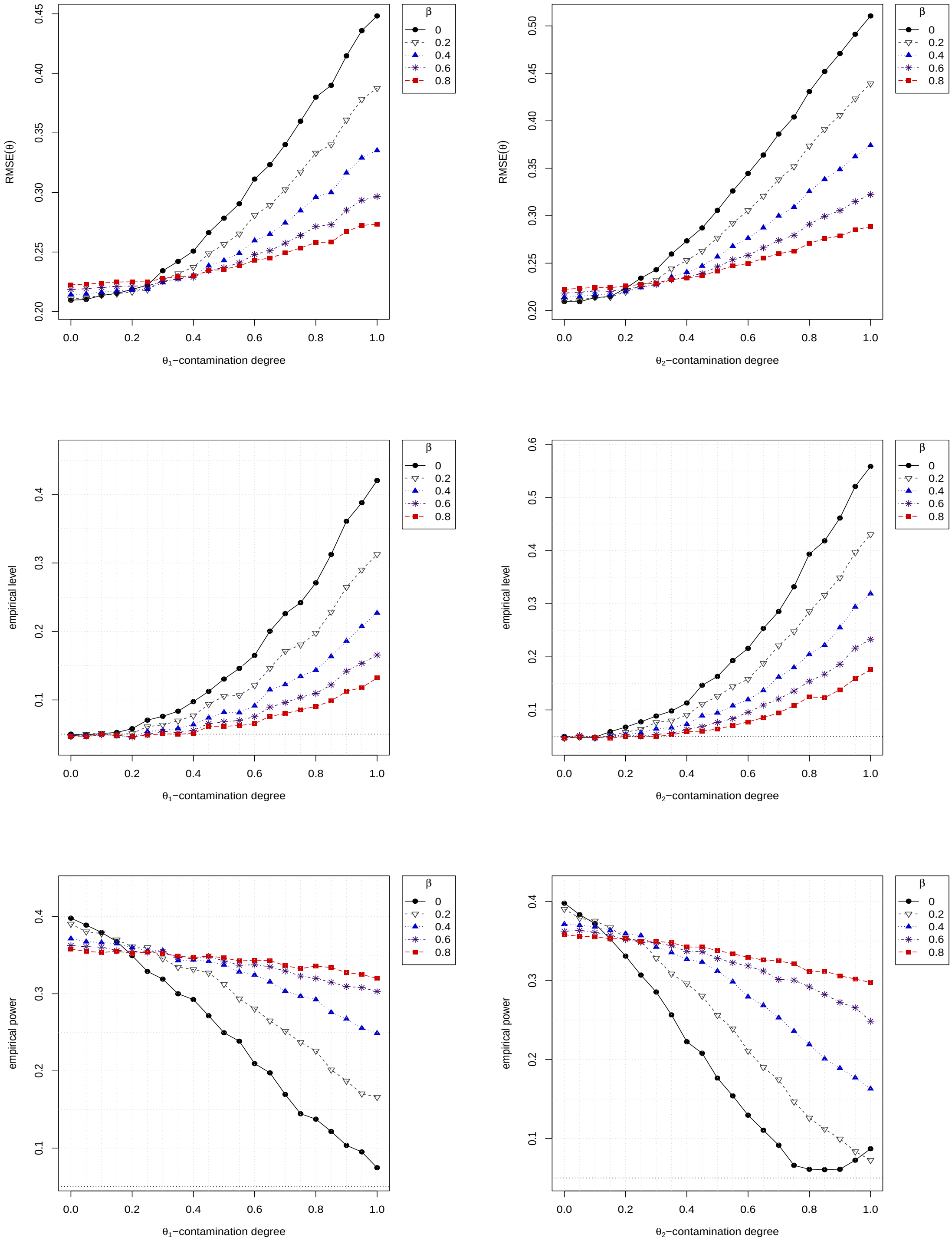

Figure 3.4.2: Exponential distribution at multiple stress levels: RMSEs (top panel) of the weighted minimum DPD estimators of $\boldsymbol{\theta}$, the simulated levels (middle panel) and powers (bottom panel) of the Wald-type tests under the θ_1-contaminated data (left) and under the θ_2-contaminated data(right).

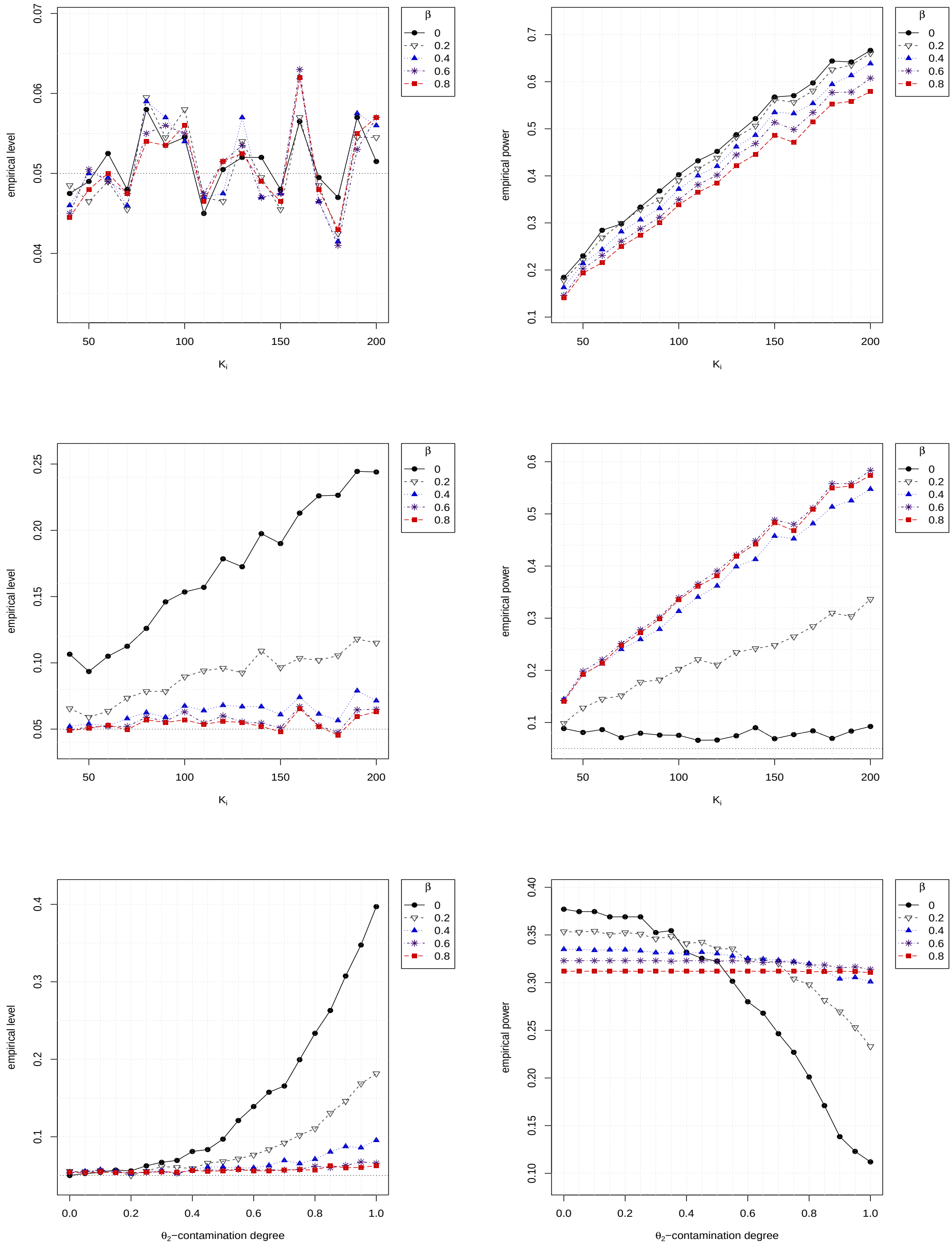

Figure 3.4.3: Exponential distribution at multiple stress levels: Empirical levels (left) and powers (right) under the pure data and under the contaminated data when parameter $\theta_0 = -6$

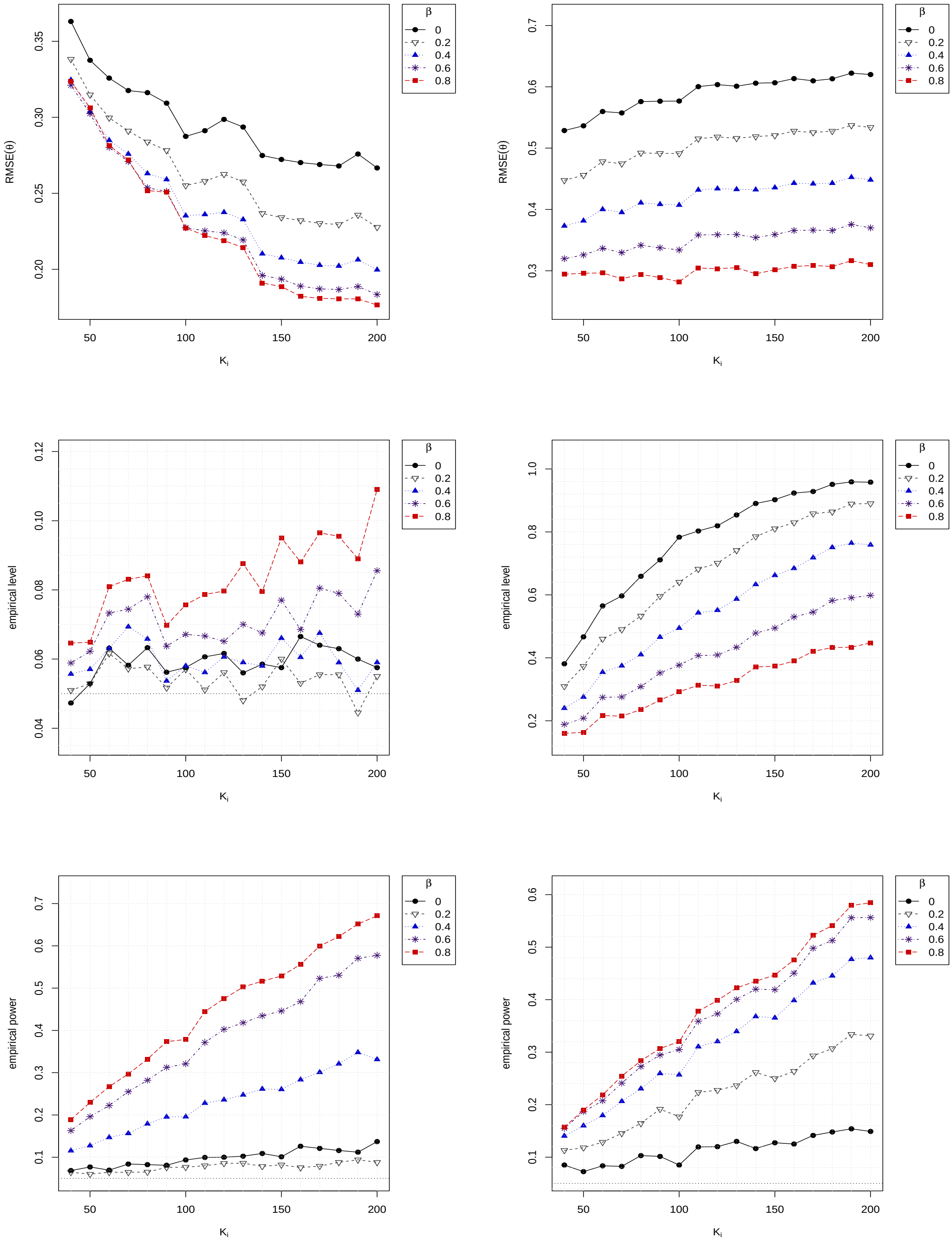

Figure 3.4.4: Exponential distribution at multiple stress levels: RMSEs (top panel), empirical levels (middle panel) and empirical powers (bottom panel) of two-cells contaminated data (left) and θ_1-θ_2-contaminated data (right), when parameter $\theta_0 = -6.5$.

Let θ_1, θ_2 and θ_3 denote the parameters corresponding to the covariates of strain of offspring, gender, and square root of concentration of the chemical of benzidine dihydrochloride in the exponential distribution given in (3.6). The weighted minimum DPD estimators with tuning parameter $\beta \in \{0, 0.2, 0.4, 0.6, 0.8\}$ were all computed and are presented in Table 3.5.1. Negative values for $\widehat{\theta}_1$ and $\widehat{\theta}_2$ indicate a greater resistance of F2 strain and male mice. As expected, a greater concentration of benzidine dihydrochloride is seen to decrease the expected lifetime.

Table 3.5.1: Mice Tumor Toxicological data: Point estimation under the exponential distribution at multiple stress levels

β	$\widehat{\theta}_0$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$
0	-4.452	-0.126	-1.201	0.133
0.2	-4.821	-0.195	-1.300	0.148
0.4	-4.784	-0.184	-1.291	0.145
0.6	-4.753	-0.176	-1.282	0.143
0.8	-4.731	-0.170	-1.275	0.141

3.5.2 Electric current data

These data (Balakrishnan and Ling [2012b]), presented in Table 3.5.2, consist of 120 one-shot devices that were divided into four accelerated conditions with higher-than-normal temperature and electric current, and inspected at three different times. By subjecting the devices to adverse conditions, we shorten the lifetimes, observing more failures in a clear example of an accelerated life test design. This numerical example also served as a basis for the Monte Carlo study carried out earlier in Section 3.4.

Table 3.5.2: Electric current data

i	IT_i	K_i	n_i	Temeperature (x_{i1})	Electric current (x_{i2})
1	2	10	0	55	70
2	2	10	4	55	100
3	2	10	4	85	70
4	2	10	7	85	100
5	5	10	4	55	70
6	5	10	7	55	100
7	5	10	8	85	70
8	5	10	8	85	100
9	8	10	3	55	70
10	8	10	9	55	100
11	8	10	9	85	70
12	8	10	10	55	100

The estimates of the model parameters are presented in Table 3.5.3, for different values of the tuning parameter $\boldsymbol{\beta}$. Reliability at different inspections times and normal testing conditions $\boldsymbol{x}_0 = (25, 35)$, as well as the mean lifetimes, are also presented. As expected, the reliability of the devices decrease when the inspection time increases. Figure 3.5.1 displays the estimated reliabilities at a pre-fixed inspection time, $t = 30$, for different values of temperature and electric current, and two different tuning parameters: $\beta = 0$ (MLE) and a high-moderate value $\beta = 0.6$. Let us denote $\widehat{R}_0^{ij}$ and $\widehat{R}_{0.6}^{ij}$ for the estimated reliability at temperature level i and electric current level j based on the weighted minimum DPD estimators with tuning parameter $\beta = 0$ and $\beta = 0.6$, which are represented in the top left and top right of Figure 3.5.1, respectively. As expected, they decrease when the testing conditions increase, becoming especially low for extreme testing levels. Left bottom of Figure 3.5.1 shows the differences between the two measures, that is, $\widehat{R}_{0.6}^{ij} - \widehat{R}_0^{ij}$,

Table 3.5.3: Electric current data: Point estimation of parameters and reliabilities at time $t \in \{10, 30, 60\}$ and mean lifetimes for different tuning parameters at normal conditions $\boldsymbol{x}_0 = (25, 35)$.

β	$\widehat{\theta}_0$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$R(10, 25, 35)$	$R(30, 25, 35)$	$R(60, 25, 35)$	$\widehat{T}$
0	−6.5128	0.0301	0.0340	0.9018	0.7334	0.5379	96.74
0.1	−6.6100	0.0308	0.0346	0.9069	0.7460	0.5565	102.38
0.2	−6.7178	0.0315	0.0354	0.9123	0.7594	0.5767	109.00
0.3	−6.8327	0.0323	0.0362	0.9178	0.7730	0.5975	116.51
0.4	−6.9549	0.0332	0.0370	0.9232	0.7868	0.6190	125.09
0.5	−7.0759	0.0340	0.0379	0.9282	0.7997	0.6395	134.21
0.6	−7.1920	0.0348	0.0387	0.9327	0.8115	0.6585	143.60
0.7	−7.2915	0.0355	0.0394	0.9364	0.8211	0.6742	152.17
0.8	−7.3740	0.0361	0.0400	0.9393	0.8287	0.6867	159.65
0.9	−7.4387	0.0365	0.0404	0.9415	0.8345	0.6964	165.79
1	−7.4869	0.0369	0.0407	0.9430	0.8387	0.7034	170.52

while right bottom of Figure 3.5.1 shows the standardized differences $(\widehat{R}^{ij}_{0.6} - \widehat{R}^{ij}_0)/\widehat{R}^{ij}_0$. While in absolute value the biggest differences are given for moderate values of temperature and current electricity (where reliabilities are higher), the most remarkable difference (that is measured with independence on the scale) is obtained for extreme conditions both of current and temperature. Note that these are the only cases when the estimated reliability based on the MLEs is higher than the one based on the weighted minimum DPD estimators with tuning parameter $\beta = 0.6$. Table 3.5.4 shows the estimated probabilities of the weighted minimum DPD estimators with different tuning parameters $\beta \in [0, 1]$, compared with the observed probabilities. Last row in Table 3.5.4 shows the estimated mean absolute error of each weighted minimum DPD estimator considered here, e^{β}_i. MLE ($\beta = 0$) seems, in general, to be one of the worst choices to predict each testing condition. In particular, we can say that weighted minimum DPD estimators with high or moderate value of the tuning parameter seem to have a better behavior than the MLEs when higher-than-normal testing conditions are considered, just as we observed a greater difference in terms of reliability (Figure 3.5.1).

Table 3.5.4: Electric current data: Estimated probabilities for different weighted minimum DPD estimators

		$\widehat{\pi}^{\beta}_i$										
i	$\frac{n_i}{K_i}$	$\beta = 0$	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.3$	$\beta = 0.4$	$\beta = 0.5$	$\beta = 0.6$	$\beta = 0.7$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 1$
1	0	0.154	0.152	0.150	0.148	0.146	0.144	0.142	0.141	0.139	0.138	**0.137**
2	0.4	0.338	0.340	0.343	0.346	0.348	0.351	0.354	0.356	0.358	0.359	**0.360**
3	0.4	0.371	0.373	0.376	0.378	0.381	0.384	0.387	0.389	0.391	0.393	**0.394**
4	0.7	0.681	0.691	**0.703**	0.715	0.728	0.740	0.752	0.761	0.769	0.776	0.780
5	0.4	**0.342**	0.338	0.335	0.331	0.327	0.322	0.319	0.315	0.312	0.310	0.309
6	0.7	0.644	0.647	0.650	0.654	0.657	0.661	0.664	0.667	0.669	0.671	**0.672**
7	0.8	0.686	0.689	0.692	0.695	0.699	0.703	0.706	0.709	0.711	0.713	**0.714**
8	0.8	**0.943**	0.947	0.952	0.957	0.961	0.965	0.969	0.972	0.974	0.976	0.977
9	0.3	0.488	0.484	0.479	0.474	0.469	0.464	0.459	0.454	0.451	0.448	**0.446**
10	0.9	0.808	0.811	0.814	0.817	0.820	0.823	0.825	0.828	0.830	0.831	**0.832**
11	0.9	0.843	0.846	0.848	0.851	0.854	0.856	0.859	0.861	0.863	0.864	**0.865**
12	1	0.990	0.991	0.992	0.993	0.994	0.995	0.996	0.997	0.997	0.997	**0.997**
	e^{β}_i	0.082	0.080	0.078	0.077	0.077	0.076	0.076	0.076	0.075	0.075	**0.075**

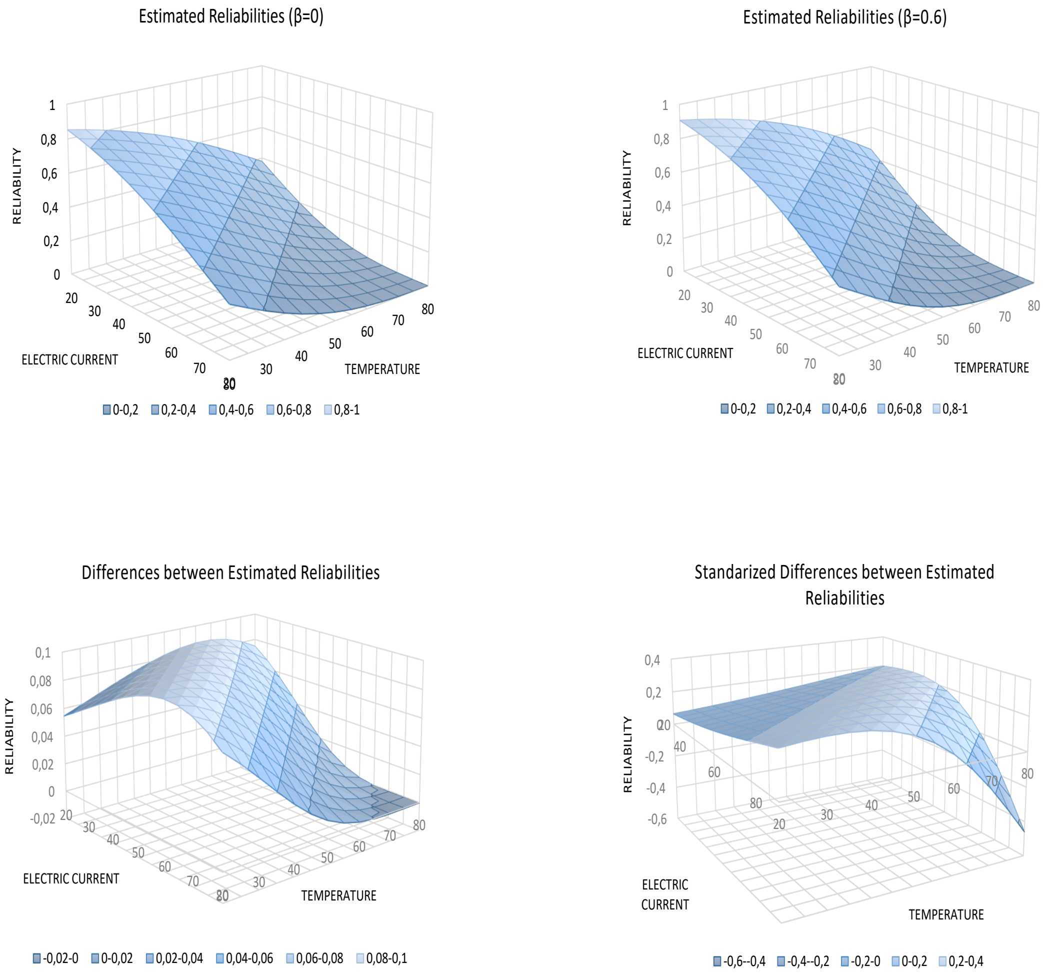

Figure 3.5.1: Electric current data: Estimated reliabilities based on weighted minimum DPD estimators with tuning parameters $\beta = 0$ (top left) and $\beta = 0.6$ (top right) and their differences (bottom left) and standardized differences (bottom right)

Chapter 4

Robust inference for one-shot device testing under gamma distribution

4.1 Introduction

Gamma distribution is commonly used for fitting lifetime data in reliability and survival studies due to its flexibility. Its hazard function can be increasing, decreasing, and constant. When the hazard function of gamma distribution is a constant, it corresponds to the exponential distribution. In addition to the exponential distribution, the gamma distribution also includes the Chi-square distribution as a special case. The gamma distribution has found a number of applications in different fields. For example, Husak et al. [2007] used it to describe monthly rainfall in Africa for the management of water and agricultural resources, as well as food reserves. Kwon and Frangopol [2010] assessed and predicted bridge fatigue reliabilities of two existing bridges, the Neville Island Bridge and the Birmingham Bridge, based on long-term monitoring data. They made use of log-normal, Weibull, and gamma distributions to estimate the mean and standard deviation of the stress range. Tseng et al. [2009] proposed an optimal step-stress accelerated degradation testing plan for assessing the lifetime distribution of products with longer lifetime based on a gamma process.

In this chapter, we extend the results of Chapter 3 by assuming that the lifetimes follow a gamma distribution. With this premise, weighted minimum DPD estimators, their estimating equations and asymptotic distribution are developed in Section 4.2. In this section, robust Wald-type tests are also presented. A simulation study is provided in Section 4.3 and a real example is presented in Section 4.4.

The results of this Chapter have been published in the form of a paper (Balakrishnan et al. [2019a]).

4.1.1 The gamma distribution

Let us denote by $\boldsymbol{\theta} = (a_0, \ldots, a_J, b_0, \ldots, b_J)^T$ the model parameter vector. We shall then assume that the lifetimes of the units, under the testing condition i, follow gamma distribution with corresponding probability density function and cumulative distribution function as

$$f(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{t^{\alpha_i - 1}}{\lambda_i^{\alpha_i}\Gamma(\alpha_i)} \exp\left(-\frac{t}{\lambda_i}\right), \; t > 0,$$

and

$$F(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \int_0^t \frac{y^{\alpha_i - 1}}{\lambda_i^{\alpha_i}\Gamma(\alpha_i)} \exp\left(-\frac{y}{\lambda_i}\right) dy, \; t > 0, \tag{4.1}$$

where $\alpha_i > 0$ and $\lambda_i > 0$ are, respectively, the shape and scale parameters at condition i, which we assume are related to the stress factors in log–linear forms as

$$\alpha_i = \exp\left\{\sum_{j=0}^{J} a_j x_{ij}\right\} \quad \text{and} \quad \lambda_i = \exp\left\{\sum_{j=0}^{J} b_j x_{ij}\right\},$$

with $x_{i0} = 1$ for all i. Let us denote by $R_T(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - F(t; \boldsymbol{x}_i, \boldsymbol{\theta})$ the reliability function, the probability that the unit lasts lifetime t.

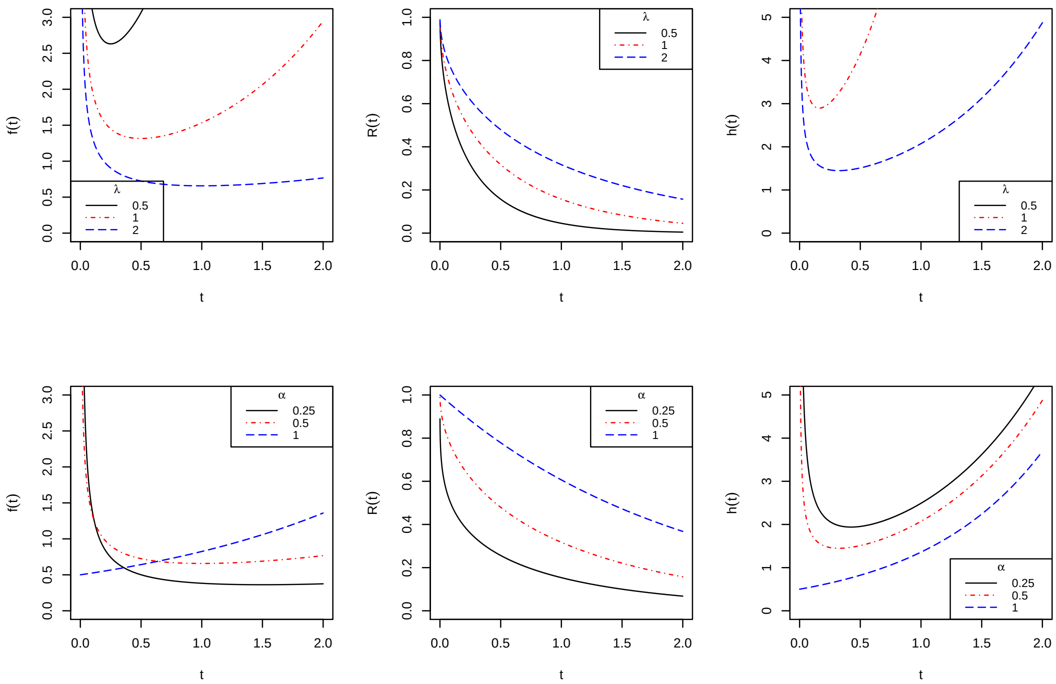

Figure 4.1.1: Gamma distributions for different values of shape and scale parameters.

4.2 Inference under the gamma distribution

Theorem 4.1 *For $\beta \geq 0$, the estimating equations are given by*

$$\sum_{i=1}^{I} l_i \left(K_i F\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) - n_i\right) \left(F^{\beta-1}\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) + \left(1 - F\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right)^{\beta-1}\right) \boldsymbol{x}_i = \boldsymbol{0}_{J+1},$$
$$\sum_{i=1}^{I} s_i \left(K_i F\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) - n_i\right) \left(F^{\beta-1}\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) + \left(1 - F\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right)^{\beta-1}\right) \boldsymbol{x}_i = \boldsymbol{0}_{J+1},$$

where

$$l_i = \alpha_i \left\{ -\Psi\left(\alpha_i\right) \pi_{i1}(\boldsymbol{\theta}) + \log\left(\frac{IT_i}{\lambda_i}\right) \pi_{i1}(\boldsymbol{\theta}) - \frac{\left(\frac{IT_i}{\lambda_i}\right)^{\alpha_i}}{\alpha_i^2 \Gamma(\alpha_i)} {}_2F_2\left(\alpha_i, \alpha_i; 1+\alpha_i, 1+\alpha_i; -\frac{IT_i}{\lambda_i}\right) \right\} \quad (4.2)$$

and

$$s_i = -f\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) IT_i, \quad (4.3)$$

where $F\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right)$ was given in (4.1). Here, ${}_nF_m(a_1, \ldots, a_n; b_1, \ldots, b_m; z)$ denotes the Gaussian hypergeometric function. For more details about the Gaussian hypergeometric function, one may refer to Seaborn [1991].

Proof. The estimating equations are given by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{I} \frac{K_i}{K} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \sum_{i=1}^{I} \frac{K_i}{K} \frac{\partial}{\partial \boldsymbol{\theta}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \mathbf{0}_{2(J+1)},$$

with

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \left(\frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i2}^{\beta+1}(\boldsymbol{\theta}) \right) - \frac{\beta+1}{\beta} \left(\widehat{p}_{i1} \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2} \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i2}^{\beta}(\boldsymbol{\theta}) \right) \\
&= (\beta+1) \left(\pi_{i1}^{\beta}(\boldsymbol{\theta}) - \pi_{i2}^{\beta}(\boldsymbol{\theta}) - \widehat{p}_{i1} \pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \widehat{p}_{i2} \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) \\
&= (\beta+1) \left(\left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right) \pi_{i1}^{\beta-1}(\boldsymbol{\theta}) - \left(\pi_{i2}(\boldsymbol{\theta}) - \widehat{p}_{i2}\right) \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) \\
&= (\beta+1) \left(\left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right) \pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right) \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) \\
&= (\beta+1) \left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right) \left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta}) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}). \qquad (4.4)
\end{aligned}$$

The required result follows taking into account that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) = (l_i \boldsymbol{x}_i^T, s_i \boldsymbol{x}_i^T)^T.$$

■

In the following theorem, the asymptotic distribution of the weighted minimum DPD estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_\beta$, is presented for one-shot device testing data under gamma lifetimes.

Theorem 4.2 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter $\boldsymbol{\theta}$. The asymptotic distribution of the weighted minimum DPD estimator, $\widehat{\boldsymbol{\theta}}_\beta$, is given by*

$$\sqrt{K} \left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0 \right) \xrightarrow[K \to \infty]{\mathcal{L}} \mathcal{N} \left(\mathbf{0}_{2(J+1)}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \boldsymbol{K}_\beta(\boldsymbol{\theta}^0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0) \right),$$

with

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} \boldsymbol{\Psi}_i \left(F^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}))^{\beta-1} \right), \qquad (4.5)$$

$$\begin{aligned}
\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} & \boldsymbol{\Psi}_i F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) \left(1 - F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta})\right) \\
& \times \left(F^{\beta-1}(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - F(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}))^{\beta-1} \right)^2, \qquad (4.6)
\end{aligned}$$

and

$$\boldsymbol{\Psi}_i = \begin{pmatrix} l_i^2 \boldsymbol{x}_i \boldsymbol{x}_i^T & l_i s_i \boldsymbol{x}_i \boldsymbol{x}_i^T \\ l_i s_i \boldsymbol{x}_i \boldsymbol{x}_i^T & s_i^2 \boldsymbol{x}_i \boldsymbol{x}_i^T \end{pmatrix},$$

with l_i and s_i as given in (4.2) and (4.3), respectively.

Proof. Let us denote

$$\begin{aligned}
\boldsymbol{u}_{ij}(\boldsymbol{\theta}) &= \left(\frac{\partial \log \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{a}}, \frac{\partial \log \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \right)^T = \left(\frac{1}{\pi_{ij}(\boldsymbol{\theta})} \frac{\partial \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{a}}, \frac{1}{\pi_{ij}(\boldsymbol{\theta})} \frac{\partial \pi_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{b}} \right)^T \\
&= \left(\frac{(-1)^{j+1}}{\pi_{ij}(\boldsymbol{\theta})} l_i \boldsymbol{x}_i, \frac{(-1)^{j+1}}{\pi_{ij}(\boldsymbol{\theta})} s_i \boldsymbol{x}_i \right)^T,
\end{aligned}$$

with l_i and s_i as given in (4.2) and (4.3), see Balakrishnan and Ling [2014a] for more details. Upon using Theorem 3.1 of Ghosh et al. [2013], we have

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}^0\right)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\mathcal{N}\left(\mathbf{0}_{2(J+1)},\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta})=\sum_{i=1}^{I}\sum_{j=1}^{2}\frac{K_i}{K}\boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta})\pi_{ij}^{\beta+1}(\boldsymbol{\theta}),$$

$$\boldsymbol{K}_\beta(\boldsymbol{\theta})=\left(\sum_{i=1}^{I}\sum_{j=1}^{2}\frac{K_i}{K}\boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta})\pi_{ij}^{2\beta+1}(\boldsymbol{\theta})-\sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})\boldsymbol{\xi}_{i,\beta}^T(\boldsymbol{\theta})\right),$$

with

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})=\sum_{j=1}^{2}\boldsymbol{u}_{ij}(\boldsymbol{\theta})\pi_{ij}^{\beta+1}(\boldsymbol{\theta})=(l_i\boldsymbol{x}_i,s_i\boldsymbol{x}_i)^T\sum_{j=1}^{2}(-1)^{j+1}\pi_{ij}^{\beta}(\boldsymbol{\theta}).$$

Now, for $\boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta})$, we have

$$\boldsymbol{u}_{ij}(\boldsymbol{\theta})\boldsymbol{u}_{ij}^T(\boldsymbol{\theta})=\frac{1}{\pi_{ij}^2(\boldsymbol{\theta})}\begin{pmatrix} l_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i & l_is_i\boldsymbol{x}_i^T\boldsymbol{x}_i\\ l_is_i\boldsymbol{x}_i^T\boldsymbol{x}_i & s_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i\end{pmatrix}=\frac{1}{\pi_{ij}^2(\boldsymbol{\theta})}\boldsymbol{\Psi}_i,$$

with

$$\boldsymbol{\Psi}_i=\begin{pmatrix} l_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i & l_is_i\boldsymbol{x}_i^T\boldsymbol{x}_i\\ l_is_i\boldsymbol{x}_i^T\boldsymbol{x}_i & s_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i\end{pmatrix}.$$

It then follows that

$$\boldsymbol{J}_\beta(\boldsymbol{\theta})=\sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Psi}_i\sum_{j=1}^{2}\pi_{ij}^{\beta-1}(\boldsymbol{\theta})=\sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Psi}_i\left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta})+\pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right).$$

In a similar manner,

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})\boldsymbol{\xi}_{i,\beta}^T(\boldsymbol{\theta})=\boldsymbol{\Psi}_i\left(\sum_{j=1}^{2}(-1)^{j+1}\pi_{ij}^{\beta}(\boldsymbol{\theta})\right)^2$$

and

$$\boldsymbol{K}_\beta(\boldsymbol{\theta})=\sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Psi}_i\left(\sum_{j=1}^{2}\pi_{ij}^{2\beta-1}(\boldsymbol{\theta})-\left(\sum_{j=1}^{2}(-1)^{j+1}\pi_{ij}^{\beta}(\boldsymbol{\theta})\right)^2\right).$$

Since

$$\sum_{j=1}^{2}\pi_{ij}^{2\beta-1}(\boldsymbol{\theta})-\left(\sum_{j=1}^{2}(-1)^{j+1}\pi_{ij}^{\beta}(\boldsymbol{\theta})\right)^2=\pi_{i1}(\boldsymbol{\theta})\pi_{i2}(\boldsymbol{\theta})\left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta})+\pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)^2,$$

we have

$$\boldsymbol{K}_\beta(\boldsymbol{\theta})=\sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Psi}_i\pi_{i1}(\boldsymbol{\theta})\pi_{i2}(\boldsymbol{\theta})\left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta})+\pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)^2.$$

■

Now, we present the IF of the proposed estimators:

Theorem 4.3 *Let us consider the one-shot device testing under the gamma distribution with multiple stress factors. The IF with respect to the k−th observation of the i_0−th group is given by*

$$\begin{aligned}IF(t_{i_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0})=&\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)(l_{i_0}\boldsymbol{x}_{i_0},s_{i_0}\boldsymbol{x}_{i_0})^T\\&\times\left(F^{\beta-1}(IT_{i_0};\boldsymbol{x}_{i_0},\boldsymbol{\theta}^0)+R^{\beta-1}(IT_{i_0};\boldsymbol{x}_{i_0},\boldsymbol{\theta}^0)\right)\left(F(IT_{i_0};\boldsymbol{x}_{i_0},\boldsymbol{\theta}^0)-\Delta_{t_{i_0}}^{(1)}\right),\end{aligned}\tag{4.7}$$

where $\Delta_{t_{i_0},k}^{(1)}$ is the degenerating function at point (t_{i_0},k).

Proof. Straightforward following results in Section 2.5. ■

Theorem 4.4 *Let us consider the one-shot device testing under the gamma distribution with multiple stress factors. The IF with respect to all the observations is given by*

$$
\begin{aligned}
IF(\boldsymbol{t},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^0}) =&\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\sum_{i=1}^{I}\frac{K_i}{K}(l_i\boldsymbol{x}_i,s_i\boldsymbol{x}_i)^T \\
&\times\left(F^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}^0)+R^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}^0)\right)\left(F(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}^0)-\Delta_{t_i}^{(1)}\right),
\end{aligned} \tag{4.8}
$$

where $\Delta_{t_i}^{(1)}=\sum_{k=1}^{K_i}\Delta_{t_{i,k}}^{(1)}$.

Proof. Straightforward following results in Section 2.5. ■

4.2.1 Wald-type tests

From Theorem 4.2, where the asymptotic distribution of the proposed weighted minimum DPD estimators is presented, we can develop Wald-type tests for testing composite null hypotheses.

Let us consider the function $\boldsymbol{m}:\mathbb{R}^{2(J+1)}\longrightarrow\mathbb{R}^r$, where $r\leq 2(J+1)$. Then, $\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{0}_r$ represents a composite null hypothesis. We assume that the $2(J+1)\times r$ matrix

$$\boldsymbol{M}(\boldsymbol{\theta})=\frac{\partial\boldsymbol{m}^T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

exists and is continuous in $\boldsymbol{\theta}$ and with rank $\boldsymbol{M}(\boldsymbol{\theta})=r$. For testing

$$H_0:\boldsymbol{\theta}\in\boldsymbol{\Theta}_0 \text{ against } H_1:\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0, \tag{4.9}$$

where $\boldsymbol{\Theta}_0=\left\{\boldsymbol{\theta}\in\mathbb{R}^{2(J+1)}:\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{0}_r\right\}$, we can consider the following Wald-type test statistics

$$W_K(\widehat{\boldsymbol{\theta}}_\beta)=K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta), \tag{4.10}$$

where $\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)=\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta(\widehat{\boldsymbol{\theta}}_\beta)$ and $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are as in (4.5) and (4.6), respectively.

In the following theorem, we present the asymptotic distribution of $W_K(\widehat{\boldsymbol{\theta}}_\beta)$.

Theorem 4.5 *The asymptotic null distribution of the proposed Wald-type test statistics, given in Equation (4.10), is a chi-squared (χ^2) distribution with r degrees of freedom. This is,*

$$W_K(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\chi^2_r.$$

Proof. Let $\boldsymbol{\theta}^0\in\Theta_0$ be the true value of parameter $\boldsymbol{\theta}$. $\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}^0\right)\underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\left(\boldsymbol{0}_{2(J+1)},\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\right)$. Therefore, under H_0, we have

$$\sqrt{K}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\left(\boldsymbol{0}_r,\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)$$

and taking into account that $rank(\boldsymbol{M}(\boldsymbol{\theta}^0))=r$, we obtain

$$K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\chi^2_r.$$

But, $\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}$ is a consistent estimator of $\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}$ and, therefore, $W_K(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\chi^2_r$. ■

Based on Theorem 4.5, we will reject the null hypothesis in (4.9) if $W_K(\widehat{\boldsymbol{\theta}}_\beta)>\chi^2_{r,\alpha}$, where $\chi^2_{r,\alpha}$ is the upper percentage point of order α of χ^2_r distribution.

Results concerning the power function of the proposed Wald-type tests could be obtained in a similar manner to previous chapters.

As happened under the exponential distribution, it becomes necessary to consider the second-order IF of the proposed Wald-type tests, as presented in the following result

Theorem 4.6 *The second-order IF of the functional associated with the Wald-type test statistics, with respect to the $k-th$ observation of the i_0-th group of observations, is given by*

$$\begin{aligned}&IF_2(t_{i_0,k}, W_K, F_{\boldsymbol{\theta}^0})\\&= 2\ IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})\boldsymbol{m}^T(\boldsymbol{\theta}^0)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\boldsymbol{\theta}^0)IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}),\end{aligned}$$

where $IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})$ is given in (5.8).

Proof. Straightforward following results on Section 3.3.3. ■

Similarly, for all the indices:

Theorem 4.7 *The second-order IF of the functional associated with the Wald-type test statistics, with respect to all the observations, is given by*

$$\begin{aligned}&IF_2(\boldsymbol{t}, W_K, F_{\boldsymbol{\theta}^0})\\&= 2\ IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})\boldsymbol{m}^T(\boldsymbol{\theta}^0)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\boldsymbol{\theta}^0)IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}),\end{aligned}$$

where $IF(\boldsymbol{t}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0})$ is given in (5.9).

Proof. Straightforward following results on Section 3.3.3. ■

4.3 Simulation study

In this section, Monte Carlo simulations of size 2.500 are carried out to examine the behavior of the weighted minimum DPD estimators and Wald-type tests discussed in the preceding sections.

4.3.1 Weighted minimum DPD estimators

Based on the simulation experiment proposed by Balakrishnan and Ling [2014a], we consider the devices to have gamma lifetimes, under 4 different conditions with 2 stress factors at 2 levels, taken to be $\{(30, 40), (40, 40), (30, 50), (40, 50)\}$. Then, all devices under each condition are tested at 3 different inspection times, depending on the reliability considered. The model parameters were set as $(a_1, a_2, b_0, b_1, b_2) = (-0.06, -0.06, -0.36, 0.04, -0.01)$ while $a_0 = 6.5, 7$ or 7.5, corresponding to low, moderate and high reliability, respectively. In order to study the robustness of the weighted minimum DPD estimators, we consider a contaminated scheme, wherein the first "cell" is generated under $\tilde{a}_1 = -0.035$.

Bias of estimates of reliabilities at normal conditions and different times, as well as the RMSE of the parameter estimates, are computed with the same sample size for each condition $K = \{50, 100, 150\}$, and those are presented in Table 4.3.1, 4.3.2 and 4.3.3.

It can be seen that, while for the non-contaminated scheme, the MLE generally possesses the best behaviour, weighted minimum DPD estimators with medium β are a better option in the contamination scenario. This robustness is in accordance with the earlier finding for the case of one-shot device testing based on exponential lifetimes.

Table 4.3.1: Gamma distribution at multiple stress levels: Bias of the estimates of reliabilities for pure and contaminated data in the case of low reliability.

Low reliability		Pure data				Contaminated data			
	True value	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
k=50									
$R(10; (25, 30))$	0.8197	-0,0101	-0,0085	-0,0135	-0,3130	0,1216	0,1137	0,1141	-0,0572
$R(20; (25, 30))$	0.6168	-0,0037	-0,0058	-0,0064	-0,2287	0,1162	0,1027	0,0993	-0,0294
$R(30; (25, 30))$	0.4497	-0,009	-0,0134	-0,0114	-0,1688	0,0037	-0,0009	-0,0063	-0,0788
$R(40; (25, 30))$	0.3220	-0,0091	-0,0145	-0,0109	-0,1214	-0,0677	-0,0655	-0,0704	-0,0106
$R(50; (25, 30))$	0.2278	-0,0037	-0,0092	-0,0048	-0,0830	-0,0849	-0,0815	-0,0852	-0,1013
RMSE(θ)	-	0,9933	0,9737	1,0207	2,2000	1,8496	1,7226	1,7652	1,9626
k=100									
$R(10; (25, 30))$	0.8197	-0,0055	-0,003	-0,0056	-0,2749	0,1291	0,1193	0,1216	0,0100
$R(20; (25, 30))$	0.6168	-0,0027	-0,0035	-0,0031	-0,2016	0,1229	0,1084	0,106	0,0194
$R(30; (25, 30))$	0.4497	-0,0055	-0,0088	-0,0061	-0,1456	0,0099	0,0064	-0,0007	-0,0498
$R(40; (25, 30))$	0.3220	-0,0057	-0,0105	-0,0063	-0,1022	-0,0700	-0,0648	-0,0728	-0,0949
$R(50; (25, 30))$	0.2278	-0,0029	-0,0082	-0,0033	-0,0686	-0,0940	-0,0867	-0,0938	-0,1018
RMSE(θ)	-	0,706	0,6919	0,7174	1,4967	1,7763	1,6122	1,6848	1,6592
k=150									
$R(10; (25, 30))$	0.8197	-0,0055	-0,0021	-0,0061	-0,2563	0,1317	0,1214	0,1234	0,0697
$R(20; (25, 30))$	0.6168	-0,0028	-0,0024	-0,0034	-0,1887	0,1237	0,1092	0,1063	0,0632
$R(30; (25, 30))$	0.4497	-0,0039	-0,0064	-0,0042	-0,1357	0,0120	0,0095	0,0017	-0,0215
$R(40; (25, 30))$	0.3220	-0,0037	-0,0081	-0,0039	-0,0950	-0,0710	-0,0638	-0,0730	-0,0810
$R(50; (25, 30))$	0.2278	-0,0019	-0,0070	-0,0018	-0,0640	-0,0991	-0,0893	-0,0977	-0,0982
RMSE(θ)	-	0,5714	0,5754	0,5785	1,1034	1,7675	1,5892	1,6629	1,5933

Table 4.3.2: Gamma distribution at multiple stress levels: Bias of the estimates of reliabilities for pure and contaminated data in the case of moderate reliability.

Moderate reliability		Pure data				Contaminated data			
	True value	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
k=50									
$R(40; (25, 30))$	0.5406	-0,0322	-0,0331	-0,0345	-0,0356	-0,0965	-0,0859	-0,0764	-0,0716
$R(50; (25, 30))$	0.4449	-0,0317	-0,0318	-0,0335	-0,0342	-0,1242	-0,1108	-0,0989	-0,0917
$R(60; (25, 30))$	0.3638	-0,0255	-0,0248	-0,0266	-0,0269	-0,1282	-0,1145	-0,1024	-0,0942
$R(70; (25, 30))$	0.2960	-0,0165	-0,0151	-0,0169	-0,0169	-0,1195	-0,1067	-0,0955	-0,0872
$R(80; (25, 30))$	0.2399	-0,0066	-0,0048	-0,0064	-0,0063	-0,1053	-0,0936	-0,0837	-0,0758
RMSE(θ)	-	1,1827	1,1857	1,2093	1,2265	1,7034	1,5738	1,4878	1,4367
k=100									
$R(40; (25, 30))$	0.5406	-0,0169	-0,0177	-0,0174	-0,0184	-0,0774	-0,0666	-0,0552	-0,0512
$R(50; (25, 30))$	0.4449	-0,0189	-0,0193	-0,0191	-0,0208	-0,1177	-0,1028	-0,0875	-0,0811
$R(60; (25, 30))$	0.3638	-0,0173	-0,0174	-0,0171	-0,0191	-0,1330	-0,1175	-0,1016	-0,0940
$R(70; (25, 30))$	0.2960	-0,0132	-0,0130	-0,0127	-0,0147	-0,1319	-0,1177	-0,1028	-0,0951
$R(80; (25, 30))$	0.2399	-0,0080	-0,0074	-0,0072	-0,0091	-0,1220	-0,1096	-0,0965	-0,0892
RMSE(θ)	-	0,8056	0,8074	0,8144	0,8344	1,4918	1,3356	1,2141	1,1558
k=150									
$R(40; (25, 30))$	0.5406	-0,0128	-0,0128	-0,0131	-0,0132	-0,0713	-0,0592	-0,0483	-0,0432
$R(50; (25, 30))$	0.4449	-0,0143	-0,0138	-0,0147	-0,0149	-0,1160	-0,0988	-0,0836	-0,0754
$R(60; (25, 30))$	0.3638	-0,0133	-0,0124	-0,0138	-0,0139	-0,1356	-0,1175	-0,1012	-0,0917
$R(70; (25, 30))$	0.2960	-0,0105	-0,0092	-0,0109	-0,0109	-0,1374	-0,1209	-0,1053	-0,0959
$R(80; (25, 30))$	0.2399	-0,0067	-0,0052	-0,0069	-0,0068	-0,1289	-0,1148	-0,1010	-0,0922
RMSE(θ)	-	0,6769	0,6788	0,6903	0,7028	1,4402	1,2655	1,1330	1,0542

Table 4.3.3: Gamma distribution at multiple stress levels: Bias of the estimates of reliabilities for pure and contaminated data in the case of high reliability.

High reliability		Pure data				Contaminated data			
	True value	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
k=50									
$R(70; (25, 30))$	0.5157	-0,0587	-0,0551	-0,0580	-0,0623	-0,1591	-0,1374	-0,1259	-0,1101
$R(80; (25, 30))$	0.4581	-0,0493	-0,0458	-0,0496	-0,0554	-0,1530	-0,1326	-0,1214	-0,1067
$R(90; (25, 30))$	0.4059	-0,0385	-0,0349	-0,0395	-0,0465	-0,1431	-0,1240	-0,1133	-0,1000
$R(100; (25, 30))$	0.3590	-0,0270	-0,0233	-0,0284	-0,0364	-0,1311	-0,1131	-0,1030	-0,0912
$R(110; (25, 30))$	0.3169	-0,0154	-0,0116	-0,0170	-0,0258	-0,1181	-0,1012	-0,0915	-0,0811
RMSE(θ)	-	1,7033	1,7030	1,7033	1,6846	1,8756	1,7491	1,7106	1,5983
k=100									
$R(70; (25, 30))$	0.5157	-0,0451	-0,0463	-0,0492	-0,0535	-0,1652	-0,1405	-0,1243	-0,1107
$R(80; (25, 30))$	0.4581	-0,0421	-0,0441	-0,0475	-0,0525	-0,1677	-0,1438	-0,1278	-0,1157
$R(90; (25, 30))$	0.4059	-0,0369	-0,0396	-0,0433	-0,0489	-0,1643	-0,1416	-0,1263	-0,1158
$R(100; (25, 30))$	0.3590	-0,0304	-0,0335	-0,0374	-0,0434	-0,1569	-0,1356	-0,1213	-0,1123
$R(110; (25, 30))$	0.3169	-0,0231	-0,0265	-0,0305	-0,0367	-0,1471	-0,1273	-0,1141	-0,1063
RMSE(θ)	-	1,1432	1,1406	1,1653	1,1651	1,5383	1,3921	1,3102	1,2219
k=150									
$R(70; (25, 30))$	0.5157	-0,0376	-0,0393	-0,0419	-0,0485	-0,1639	-0,1379	-0,1194	-0,1080
$R(80; (25, 30))$	0.4581	-0,0369	-0,0391	-0,0420	-0,0497	-0,1705	-0,1458	-0,1273	-0,1164
$R(90; (25, 30))$	0.4059	-0,0341	-0,0367	-0,0398	-0,0483	-0,1702	-0,1474	-0,1295	-0,1195
$R(100; (25, 30))$	0.3590	-0,0298	-0,0326	-0,0359	-0,0448	-0,1651	-0,1444	-0,1275	-0,1184
$R(110; (25, 30))$	0.3169	-0,0246	-0,0275	-0,0307	-0,0400	-0,1569	-0,1382	-0,1225	-0,1144
RMSE(θ)	-	0,9471	0,9525	0,9749	0,9772	1,4014	1,2390	1,1308	1,0681

4.3.2 Wald-type tests

Let us now empirically evaluate the robustness of the Wald-type tests developed. The simulation is performed under the low-reliability model described before.

We first study the observed level (measured as the proportion of test statistics exceeding the corresponding chi-square critical value) of the test under the true null hypothesis $H_0 : a_1 = -0.06$ against the alternative $H_1 : a_1 \neq -0.06$. In the top of Figure 4.3.1, these levels are plotted for different values of the samples sizes, pure data (left) and contaminated data ($\tilde{a}_1 = -0.035$, right). Notice that in the case of pure data considered, all the observed levels are close to the nominal level of 0.05. In the case of contaminated data, the level of the classical Wald test (at $\beta = 0$) displays a lack of robustness, while the weighted minimum DPD estimators based Wald-type tests for moderate and large positive β possess levels closer to the nominal level.

To investigate the power of these tests (obtained in a similar manner), we change the true data generating parameters value to $\boldsymbol{\theta} = (6.5, -0.06, -0.035, -0.36, 0.04, -0.01)$, and $\tilde{a}_1 = -0.45$ in a contaminated scenario, nearer to the null hypothesis. The resulting empirical powers are plotted in the bottom of Figure 4.3.1. When there are no outliers in the data, the classical Wald test (at $\beta = 0$) is quite similar, not even the most powerful, to other tests. On the other hand, when there are outliers in the data, the Wald-type tests with larger $\beta > 0$ provides a significantly better power.

4.4 Real data example: application to a tumor toxicological data

Survival analysis usually faces problems associated with interval censoring. One extreme situation is the one in which the only available information on a survival variable is whether or not it exceeds a monitoring time. This form of censoring, known as current status data, can be seen as one-shot

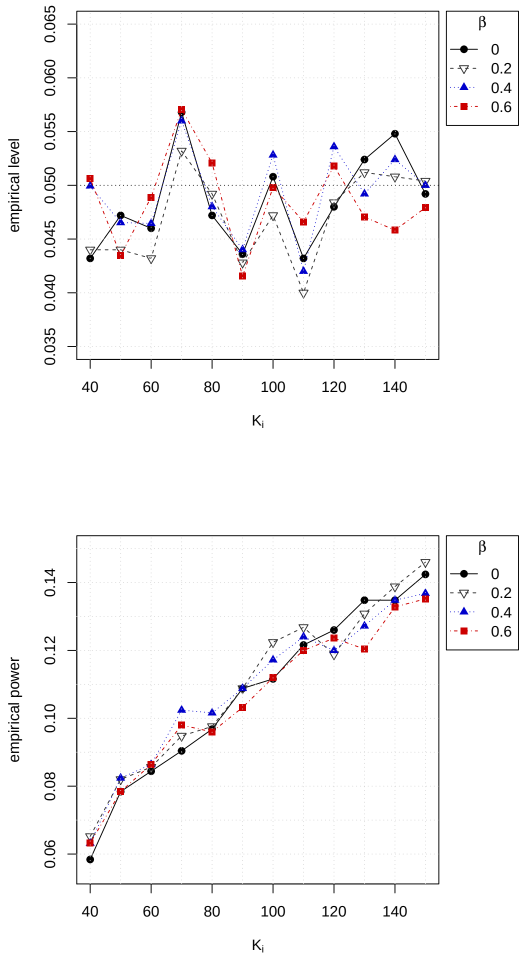

Figure 4.3.1: Gamma distribution at multiple stress levels: Levels and powers for pure (left) and contaminated data (right).

device testing data, and so we can apply the methods developed in the preceding sections to a real current status data from a tumor toxicological experiment.

The data considered, taken from the National Center for Toxicological Research, was originally reported by Kodell and Nelson [1980] and recently analyzed by Balakrishnan and Ling [2013, 2014a] using MLE under a one-shot device model. In Chapter 3, these data were analyzed using weighted minimum DPD estimators, but under the assumption of exponential lifetimes. However, the gamma distribution is a better lifetime model for these data (Balakrishnan and Ling [2014a]). These data consisted of 1816 mice, of which 553 had tumors, involving the strain of offspring (F1 or F2), gender (females or males), and concentration of benzidine dihydrochloride (60 ppm, 120 ppm, 200 ppm or 400 ppm) as the stress factors. For each testing condition, the numbers of mice tested and the numbers of mice having tumors were all recorded.

Let a_1, a_2 and a_3 denote the parameters corresponding to the covariates of strain of offspring, gender and square root of concentration of the chemical of benzidine dihydrochloride in the shape parameter of the gamma distribution, while b_1, b_2 and b_3 denote similarly for the scale parameter, respectively. The weighted minimum DPD estimators of model parameters as well as of the mean time to occurrence of tumors for each group, for different values of $\beta \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, are computed and these are presented in Tables

Table 4.4.1: Gamma distribution at multiple stress levels: weighted minimum DPD estimators of the model parameters.

β	a_0	a_1	a_2	a_3	b_0	b_1	b_2	b_3	MAB($p(\widehat{\boldsymbol{\theta}})$)	RMSE($p(\widehat{\boldsymbol{\theta}})$)
0	2.4066	-0.1875	-1.0099	0.0359	0.8730	0.2419	1.5545	-0.0901	0.2758	0.3950
0.1	2.8958	-0.1743	-1.2198	0.0136	0.3678	0.2196	1.7680	-0.0670	0.2701	0.3931
0.2	2.8644	-0.1477	-1.3185	0.0219	0.4050	0.1885	1.8742	-0.0758	0.2667	0.3926
0.3	2.7834	-0.1375	-1.3920	0.0332	0.4935	0.1756	1.9535	-0.0877	0.2638	0.3922
0.4	2.6980	-0.1275	-1.4569	0.0443	0.5847	0.1635	2.0217	-0.0994	0.2616	0.3919
0.5	2.6343	-0.1205	-1.5071	0.0529	0.6517	0.1549	2.0732	-0.1082	0.2601	0.3917
0.6	2.5965	-0.1189	-1.5404	0.0585	0.6912	0.1525	2.1067	-0.1139	0.2593	0.3917
0.7	2.5758	-0.1219	-1.5603	0.0619	0.7126	0.1551	2.1263	-0.1173	0.2590	0.3918
0.8	2.5636	-0.1267	-1.5716	0.0640	0.7253	0.1598	2.1370	-0.1194	0.2588	0.3919
0.9	2.5597	-0.1328	-1.5771	0.0651	0.7293	0.1659	2.1420	-0.1205	0.2588	0.3920
1	2.5570	-0.1384	-1.5787	0.0658	0.7321	0.1716	2.1431	-0.1212	0.2588	0.3921

Table 4.4.2: Gamma distribution at multiple stress levels: weighted minimum DPD estimators of the mean time to the occurrence of tumors (in months), $\widehat{E}[T]$.

			$\widehat{E}[T]$					
strain	gender	conc	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0.8$	$\beta = 1$
0	0	60	17.461	17.316	17.395	17.433	17.449	17.455
0	1	60	30.103	30.188	30.598	30.712	30.713	30.694
1	0	60	18.438	18.039	18.031	18.029	18.036	18.044
1	1	60	31.787	31.447	31.719	31.762	31.747	31.728
0	0	120	14.676	14.565	14.577	14.594	14.605	14.610
0	1	120	25.301	25.392	25.643	25.710	25.707	25.691
1	0	120	15.496	15.173	15.111	15.092	15.096	15.103
1	1	120	26.715	26.451	26.582	26.588	26.572	26.557
0	0	200	12.348	12.265	12.231	12.231	12.238	12.243
0	1	200	21.288	21.381	21.514	21.547	21.541	21.529
1	0	200	13.039	12.776	12.678	12.649	12.650	12.656
1	1	200	22.478	22.273	22.302	22.283	22.266	22.254
0	0	400	8.991	8.942	8.858	8.840	8.843	8.848
0	1	400	15.500	15.589	15.583	15.574	15.566	15.558
1	0	400	9.493	9.315	9.183	9.142	9.141	9.146
1	1	400	16.366	16.240	16.153	16.106	16.090	16.082

4.4.1 and 4.4.2, respectively, where strain= 0 for F1 strain of offspring, and gender= 0 for females.

There is a significant difference between genders, with males having a higher expected lifetime. Also, tumors are induced by an increase in the dosage of benzidine dihydrochloride. Empirical mean absolute error (MAB) and RMSE, measured by comparing predicted probabilities to the observed ones, are also presented in Table 4.4.1. In both cases, MLE presents the maximum error.

Chapter 5

Robust inference for one-shot device testing under Weibull distribution

5.1 Introduction

Let us consider again the one-shot device testing problem with multiple stress factors, schematized in Table 3.1.1. So far, we have considered that the lifetimes can follow an exponential distribution (Chapter 3) or gamma distribution (Chapter 4). However, in practice, the Weibull distribution is widely used as a lifetime model in engineering and physical sciences. In fact, the Weibull model is also used extensively in biomedical studies as a proportional hazards model for evaluating the effects of covariates on lifetimes, meaning that the hazard rates of any two products stay in constant ratio over time. See Meeter and Meeker [1994], Meeker et al. [1998], and references therein. However, in some situations, the assumption of constant shape parameters may not be valid; see, for example, Kodell and Nelson [1980], Nogueira et al. [2009] and Vázquez et al. [2010]. In such situations, Balakrishnan and Ling [2013] suggested using a log-link of the stress levels to model the unequal shape parameters. Based on this idea, we develop, in this chapter, robust inference for one-shot device testing under the Weibull distribution with scale and shape parameters varying over stress.

The chapter is organized as follows: in Section 5.1.1, the Weibull distribution is presented. Inference based on minimum DPD estimators under the Weibull assumption is developed in Section 5.2. A complete simulation study and three numerical examples are presented in Section 5.3 and Section 5.4, respectively.

The results of this Chapter have been published in the form of a paper (Balakrishnan et al. [2020b]).

5.1.1 The Weibull distribution

Let us denote by $\boldsymbol{\theta} = (a_0, \ldots, a_J, b_0, \ldots, b_J)^T$ the model parameter vector. We shall then assume that the lifetimes of the units, under the testing condition i, follow Weibull distribution with corresponding probability density function and cumulative distribution function as

$$f_T(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\eta_i t^{\eta_i - 1}}{\alpha_i^{\eta_i}} e^{-\left(\frac{t}{\alpha_i}\right)^{\eta_i}},\ t > 0,$$

and

$$F_T(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - e^{-\left(\frac{t}{\alpha_i}\right)^{\eta_i}},\ t > 0,$$

where $\alpha_i > 0$ and $\eta_i > 0$ are, respectively, the scale and shape parameters at condition i, which we assume are related to the stress factors in log–linear forms as

$$\alpha_i = \exp\left\{\sum_{j=0}^{J} a_j x_{ij}\right\} \quad \text{and} \quad \eta_i = \exp\left\{\sum_{j=0}^{J} b_j x_{ij}\right\},$$

with $x_{i0} = 1$ for all i. Let us denote by $R_T(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - F(t; \boldsymbol{x}_i, \boldsymbol{\theta})$ the reliability function, the probability that the unit lasts lifetime t. The hazard function, given by the ratio of the density function and the reliability function, is

$$h_T(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\eta_i t^{\eta_i - 1}}{\alpha_i^{\eta_i}}, \ t > 0.$$

When $\eta_i = 1$, the hazard rate is constant and the Weibull distribution in this case is simply exponential distribution. When $\eta_i > 1$, the unit suffers an increasing rate of failure as it ages, while the opposite is the case when $\eta_i < 1$. This last case is less common in practice, unless we only consider the early part of lifetimes of devices. Weibull density function, reliability and hazard functions, for different values of shape and scale parameters, are shown in Figure 5.1.1.

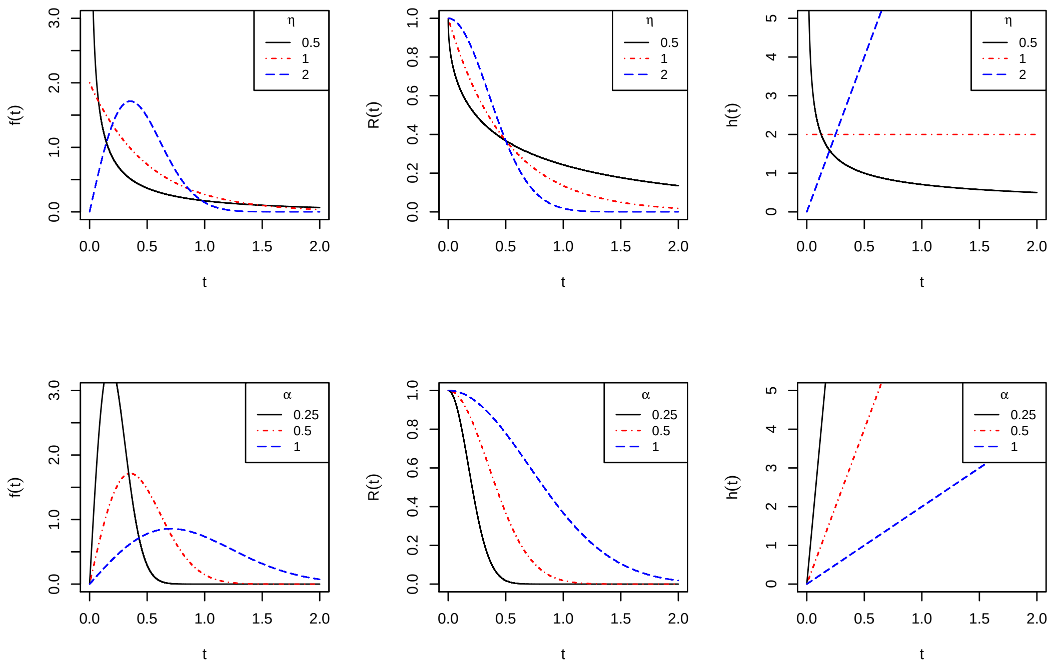

Figure 5.1.1: Weibull distributions for different values of shape and scale parameters.

Notice that, as suggested in the literature (see, for example, Meeter and Meeker [1994] and Ng et al. [2002]), it is often more convenient to work with the extreme value distribution for the log-lifetimes, as it belongs to the location-scale family rather than the Weibull distribution belonging to the scale-shape family. For this same reason, we will also consider here the extreme value distribution with the corresponding probability density, distribution and reliability functions as

$$f_W(\omega; \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{1}{\sigma_i} e^{\frac{\omega - \mu_i}{\sigma_i}} e^{-e^{\frac{\omega - \mu_i}{\sigma_i}}} = \frac{1}{\sigma_i} \xi_i e^{-\xi_i}, \ -\infty < \omega < \infty, \tag{5.1}$$

$$F_W(\omega; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - e^{-e^{\frac{\omega - \mu_i}{\sigma_i}}} = 1 - e^{-\xi_i}, \ -\infty < \omega < \infty, \tag{5.2}$$

$$R_W(\omega; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - F_W(\omega; \boldsymbol{x}_i, \boldsymbol{\theta}) = e^{-e^{\frac{\omega - \mu_i}{\sigma_i}}} = e^{-\xi_i}, \ -\infty < \omega < \infty, \tag{5.3}$$

where $\omega = \log(t)$, $\xi_i = e^{\frac{\omega - \mu_i}{\sigma_i}}$, the location parameter $\mu_i = log(\alpha_i) = \sum_{j=0}^{J} a_j x_{ij}$, and the scale parameter $\sigma_i = \eta_i^{-1} = \exp\{-\sum_{j=0}^{J} b_j x_{ij}\}$.

5.2 Inference under the Weibull distribution

Let us consider the weighted minimum DPD estimator for $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_\beta$, given in Definition 3.1, where $\pi_{i1}(\boldsymbol{\theta})$ and $\pi_{i2}(\boldsymbol{\theta})$ are given in (5.2) and (5.3), respectively.

Let us first develop the estimating equations of the minimum DPD estimators under the Weibull distribution, as well as its asymptotic distribution.

Theorem 5.1 *For $\beta \geq 0$, the estimating equations are given by*

$$\sum_{i=1}^{I} l_i \left(K_i F_W\left(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) - n_i\right)\left(F_W^{\beta-1}\left(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) + R_W^{\beta-1}\left(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right)\boldsymbol{x}_i = \mathbf{0}_{J+1},$$
$$\sum_{i=1}^{I} s_i \left(K_i F_W\left(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) - n_i\right)\left(F_W^{\beta-1}\left(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) + R_W^{\beta-1}\left(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right)\boldsymbol{x}_i = \mathbf{0}_{J+1},$$

where $F_W(lIT_i; S_i, \boldsymbol{\theta})$, $F_W(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta})$ and $R_W(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta})$ are as given in (5.1), (5.2) and (5.3), respectively, and

$$l_i = -\{\xi_i e^{-\xi_i}\}/\sigma_i, \quad s_i = \xi_i e^{-\xi_i} log(\xi_i), \quad i = 1, \ldots, I.$$

Proof. The proof is straightforward following (4.4) and taking into account

$$\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta}) = l_i = -\{\xi_i e^{-\xi_i}\}/\sigma_i, \tag{5.4}$$
$$\frac{\partial}{\partial \boldsymbol{b}}\pi_{i1}(\boldsymbol{\theta}) = s_i = \xi_i e^{-\xi_i} log(\xi_i). \tag{5.5}$$

■

Theorem 5.2 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter. The asymptotic distribution of the minimum DPD estimator, $\widehat{\boldsymbol{\theta}}_\beta$, is given by*

$$\sqrt{K}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0) \xrightarrow[K\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_{2(J+1)}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are given by

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Psi}_i\left(F_W^{\beta-1}(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + R_W^{\beta-1}(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta})\right), \tag{5.6}$$
$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\Psi}_i F_W(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) R_W(lIT_i; S_I, \boldsymbol{\theta})\left(F_W^{\beta-1}(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta}) + R_W^{\beta-1}(lIT_i; \boldsymbol{x}_i, \boldsymbol{\theta})\right)^2, \tag{5.7}$$

with

$$\boldsymbol{\Psi}_i = \begin{pmatrix} l_i^2\boldsymbol{x}_i\boldsymbol{x}_i^T & l_i s_i \boldsymbol{x}_i\boldsymbol{x}_i^T \\ l_i s_i \boldsymbol{x}_i\boldsymbol{x}_i^T & s_i^2\boldsymbol{x}_i\boldsymbol{x}_i^T \end{pmatrix}.$$

Proof. Straightforward following proof of Theorem 4.2 and equations (5.4) and (5.5). ■

Now, we present the IF of the proposed estimators:

Theorem 5.3 *Let us consider the one-shot device testing under the Weibull distribution with multiple stress factors. The IF with respect to the $k-th$ observation of the i_0-th group is given by*

$$\begin{aligned} IF(t_{i_0,k}, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}^0}) =& \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)(l_{i_0}\boldsymbol{x}_{i_0}, s_{i_0}\boldsymbol{x}_{i_0})^T \\ &\times \left(F^{\beta-1}(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0) + R^{\beta-1}(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0)\right)\left(F(IT_{i_0}; \boldsymbol{x}_{i_0}, \boldsymbol{\theta}^0) - \Delta_{t_{i_0}}^{(1)}\right), \end{aligned} \tag{5.8}$$

where $\Delta_{t_{i_0},k}^{(1)}$ is the degenerating function at point (t_{i_0}, k).

Proof. Straightforward following results in Section 2.5. ■

Theorem 5.4 *Let us consider the one-shot device testing under the Weibull distribution with multiple stress factors. The IF with respect to all the observations is given by*

$$\begin{aligned} IF(\boldsymbol{t},\boldsymbol{U}_{\beta},F_{\boldsymbol{\theta}^0}) =& \boldsymbol{J}_{\beta}^{-1}(\boldsymbol{\theta}^0)\sum_{i=1}^{I}\frac{K_i}{K}(l_i\boldsymbol{x}_i, s_i\boldsymbol{x}_i)^T \\ & \times \left(F^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}^0)+R^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}^0)\right)\left(F(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}^0)-\Delta_{t_i}^{(1)}\right), \end{aligned} \tag{5.9}$$

where $\Delta_{t_i}^{(1)}=\sum_{k=1}^{K_i}\Delta_{t_{i,k}}^{(1)}$.

Proof. Straightforward following results in Section 2.5. ■

5.2.1 Wald-type tests

From Theorem 5.2, and following the idea on previous chapters, we can develop Wald-type tests for testing composite null hypotheses.

Let us consider the function $\boldsymbol{m}:\mathbb{R}^{2(J+1)}\longrightarrow\mathbb{R}^r$, where $r\leq 2(J+1)$. Then, $\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{0}_r$ represents a composite null hypothesis. We assume that the $2(J+1)\times r$ matrix

$$\boldsymbol{M}(\boldsymbol{\theta})=\frac{\partial \boldsymbol{m}^T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

exists and is continuous in $\boldsymbol{\theta}$ and with rank $\boldsymbol{M}(\boldsymbol{\theta})=r$. For testing

$$H_0:\boldsymbol{\theta}\in\boldsymbol{\Theta}_0 \text{ against } H_1:\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0, \tag{5.10}$$

where $\boldsymbol{\Theta}_0=\left\{\boldsymbol{\theta}\in\mathbb{R}^{2(J+1)}:\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{0}_r\right\}$, we can consider the following Wald-type test statistics

$$W_K(\widehat{\boldsymbol{\theta}}_\beta)=K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta), \tag{5.11}$$

where $\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)=\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)$ and $\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are as in (5.6) and (5.7), respectively.

In the following theorem, we present the asymptotic distribution of $W_K(\widehat{\boldsymbol{\theta}}_\beta)$.

Theorem 5.5 *The asymptotic null distribution of the proposed Wald-type test statistics, given in Equation (5.11), is a chi-squared (χ^2) distribution with r degrees of freedom. This is,*

$$W_K(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\chi^2_r.$$

Proof. Let $\boldsymbol{\theta}^0\in\Theta_0$ be the true value of parameter $\boldsymbol{\theta}$. $\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta-\boldsymbol{\theta}^0\right)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\mathcal{N}\left(\boldsymbol{0}_{2(J+1)},\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\right)$. Therefore, under H_0, we have

$$\sqrt{K}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\mathcal{N}\left(\boldsymbol{0}_r,\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)$$

and taking into account that $rank(\boldsymbol{M}(\boldsymbol{\theta}^0))=r$, we obtain

$$K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\chi^2_r.$$

But, $\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}$ is a consistent estimator of $\left(\boldsymbol{M}^T(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}^0)\boldsymbol{M}(\boldsymbol{\theta}^0)\right)^{-1}$ and, therefore, $W_K(\widehat{\boldsymbol{\theta}}_\beta)\underset{K\to\infty}{\xrightarrow{\mathcal{L}}}\chi^2_r$. ■

Based on Theorem 5.5, we will reject the null hypothesis in (5.10) if $W_K(\widehat{\boldsymbol{\theta}}_\beta)>\chi^2_{r,\alpha}$, where $\chi^2_{r,\alpha}$ is the upper percentage point of order α of χ^2_r distribution.

Results concerning the power function of the proposed Wald-type tests could be obtained in a similar manner to previous chapters.

5.3 Simulation Study

In this section, a Monte Carlo simulation study that examines the accuracy of the proposed minimum weighted DPD estimators and Wald-type tests is presented. Section 5.3.1 focuses on the efficiency, measured in terms of MSE and mean absolute error (MAE), of the estimators of model parameters and reliabilities, while Section 5.3.2 examines the behavior of the Wald-type tests developed in preceding sections. Every condition of simulation were tested until $R = 2,500$ regular observations were obtained.

5.3.1 Weighted minimum DPD estimators

The lifetimes of devices are simulated from the Weibull distribution, for different levels of reliability and different sample sizes, under 3 different stress conditions with 1 stress factor at 3 levels, taken to be $\{x_1, x_2, x_3\} = \{30, 40, 50\}$. Then, all devices under each stress condition are tested at 3 different inspection times $IT = \{IT_1, IT_2, IT_3\}$, depending on the level of reliability. Our data will then be collected under 9 testing conditions $S_1 = \{x_1, IT_1\}, \dots, S_9 = \{x_3, IT_3\}$.

A. Balanced data

Firstly, a balanced data with equal sample size for each group was considered. K_i was taken to range from small to large sample sizes, and the model parameters were set to $\boldsymbol{\theta}^T = (a_0, a_1, b_0, b_1)$ $= (a_0, -0.05, -0.6, 0.03)$, while a_0 was chosen to be 4.9, 5.3, and 5.7 corresponding to devices with low, moderate, and high reliability, respectively. To prevent many zero-observations in test groups, the inspection times were set as $IT = \{5, 10, 15\}$ for the case of low reliability, $IT = \{8, 16, 24\}$ for the case of moderate reliability, and $IT = \{12, 24, 36\}$ for the case of high reliability. To evaluate the robustness of the weighted minimum DPD estimators, we examine the behavior of this model in the presence of an outlying cell for the first testing condition $S_1 = \{x_1, IT_1\}$ in our table. This cell is generated under the parameters $\tilde{\boldsymbol{\theta}}^T = (a_0, \tilde{a}_1, b_0, b_1) = (4.9, -0.025, -0.6, 0.03)$. The setting used is now summarized in Table 5.3.1. While ALT data are based on extreme observations of the stress factor, and therefore on low values of the inspection times, we are interested in testing the accuracy of our estimators under normal conditions. MSEs of estimated reliabilities under the pure and the contaminated settings are computed and are presented in Table 5.3.3, for different values of the sample size $K_i \in \{50, 100\}$. As expected, MLE presents the best behavior in the case of pure data, while a gradual decrease in efficiency occurs with greater values of β. It is almost the opposite in the case of the contaminated scheme. This behaviour is corroborated when computing the MAEs and MSEs of the model parameter vector $\boldsymbol{\theta}$, as can be seen in Figures 5.3.1 and 5.3.2, respectively. Here, just the weighted minimum DPD estimators with tuning parameters $\beta \in \{0, 0.4, 0.8\}$ are represented in order to demonstrate the general robustness feature of the proposed estimators.

B. Unbalanced data

Now we consider an unbalanced data, which does not have equal sample size for all the groups. This data consists of a total of $K = 300$ observations, and is presented in Table 5.3.2. Here the vector of true parameters is $\boldsymbol{\theta}^T = (5.3, -0.025, -0.6, 0.03)$ (moderate reliability). To examine the robustness in this ALT plan, we increase each one of the parameters of the outlying first cell, denoted by $\tilde{a}_0$, $\tilde{a}_1$, $\tilde{b}_0$ and $\tilde{b}_1$. MSEs of the vector of parameters are plotted in Figure 5.3.3. In all the cases, we can see how the MLEs and the weighted minimum DPD estimators with small values of tuning parameter β present the smallest MSEs for weak outliers, i.e., when $\tilde{a}_0$ is near a_0 (and respectively with the other parameters). On the other hand, large values of tuning parameter β make the weighted minimum DPD estimators to yield the smallest MSEs, for medium and strong outliers.

Table 5.3.1: Weibull distribution at multiple stress levels: parameter values used in the simulation study.

Parameters	Symbols	Values
	Low reliability	
Inspection times	$IT = \{IT_1, IT_2, IT_3\}$	$\{5, 10, 15\}$
Model Par.	$\boldsymbol{\theta}^T$=$(a_0$,a_1,b_0,$b_1)$	$(4.9, -0.05, -0.6, 0.03)$
Outlying Par.	$\tilde{\boldsymbol{\theta}}^T$=$(a_0$,$\tilde{a}_1$,$b_0$,$b_1)$	$(4.9, -0.025, -0.6, 0.03)$
	Moderate reliability	
Inspection times	$IT = \{IT_1, IT_2, IT_3\}$	$\{8, 16, 24\}$
Model Par.	$\boldsymbol{\theta}^T$=$(a_0$,a_1,b_0,$b_1)$	$(5.3, -0.05, -0.6, 0.03)$
Outlying Par.	$\tilde{\boldsymbol{\theta}}^T$=$(a_0$,$\tilde{a}_1$,$b_0$,$b_1)$	$(5.3, -0.025, -0.6, 0.03)$
	High reliability	
Inspection times	$IT = \{IT_1, IT_2, IT_3\}$	$\{12, 24, 36\}$
Model Par.	$\boldsymbol{\theta}^T$=$(a_0$,a_1,b_0,$b_1)$	$(5.7, -0.05, -0.6, 0.03)$
Outlying Par.	$\tilde{\boldsymbol{\theta}}^T$=$(a_0$,$\tilde{a}_1$,$b_0$,$b_1)$	$(5.7, -0.025, -0.6, 0.03)$

Table 5.3.2: Weibull distribution at multiple stress levels: ALT plan, unbalanced data.

i	x_i	IT_i	K_i
1	30	8	60
2	40	8	40
3	50	8	20
4	30	16	60
5	40	16	20
6	50	16	20
7	30	24	40
8	40	24	20
9	50	24	20

It seems clear that the weighted minimum DPD estimators can be a robust alternative to MLE in terms of efficiency, overall when working with potential outlying data. It is important now to confirm this robustness when working with the Wald-type tests proposed in preceding sections.

5.3.2 Wald-type tests

To compute the accuracy in terms of contrast, we now consider the testing problem

$$H_0 : a_1 = -0.05 \quad \text{vs.} \quad H_1 : a_1 \neq -0.05. \tag{5.12}$$

The level of significance of a test is defined as the probability of rejecting the null hypothesis by the test when it is really true, while the power of a test is the probability of rejecting the null hypothesis when a specific hypothesis is true. For computing the empirical test level, we measure the proportion of test statistics exceeding the corresponding chi-square critical value. For a nominal size $\alpha = 0.05$, with the model under the null hypothesis given in (5.12), the estimated significance test levels for different Wald-type test statistics are given by

$$\widehat{\alpha}_K^{(\beta)} = \widehat{Pr}(W_K(\boldsymbol{\theta}_\beta) > \chi^2_{1,0.05}|H_0) = \frac{1}{R}\sum_{i=1}^{R} I(W_{K,i}(\widehat{\boldsymbol{\theta}}_\beta) > \chi^2_{1,0.05}|H_0),$$

with I(S) being the indicator function (with a value of one if S is true and zero otherwise). The simulated test powers will be obtained under H_1 in (5.12) in a similar way.

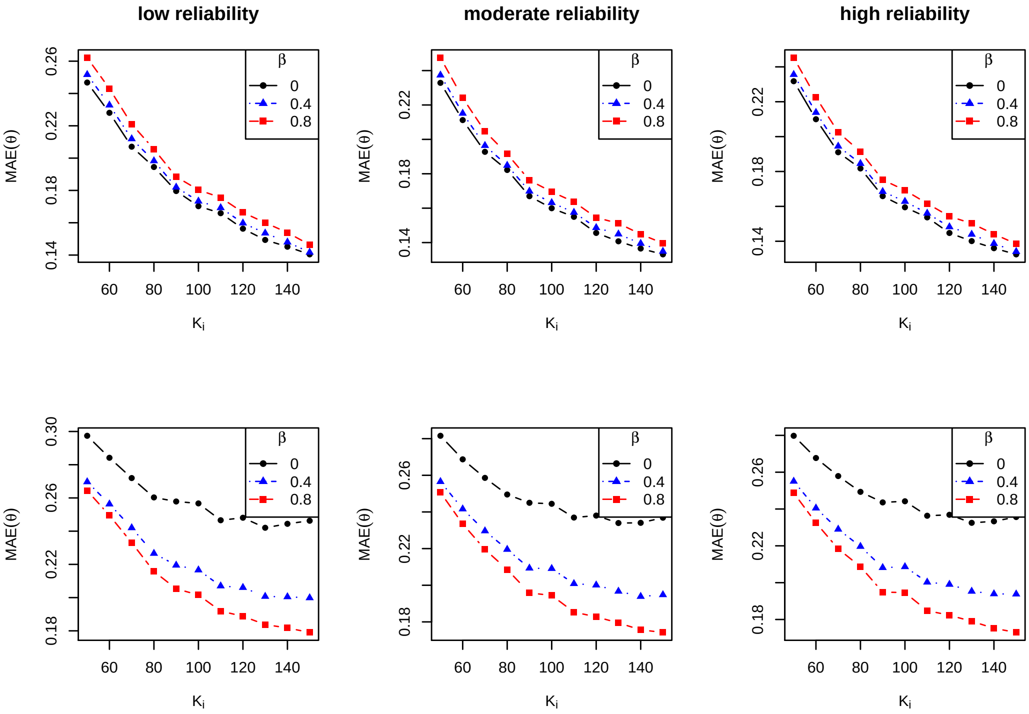

Figure 5.3.1: Weibull distribution at multiple stress levels: MAE of the estimates of parameters for different reliabilities under pure (top) and contaminated data (bottom)

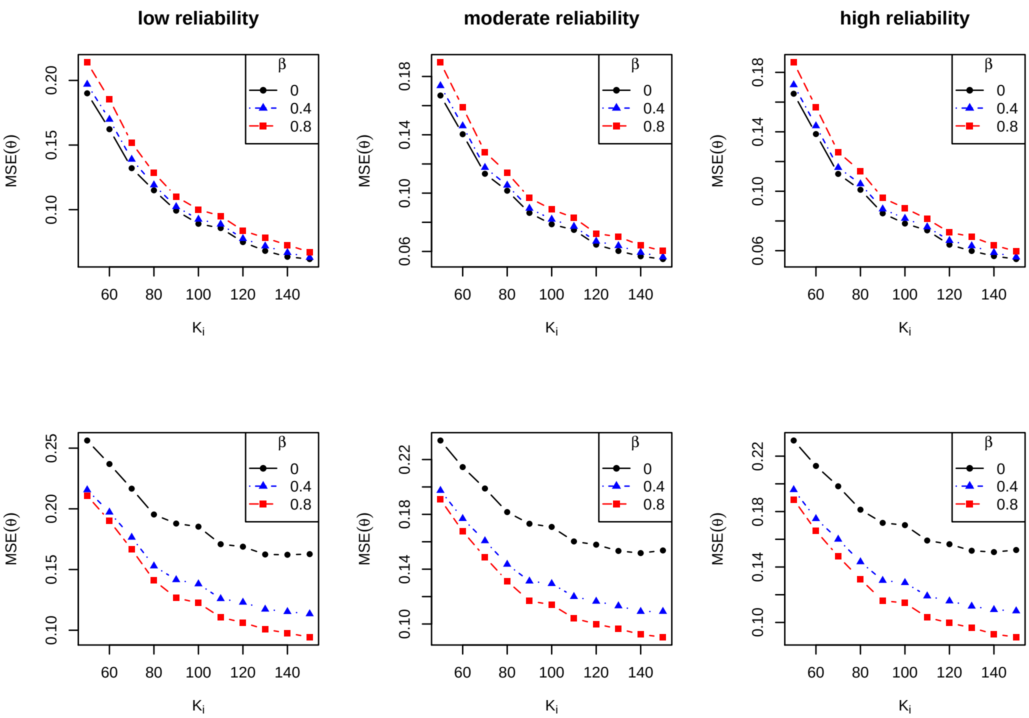

Figure 5.3.2: Weibull distribution at multiple stress levels: MSE of the estimates of parameters for different reliabilities under pure (top) and contaminated data (bottom)

Table 5.3.3: Weibull distribution at multiple stress levels: MSEs of estimates of reliabilities for different sample sizes and low reliability

	Pure data			Contaminated data		
$K = 50$	$R(10, 25)$	$R(20, 25)$	$R(30, 25)$	$R(10, 25)$	$R(20, 25)$	$R(30, 25)$
$\beta = 0$	0.0223	0.0264	0.0458	0.0290	0.0314	0.0670
0.1	0.0222	0.0264	0.0457	0.0281	0.0310	0.0647
0.2	0.0223	0.0264	0.0458	0.0272	0.0307	0.0629
0.3	0.0224	0.0264	0.0460	0.0266	0.0304	0.0615
0.4	0.0226	0.0265	0.0465	0.0261	0.0302	0.0605
0.5	0.0229	0.0266	0.0470	0.0257	0.0301	0.0599
0.6	0.0232	0.0268	0.0476	0.0255	0.0300	0.0594
0.7	0.0236	0.0270	0.0482	0.0253	0.0299	0.0591
0.8	0.0239	0.0272	0.0488	0.0253	0.0299	0.0589
0.9	0.0243	0.0273	0.0494	0.0252	0.0299	0.0588
1	0.0245	0.0275	0.0499	0.0253	0.0299	0.0588
$K = 100$						
$\beta = 0$	0.0155	0.0185	0.0314	0.0198	0.1064	0.1490
0.1	0.0155	0.0185	0.0313	0.0201	0.1057	0.1479
0.2	0.0156	0.0185	0.0314	0.0205	0.1048	0.1466
0.3	0.0157	0.0185	0.0315	0.0208	0.1039	0.1450
0.4	0.0158	0.0186	0.0317	0.0212	0.1030	0.1436
0.5	0.0159	0.0186	0.0319	0.0215	0.1020	0.1419
0.6	0.0161	0.0187	0.0321	0.0219	0.1011	0.1404
0.7	0.0163	0.0187	0.0324	0.0222	0.1002	0.1389
0.8	0.0165	0.0188	0.0327	0.0226	0.0994	0.1375
0.9	0.0167	0.0188	0.0329	0.0229	0.0986	0.1363
1	0.0169	0.0189	0.0331	0.0231	0.0980	0.1352

A. Balanced data

We compute empirical Wald-type test levels under the same parameters of low reliability model in the previous section for testing (5.12). Test powers are computed under the true parameter vector $\boldsymbol{\theta}^T = (4.9, -0.039, -0.6, 0.03)$. In the contaminated scheme, first cell will be generated from $\tilde{\boldsymbol{\theta}}^T = (4.9, -0.048, -0.6, 0.03)$, with the contrasted term nearer to the null hypothesis. This setting is summarized in Table 5.3.4. Considering samples sizes ranging from $K_i = 30$ to $K_i = 150$, results are shown in Figure 5.3.4.

For the pure setting, levels of the corresponding Wald-type tests based on different values of β seem to have the same behaviour. As the sample size increases, test levels get closer (with some exception, probably associated to empirical statistical error) to nominal level. In the case of contaminated data, the estimated test level, of classical Wald test ($\beta = 0$) is far away from the nominal level, while medium-high values of β have much more stable robustness properties. With respect to the power, and focusing on the Wald-type test based on $\beta = 0.8$, the same behaviour is observed, with an optimum classical Wald test for pure data, albeit its lack of robustness when contamination is present. Meanwhile, Wald-type test based on weighted minimum DPD estimator with $\beta = 0.8$ is seen to perform the best both in the pure and the contaminated sample cases.

B. Unbalanced data

For the unbalanced ALT plan presented in previous subsection, we compute empirical Wald-type test levels for the testing problem in (5.12). For illustrating robustness, we consider again an

Table 5.3.4: Weibull distribution at multiple stress levels: parameter values used in the simulation study

Parameters	Symbols	Values
	Levels	
Model True Parameters	$\boldsymbol{\theta}^T$=$(a_0$,a_1,b_0,$b_1)$	$(4.9, -0.05, -0.6, 0.03)$
Outlying Parameters	$\tilde{\boldsymbol{\theta}}^T$=$(a_0$,$\tilde{a}_1$,$b_0$,$b_1)$	$(4.9, -0.025, -0.6, 0.03)$
	Powers	
Model True Parameters	$\boldsymbol{\theta}^T$=$(a_0$,a_1,b_0,$b_1)$	$(4.9, -0.039, -0.6, 0.03)$
Outlying Parameters	$\tilde{\boldsymbol{\theta}}^T$=$(a_0$,$\tilde{a}_1$,$b_0$,$b_1)$	$(4.9, -0.048, -0.6, 0.03)$

increment of each one of the parameters of the outlying cell. Results, presented in Figure 5.3.5, illustrate again the lack of robustness of MLE for medium and strong outliers.

5.4 Real Data Examples

5.4.1 Glass Capacitors

Zelen [1959] presented data from a life test of glass capacitors at higher than usual levels of temperature (in °C), $T = \{170, 180\}$, and voltage $V = \{350, 300, 250, 200\}$. At each of the eight combinations of temperature and voltage, eight items were tested. We adapt these data to the one-shot device model taking the inspection times to be $IT = \{258, 315, 455, 1065\}$, respectively. Logically, higher inspection times are needed when applying less extreme voltages. These data and its relation with the Weibull distribution has been widely studied in the literature; see, for example, Meeker et al. [1998] and Rigdon et al. [2012]. As suggested in these papers, we have used the predictors as $\log(V)$ and $1/T_K$, where T_K is the temperature in degrees Kelvin.

Weighted minimum DPD estimators are computed for different values of the tuning parameter, β, and predicted probabilities are compared to the observed ones (top of Figure 5.4.3). The MAEs and RMSEs are presented in this figure as well. It is easily seen that the MLE seems to be either the worst, or one of the worst estimators in this case.

5.4.2 Solder Joints

Lau et al. [1988] described an experiment in which the reliability of 90 solder joints was studied under the effect of three types of printed circuit boards (PCBs) at three different temperatures. The lifetime was measured as the number of cycles until the solder joint failed, while the failure of a solder joint is defined as a 10% increase in measured resistance.

A simplified data set is derived from the original one and presented in Table 5.4.1. In it, the two stress factors considered are the temperature, $Temp$, and a dichotomous variable, PCB, indicating if the PCB type is "Copper-nickel-tin" ($PCB = 1$) or not.

Table 5.4.1: Solder Joints example

i	PCB_i	$Temp_i$	IT_i	n_i	K_i
1	1	20	300	4	10
2	1	60	300	4	10
3	1	100	100	6	10
4	0	20	1300	10	20
5	0	60	800	3	20
6	0	100	200	4	20

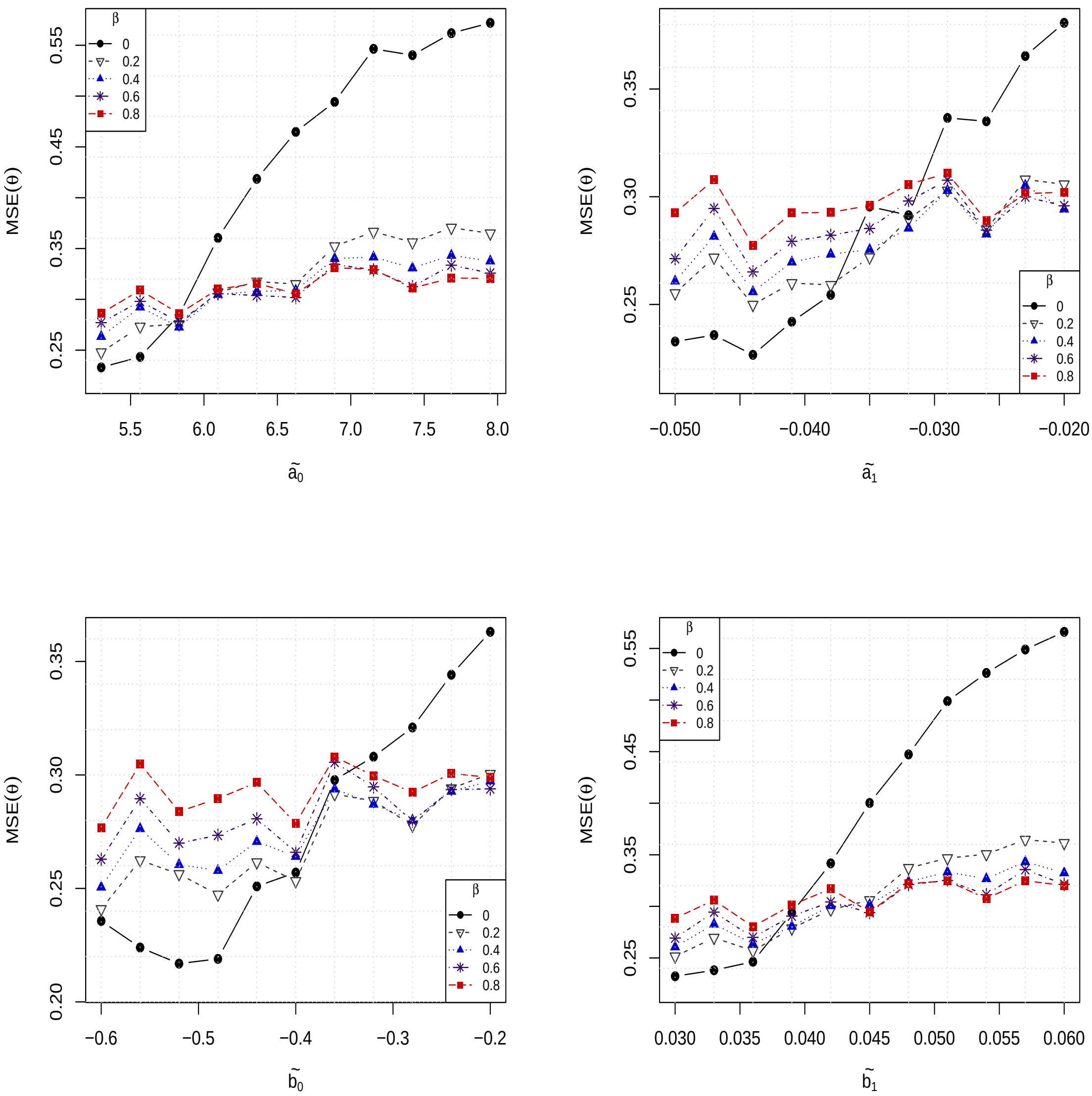

Figure 5.3.3: Weibull distribution at multiple stress levels: MSEs of the estimates of parameters. Unbalanced data

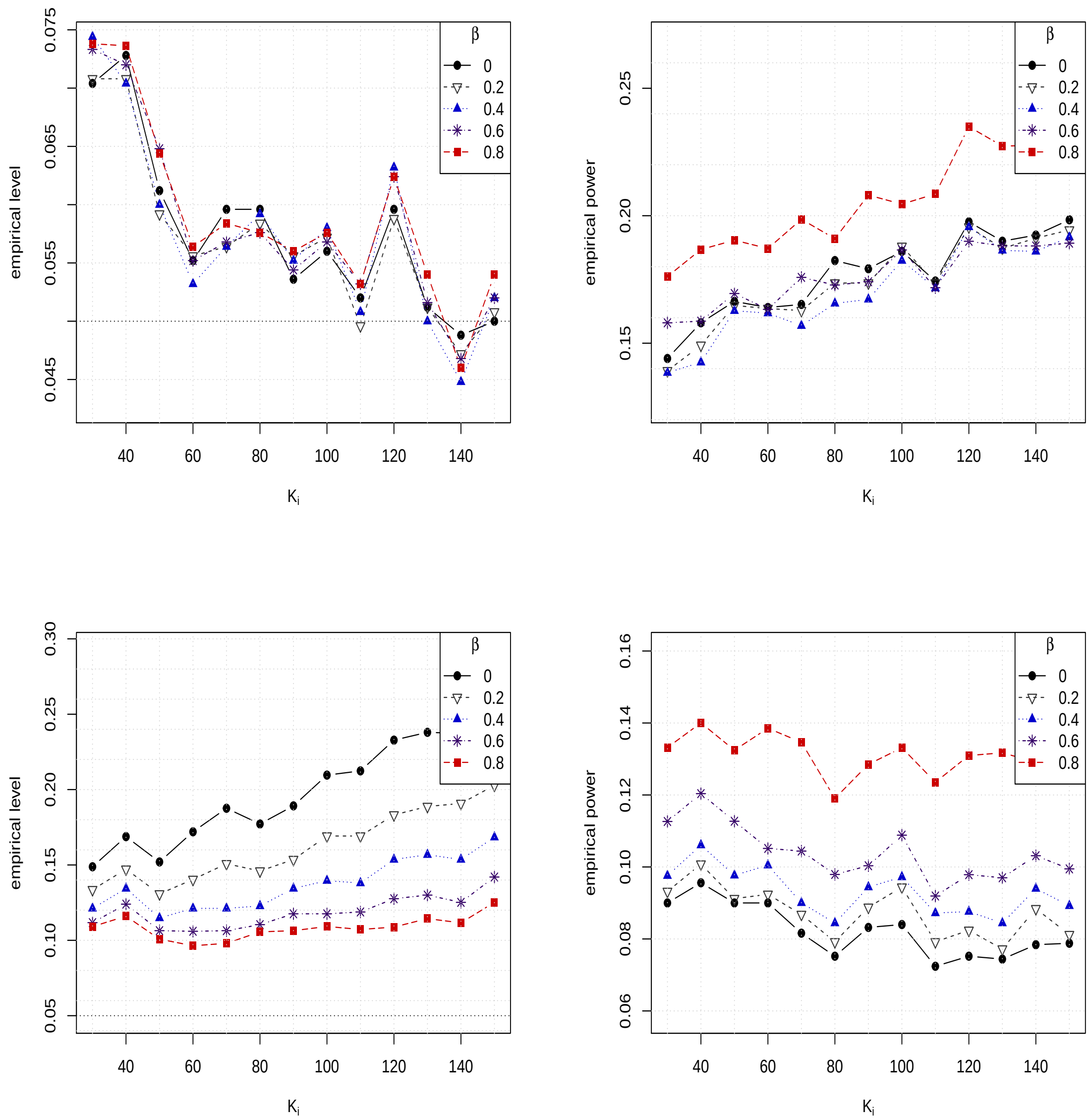

Figure 5.3.4: Weibull distribution at multiple stress levels: empirical levels and powers of pure (top) and contaminated data (bottom)

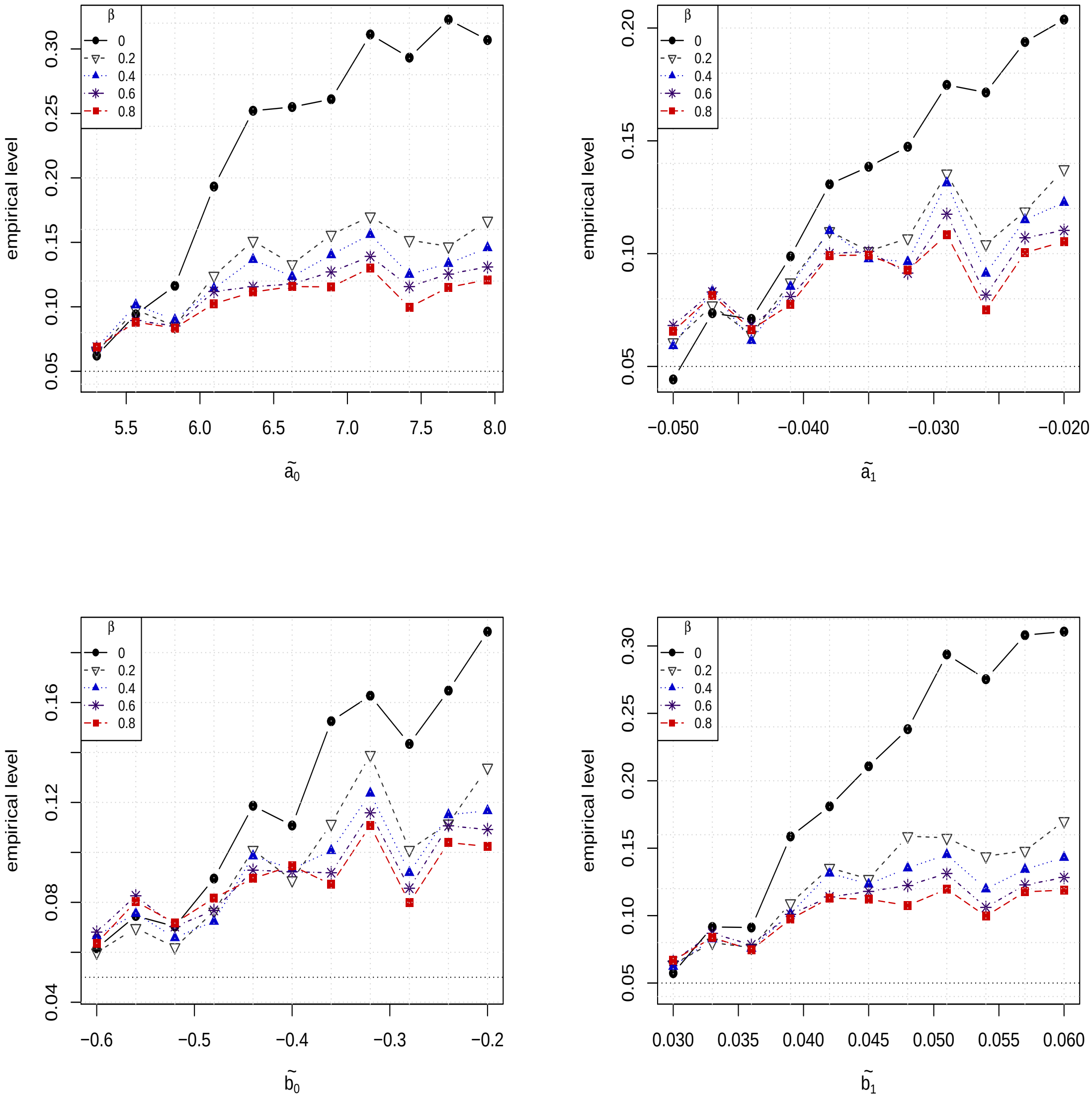

Figure 5.3.5: Weibull distribution at multiple stress levels: empirical levels for unbalanced data

Once again, predicted probabilities are compared to the observed ones for different values of the tuning parameter, β, (bottom of Figure 5.4.3). Both probability vectors are quite close and also that there is not much difference in the estimates between different choices of β, although it is seen that MLE is probably the worst estimator in this case.

5.4.3 Mice Tumor Toxicological data

Let us consider again the Mice Tumor Toxicological data (Kodell and Nelson [1980]). These data, studied for the exponential model in Section 3.5.1, consisted of 1816 mice, of which 553 had tumors, involving the strain of offspring (F1 or F2), gender (females or males), and concentration of benzidine dihydrochloridem (60 ppm, 120 ppm, 200 ppm or 400 ppm) as the stress factors. For each testing condition, the numbers of mice tested and the numbers of mice having tumors were all recorded.

Weighted minimum DPD estimators are obtained for different values of the tuning parameter, and expected lifetimes are computed for different values of concentration, gender and strain. As seen in Figure 5.4.2, there is a significant difference between genders, with males having a higher expected lifetime. This difference is even more remarkable for $\beta = 0.8$. The effect of Strain is not so clear, with slightly better results for mice from F2 group.

5.4.4 Choice of tuning parameter

Let us think now about the problem of the choice of the optimal tuning parameter, given any data set. We could apply the Algorithm 1 (Section 2.6.3) with $\boldsymbol{J}_\beta(\boldsymbol{\theta}_\beta)$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta}_\beta)$ as given in equations (5.6) and (5.7), respectively, in order to avoid complex computations.

We consider the unbalanced data studied in the Simulation Study and we apply this approach for the contamination in the first component with different choices of the pilot estimator. Results are shown in Figure 5.4.1. As contamination increases, the chosen tuning parameter tends to be larger. It seems that a moderate low value of the pilot estimator offers the best trade-off between pure and contaminated data. Although not presented here, similar conclusions were obtained under the other contamination schemes. So, we suggest to use the pilot choice $\boldsymbol{\theta}_P = \widehat{\boldsymbol{\theta}}_{0.4}$.

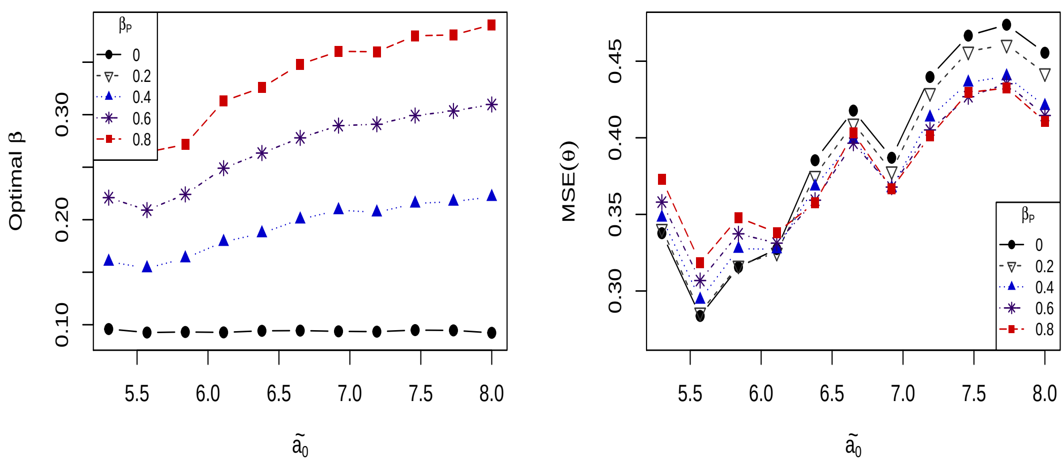

Figure 5.4.1: Weibull distribution at multiple stress levels: simulated MSEs of the weighted minimum DPD estimators at the optimally chosen β, starting from different pilot estimators

Let us apply this procedure to the previous data sets. The corresponding results are shown in Table 5.4.2. As can be seen, MLE is not the best choice in any case, which clearly demonstrates the need for the proposed weighted minimum DPD estimators.

Table 5.4.2: Weibull distribution at multiple stress levels: Optimal β for different data sets

Data	Glass Capacitors	Solder joints	Mice Tumors
β_{opt}	0.94	0.37	0.41

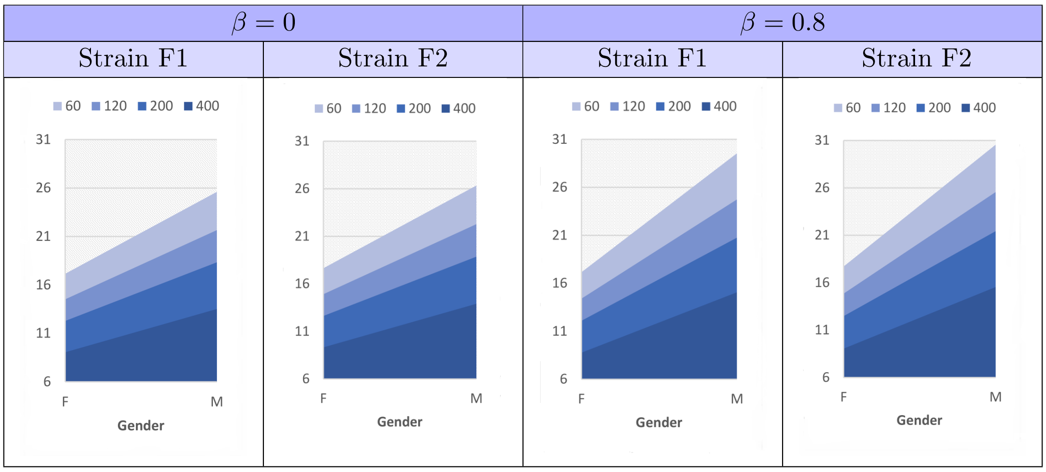

Figure 5.4.2: Estimated lifetimes in Tumor Toxicological Experiment.

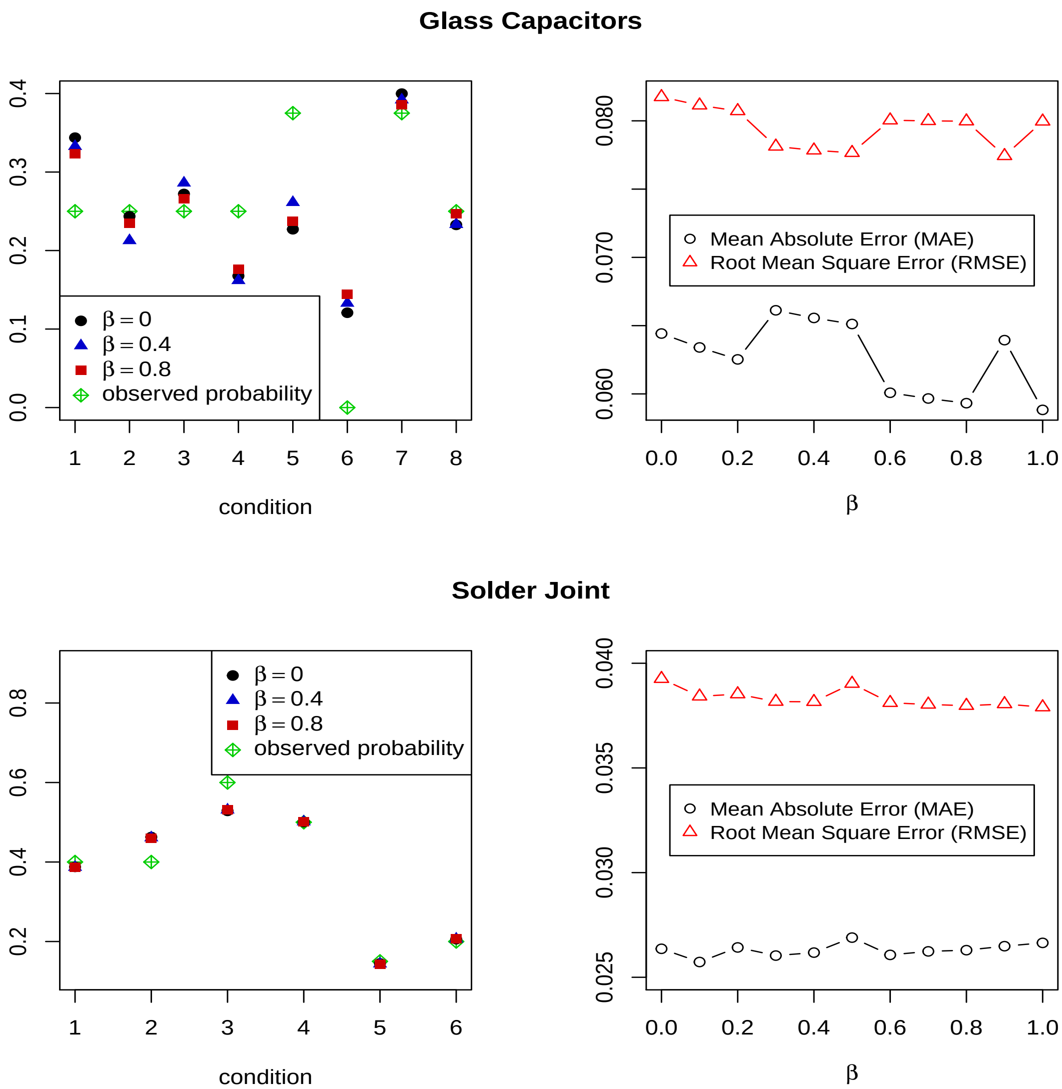

Figure 5.4.3: Glass Capacitors and Solder Joint examples. Left: estimated vs. observed probabilities. Right: RMSEs and MAEs of the estimated probabilities

Chapter 6

Robust inference for one-shot device testing under other distributions: Lindley and lognormal distributions

6.1 Introduction

We have studied the problem on one-shot device testing under the assumption of exponential, gamma and Weibull distributions. However, other distributions may be considered for modeling the lifetimes. In this chapter, we consider the Lindley and lognormal distributions.

The Lindley distribution, introduced by Lindley [1958], has shown to give better modeling that the exponential distribution in some contexts (see Ghitany et al. [2008]). Gupta and Singh [2013] studied the parametric estimation of Lindley distribution with hybrid censored data while Mazucheli and Achcar [2011] applied this distribution to competing risks in lifetime data. On the other hand, the lognormal distribution has been studied in different types of censored data. Meeker [1984] compared accelerated life-test plans for Weibull and lognormal distributions under Type-I censoring. Ng et al. [2002] developed EM algorithm for estimating the parameters of lognormal distributions based on progressively censored data.

After introducing formally both distributions, the chapter is organized as follows: in Section 6.2, inference for one-shot devices under Lindley lifetimes is developed. Same is done under the lognormal assumption in Section 6.3. Section 6.4 focuses on the development of Wald-type tests for both cases. Extensive simulation studies are presented in Section 6.5 and Section 6.6, respectively. Finally, some numerical examples are provided in Section 6.7.

6.1.1 The Lindley distribution

Let us suppose that the true lifetime follows a Lindley distribution with unknown failure rate $\lambda_i(\boldsymbol{\theta})$, related to the stress factor $\boldsymbol{x}_i$ in loglinear form as

$$\lambda_i = \lambda_i(\boldsymbol{\theta}) = \exp(\boldsymbol{x}_i^T\boldsymbol{\theta}), \tag{6.1}$$

where $\boldsymbol{x}_i = (x_{i0}, x_{i1}, \dots, x_{iJ})^T$, and $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_J)^T$. Thus, here $\Theta = \mathbb{R}^{J+1}$. The corresponding density function and distribution function are, respectively,

$$f(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\lambda_i^2}{1+\lambda_i}(1+t)\exp\{-\lambda_i t\} \tag{6.2}$$

and

$$F(t; \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - \frac{1+\lambda_i+t\lambda_i}{1+\lambda_i}\exp\{-\lambda_i t\}. \tag{6.3}$$

On the other hand, the reliability at time t and the mean lifetime under normal operating conditions $\boldsymbol{x}_i$ are given by

$$R(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = 1 - F(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \frac{1+\lambda_i+t\lambda_i}{1+\lambda_i}\exp\{-\lambda_i t\} \tag{6.4}$$

and

$$E[T_i] = \frac{2+\lambda_i}{\lambda_i(1+\lambda_i)}.$$

The hazard function, given by the ratio of the density function and the reliability function, is

$$h(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \frac{\lambda_i}{1+\lambda_i+t\lambda_i}(1+t).$$

Remark 6.1 *The density probability function (6.2) can be expressed as*

$$f(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \frac{\lambda_i}{1+\lambda_i}f_1(t;\boldsymbol{x}_i,\boldsymbol{\theta}) + \frac{1}{1+\lambda_i}f_2(t;\boldsymbol{x}_i,\boldsymbol{\theta}),$$

where

$$f_1(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \lambda_i\exp\{-\lambda_i t\} \quad \textit{and} \quad f_2(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \lambda_i^2 t\exp\{-\lambda_i t\}.$$

Thus, Lindley distribution is a mixture of exponential and gamma distributions with mixing proportions $\frac{\lambda_i}{1+\lambda_i}$ and $\frac{1}{1+\lambda_i}$, respectively.

While the gamma distribution generalizes the exponential one, it requires of numerical integration and will need of the estimation of $2(J+1)$ parameters against the $J+1$ parameters in the exponential model (see Chapter 4). Lindley distribution has advantage over the exponential distribution that the exponential distribution has constant hazard rate and mean residual life function (see Figure 6.1.1) whereas the Lindley distribution has increasing hazard rate and decreasing mean residual life function (see Shanker et al. [2015]).

6.1.2 The lognormal distribution

We can also assume that the lifetimes of the units, under the testing condition i, follow lognormal distribution with corresponding probability density function and cumulative distribution function

$$f(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_i t}exp\left\{-\frac{log(\lambda_i t)^2}{2\sigma_i^2}\right\} \tag{6.5}$$

and

$$F(t;\boldsymbol{x}_i,\boldsymbol{\theta}) = \Phi\left(\frac{log(\lambda_i t)}{\sigma_i}\right) \tag{6.6}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and λ_i and σ_i are, respectively, the scale and shape parameters, which we assume are related to the stress factors in loglinear forms as

$$\lambda_i = exp\left\{\sum_{j=0}^{J} a_j x_{ij}\right\} \quad \text{and} \quad \sigma_i = exp\left\{\sum_{j=0}^{J} b_j x_{ij}\right\}, \tag{6.7}$$

with $x_{i0} = 1$ for all i and $\boldsymbol{\theta} = (a_1, .., a_J, b_1, .., b_J)$.

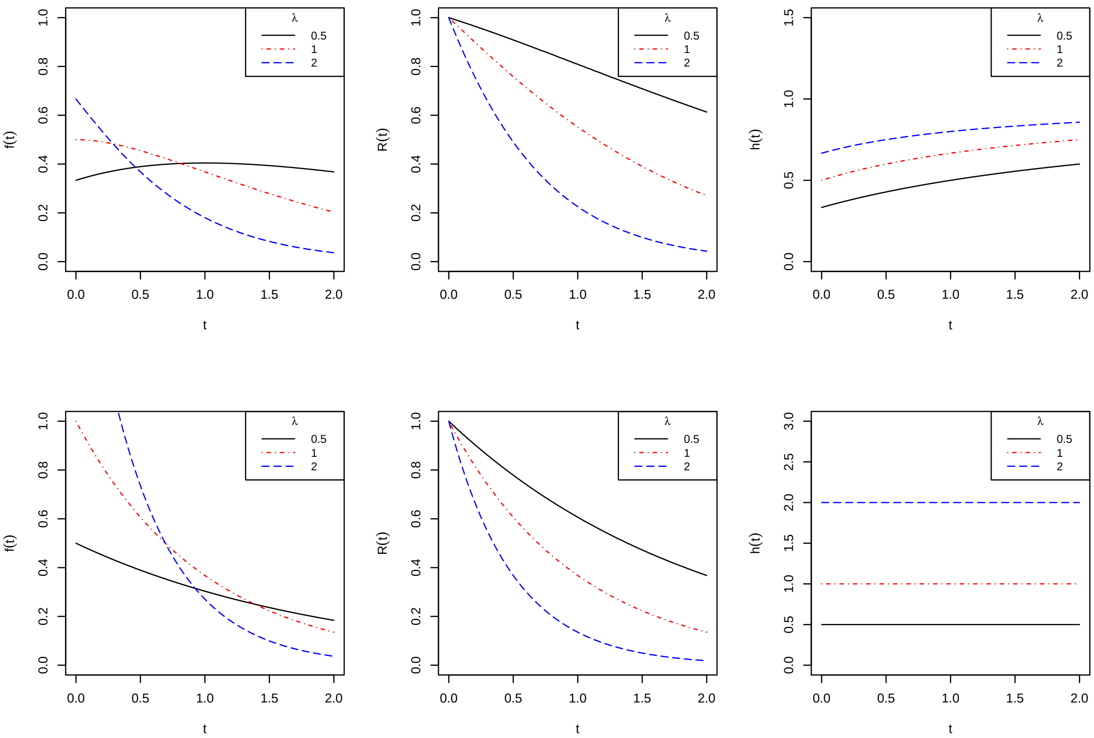

Figure 6.1.1: Denisty, reliability and hazard functions of Lindely (top) and exponential (bottom) distributions

6.2 Inference under the Lindley distribution

Theorem 6.2 *Let us consider the model described in Table 3.1.1 under the Lindley distribution. For* $\beta \geq 0$*, the estimating equations are given by*

$$\sum_{i=1}^{I} \Upsilon_i \left(K_i F\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) - n_i\right) \left[F^{\beta-1}\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right) + R^{\beta-1}\left(IT_i; \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right] \boldsymbol{x}_i = \mathbf{0}_{J+1},$$

where

$$\Upsilon_i = \left(\frac{\lambda_i}{1+\lambda_i}\right)^2 IT_i \left[(IT_i + 1)\lambda_i + IT_i + 2\right] \exp\{-\lambda_i IT_i\}. \tag{6.8}$$

Proof. Straightforward following proof of Theorem 3.2 and taking into account that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \pi_{i1}(\boldsymbol{\theta}) = \left(\frac{\lambda_i}{1+\lambda_i}\right)^2 IT_i \left[(IT_i + 1)\lambda_i + IT_i + 2\right] \exp\{-\lambda_i IT_i\} \boldsymbol{x}_i = \Upsilon_i \boldsymbol{x}_i.$$

■

In the following theorem, we establish the asymptotic distribution of the proposed weighted minimum DPD estimators.

Theorem 6.3 *Let* $\boldsymbol{\theta}^0$ *be the true value of the parameter* $\boldsymbol{\theta}$*. The asymptotic distribution of the weighted minimum DPD estimator* $\widehat{\boldsymbol{\theta}}_\beta$ *is given by*

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \xrightarrow[K\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_{J+1}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_i^I \frac{K_i}{K}\Upsilon_i^2\left(F^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}) + R^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})\right)\boldsymbol{x}_i\boldsymbol{x}_i^T, \tag{6.9}$$

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_i^I \frac{K_i}{K}\Upsilon_i^2 F(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})R(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})\left(F^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}) + R^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})\right)^2\boldsymbol{x}_i\boldsymbol{x}_i^T. \tag{6.10}$$

Proof. Straightforward following proof of Theorem 3.3. ■

6.3 Inference under the lognormal distribution

Theorem 6.4 *Let us consider the model described in Table 3.1.1 under the lognormal distribution. For $\beta \geq 0$, the estimating system of equations is given by*

$$\sum_{i=1}^I \Delta_{i1}\left(K_iF\left(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}\right) - n_i\right)\left[F^{\beta-1}\left(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}\right) + R^{\beta-1}\left(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}\right)\right]\boldsymbol{x}_i = \boldsymbol{0}_{(J+1)}$$
$$\sum_{i=1}^I \Delta_{i2}\left(K_iF\left(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}\right) - n_i\right)\left[F^{\beta-1}\left(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}\right) + R^{\beta-1}\left(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}\right)\right]\boldsymbol{x}_i = \boldsymbol{0}_{(J+1)},$$

with

$$\Delta_{i1} = \phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right) \qquad \Delta_{i2} = \frac{-log(\lambda_i IT_i)}{\sigma_i}\phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right), \tag{6.11}$$

where $\phi(\cdot)$ is the density function of the standard normal distribution

Proof. The estimating equations are given by

$$\sum_{i=1}^I \frac{K_i}{K}\frac{\partial}{\partial \boldsymbol{a}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \boldsymbol{0}_{J+1},$$
$$\sum_{i=1}^I \frac{K_i}{K}\frac{\partial}{\partial \boldsymbol{b}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \boldsymbol{0}_{J+1},$$

with

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{a}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \left(\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \frac{\partial}{\partial \boldsymbol{a}}\pi_{i2}^{\beta+1}(\boldsymbol{\theta})\right) - \frac{\beta+1}{\beta}\left(\widehat{p}_{i1}\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2}\frac{\partial}{\partial \boldsymbol{a}}\pi_{i2}^{\beta}(\boldsymbol{\theta})\right)\\
&= (\beta+1)\left(\pi_{i1}^{\beta}(\boldsymbol{\theta}) - \pi_{i2}^{\beta}(\boldsymbol{\theta}) - \widehat{p}_{i1}\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \widehat{p}_{i2}\pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta})\\
&= (\beta+1)\left(\left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right)\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) - \left(\pi_{i2}(\boldsymbol{\theta}) - \widehat{p}_{i2}\right)\pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta})\\
&= (\beta+1)\left(\left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right)\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right)\pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta})\\
&= (\beta+1)\left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right)\left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta}),\\
\frac{\partial}{\partial \boldsymbol{b}} d_\beta^*(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= (\beta+1)\left(\pi_{i1}(\boldsymbol{\theta}) - \widehat{p}_{i1}\right)\left(\pi_{i1}^{\beta-1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta-1}(\boldsymbol{\theta})\right)\frac{\partial}{\partial \boldsymbol{b}}\pi_{i1}(\boldsymbol{\theta}).
\end{aligned}$$

Then, we have to compute $\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta})$ and $\frac{\partial}{\partial \boldsymbol{b}}\pi_{i1}(\boldsymbol{\theta})$. Following the chain rule

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{a}}\pi_{i1}(\boldsymbol{\theta}) &= \left[\frac{\partial}{\partial \boldsymbol{a}}\lambda_i\right]\frac{\partial}{\partial \lambda_i}\Phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right)\\
&= \left[\frac{\partial}{\partial \boldsymbol{a}}\lambda_i\right]\left[\frac{\partial}{\partial log(\lambda_i IT_i)}\Phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right)\right]\frac{\partial}{\partial \lambda_i}log(\lambda_i IT_i)
\end{aligned}$$

$$
\begin{aligned}
&= \left[\frac{\partial}{\partial \boldsymbol{a}}\lambda_i\right]\left[\frac{\partial}{\partial log(\lambda_i IT_i)}\Phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right)\right]\frac{1}{\lambda_i} = \left[\frac{\partial}{\partial \boldsymbol{a}}\lambda_i\right]\phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right)\frac{1}{\lambda_i} \\
&= \phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right)\boldsymbol{x}_i = \Delta_{i1}\boldsymbol{x}_i \qquad (6.12)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{b}}\pi_{i1}(\boldsymbol{\theta}) &= \left[\frac{\partial}{\partial \boldsymbol{b}}\sigma_i\right]\frac{\partial}{\partial \sigma_i}\Phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right) \\
&= \left[\frac{\partial}{\partial \boldsymbol{b}}\sigma_i\right]\frac{-log(\lambda_i IT_i)}{\sigma_i^2}\phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right) \\
&= \frac{-log(\lambda_i IT_i)}{\sigma_i}\phi\left(\frac{log(\lambda_i IT_i)}{\sigma_i}\right)\boldsymbol{x}_i = \Delta_{i2}\boldsymbol{x}_i. \qquad (6.13)
\end{aligned}
$$

Then, the result follows. ■

Theorem 6.5 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter $\boldsymbol{\theta}$. The asymptotic distribution of the weighted minimum DPD estimator $\widehat{\boldsymbol{\theta}}_\beta$ is given by*

$$
\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\mathbf{0}_{2(J+1)}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),
$$

where

$$
\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_i^I \frac{K_i}{K}\boldsymbol{\Delta}_i\left(F^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}) + R^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})\right), \qquad (6.14)
$$

$$
\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_i^I \frac{K_i}{K}\boldsymbol{\Delta}_i F(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})R(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})\left(F^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta}) + R^{\beta-1}(IT_i;\boldsymbol{x}_i,\boldsymbol{\theta})\right)^2, \qquad (6.15)
$$

where

$$
\boldsymbol{\Delta}_i = \begin{pmatrix} \Delta_{i1}^2\boldsymbol{x}_i\boldsymbol{x}_i^T & \Delta_{i1}\Delta_{i2}\boldsymbol{x}_i\boldsymbol{x}_i^T \\ \Delta_{i1}\Delta_{i2}\boldsymbol{x}_i\boldsymbol{x}_i^T & \Delta_{i2}^2\boldsymbol{x}_i\boldsymbol{x}_i^T \end{pmatrix} \qquad (6.16)
$$

and Δ_{i1} and Δ_{i2} were given in (6.11).

Proof. Straightforward following proof of Theorem 4.2 and equations (6.12) and (6.13). ■

6.4 Wald-type tests

From Theorem 6.3 and Theorem 6.5, and following the idea on previous chapters, we can develop Wald-type tests for testing composite null hypotheses.

Let us consider the function $\boldsymbol{m} : \mathbb{R}^S \longrightarrow \mathbb{R}^r$, where $r \leq S$ and $S = J+1$ (Lindley distribution) or $S = 2(J+1)$ (lognormal distribution). Then, $\boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{0}_r$ represents a composite null hypothesis. We assume that the $S \times r$ matrix

$$
\boldsymbol{M}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{m}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}
$$

exists and is continuous in $\boldsymbol{\theta}$ and with rank $\boldsymbol{M}(\boldsymbol{\theta}) = r$. For testing

$$
H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \text{ against } H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0, \qquad (6.17)
$$

where $\boldsymbol{\Theta}_0 = \left\{\boldsymbol{\theta} \in \mathbb{R}^S : \boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{0}_r\right\}$, we can consider the following Wald-type test statistics

$$
W_K(\widehat{\boldsymbol{\theta}}_\beta) = K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta), \qquad (6.18)
$$

where $\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta) = \boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)$ and $\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are as in (6.9) and (6.10) (Lindley distribution) or in equations (6.14) and (6.15) (lognormal distribution), respectively.

Theorem 6.6 *The asymptotic null distribution of the proposed Wald-type test statistics, given in Equation (6.18), is a chi-squared (χ^2) distribution with p degrees of freedom. This is,*

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) \xrightarrow[K\to\infty]{\mathcal{L}} \chi^2_r.$$

Proof. Straightforward taking into account the asymptotic distribution of the proposed minimum DPD estimators. ■

Based on Theorem 5.5, we will reject the null hypothesis in (5.10) if

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) > \chi^2_{r,\alpha},$$

where $\chi^2_{r,\alpha}$ is the upper percentage point of order α of χ^2_r distribution.

6.5 Simulation study under the Lindley distribution

In this section, we develop a simulation study in order to evaluate the performance of the proposed weighted minimum DPD estimators and Wald-type tests under the assumption of Lindley lifetimes. We consider different scenarios both for balanced and unbalanced data (equal or different sample size K_i in each condition i, respectively). In the case of balanced data, different reliabilities and sample sizes are taken for both pure and contaminated data, while in the case of unbalanced data, the performance is evaluated under different degrees of contamination. The results are recorded and averaged over 1000 simulation runs, in the R statistical software.

6.5.1 The weighted minimum DPD estimators

First of all, we evaluate the robustness of the estimators by means of the RMSE of the parameter vector $\boldsymbol{\theta}$ for different values of the tuning parameter $\beta \in \{0, 0.2, 0.4, 0.6\}$. The main purpose of this study is to show how there are alternative estimators to the MLE ($\beta = 0$) that can offer a better performance in terms of robustness.

A. Balanced data

The lifetimes of devices are simulated from the Lindley distribution, for different levels of reliability and different samples sizes $K_i \in (40, ..., 150)$, under $I = 12$ conditions with two stress factors at two levels each one, tested at three different inspection times. The parameter values used in this simulation are detailed in Table 6.5.1 and the results are shown in Figure 6.5.1.

With independence of the reliability considered, we observe how, when a pure scheme is evaluated, the MLE presents the highest efficiency, while in a contamination scheme, this estimator becomes the worst, with the greatest error. Bias of estimates of reliabilities at normal conditions and different times, as well as the RMSE of the parameter estimates, is computed with the contamination of the last two cells $\tilde{\theta}_1 = 0.026$ and $\tilde{\theta}_2 = 0.026$, under medium reliability. Results are presented in Table 6.5.2, with similar conclusions.

B. Unbalanced data

In this scheme, we consider an unbalanced data, in which each condition is evaluated under a different sample size (Table 6.5.3). The data are generated under low reliability, with $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^T = (-5.5, 0.03, 0.03)^T$. To contaminate the data, the last cell is generated by $\tilde{\boldsymbol{\theta}} = (\theta_0, \tilde{\theta}_1, \theta_2)^T$ or $\tilde{\boldsymbol{\theta}} = (\theta_0, \theta_1, \tilde{\theta}_2)^T$, with the degree of contamination measured by $4(1 - \tilde{\theta}_1/\theta_1) \in (0, ..., 1)$ and $4(1 - \tilde{\theta}_2/\theta_2) \in (0, ..., 1)$. Results are shown in top of Figure 6.5.2. When a pure data scheme is considered ($(1 - \tilde{\theta}_1/\theta_1) = 0$ or $(1 - \tilde{\theta}_2/\theta_2) = 0$), the MLE presents the lowest error. This changes when increasing the dregree of contamination, with alternative estimators with $\beta > 0$ presenting a much more robust behaviour. As expected, when increasing the contamination, the error also increases, specially when $\beta = 0$.

Table 6.5.1: Lindley distribution at multiple stress levels: parameter values used in the simulation. Efficiency.

Reliability	Parameters	Symbols	Values
High reliability	Number of Conditions	I	12
	True parameters	$\theta_0,\, \theta_1,\, \theta_2$	$-6, 0.03, 0.03$
	Contamination	$\tilde{\theta}_2$	0.025
	First Stress factor	$x_{i1},\ i = 1, \dots, I$	$\{55, 85\}$
	Second Stress factor	$x_{i2},\ i = 1, \dots, I$	$\{70, 100\}$
	Inspection Time	$IT_i,\ i = 1, \dots, I$	$\{2, 5, 8\}$
Moderate reliability	Number of Conditions	I	12
	True parameters	$\theta_0,\, \theta_1,\, \theta_2$	$-5.5, 0.03, 0.03$
	Contamination	$\tilde{\theta}_2$	0.024
	First Stress factor	$x_{i1},\ i = 1, \dots, I$	$\{55, 85\}$
	Second Stress factor	$x_{i2},\ i = 1, \dots, I$	$\{70, 100\}$
	Inspection Time	$IT_i,\ i = 1, \dots, I$	$\{1.5, 4.5, 7.5\}$
Low reliability	Number of Conditions	I	12
	True parameters	$\theta_0,\, \theta_1,\, \theta_2$	$-5, 0.03, 0.03$
	Contamination	$\tilde{\theta}_2$	0.023
	First Stress factor	$x_{i1},\ i = 1, \dots, I$	$\{55, 85\}$
	Second Stress factor	$x_{i2},\ i = 1, \dots, I$	$\{70, 100\}$
	Inspection Time	$IT_i,\ i = 1, \dots, I$	$\{1, 4, 7\}$

6.5.2 The Wald-type tests

To evaluate the performance of the proposed Wald-type tests, we consider the scenario of unbalanced data proposed in the previous section. We consider the testing problem

$$H_0 : \theta_0 = -5.5 \quad \text{against} \quad H_1 : \theta_0 \neq -5.5. \tag{6.19}$$

We first evaluate the empirical levels, measured as the proportion of test statistics exceeding the corresponding chi-square critical value for a nominal size $\alpha = 0.05$. The empirical powers are computed in a similar way, with $\theta_0^0 = -0.75$. Results are shown in Figure 6.5.2.

6.6 Simulation study under the lognormal distribution

In this section, a simulation study evaluate the performance of the proposed weighted minimum DPD estimators and Wald-type tests under the assumption of lognormal lifetimes. Once, again, we consider different scenarios both for balanced and unbalanced data (equal or different sample size K_i in each condition i, respectively). In the case of balanced data, different sample sizes are taken for both pure and contaminated data, while in the case of unbalanced data, the performance is evaluated under different degrees of contamination. The results are recorded and averaged over 1000 simulation runs, in the R statistical software.

6.6.1 The weighted minimum DPD estimators

Let us evaluate the robustness of the estimators by means of the RMSE of the parameter vector $\boldsymbol{\theta}$ for different values of the tuning parameter $\beta \in \{0, 0.2, 0.4, 0.6\}$.

Table 6.5.2: Lindley distribution at multiple stress levels: bias of the estimates of reliabilities for pure and contaminated data in the case of moderate reliability. Two-cells contamination

High reliability		Pure data				Contaminated data			
	True value	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
$K_i = 50$									
$R(10; (45, 60))$	0.7738	-0.0044	-0.0051	-0.0053	-0.0054	-0.4407	-0.4081	-0.3490	-0.1816
$R(20; (45, 60))$	0.4874	-0.0023	-0.0032	-0.0035	-0.0034	-0.4188	-0.4027	-0.3675	-0.2156
$R(30; (45, 60))$	0.2791	0.0024	0.0015	0.0014	0.0016	-0.2667	-0.2618	-0.2488	-0.1595
$R(40; (45, 60))$	0.1512	0.0053	0.0047	0.0047	0.0050	-0.1491	-0.1479	-0.1438	-0.0982
$R(45; (45, 60))$	0.1097	0.0059	0.0054	0.0055	0.0057	-0.1089	-0.1083	-0.1061	-0.0741
$RMSE(\boldsymbol{\theta})$	-	0.1329	0.1332	0.1358	0.1384	1.3142	1.2285	1.0720	0.6058
$K_i = 100$									
$R(10; (45, 60))$	0.7738	-0.0018	-0.0022	-0.0024	-0.0025	-0.4394	-0.4066	-0.3479	-0.1777
$R(20; (45, 60))$	0.4874	-0.0005	-0.0010	-0.0013	-0.0012	-0.4194	-0.4033	-0.3691	-0.2196
$R(30; (45, 60))$	0.2791	0.0019	0.0014	0.0013	0.0014	-0.2671	-0.2624	-0.2503	-0.1673
$R(40; (45, 60))$	0.1512	0.0032	0.0029	0.0029	0.0031	-0.1493	-0.1481	-0.1446	-0.1053
$R(45; (45, 60))$	0.1097	0.0034	0.0032	0.0032	0.0034	-0.1090	-0.1085	-0.1066	-0.0803
$RMSE(\boldsymbol{\theta})$	-	0.0938	0.0952	0.0981	0.1006	1.3102	1.2242	1.0696	0.5912
$K_i = 150$									
$R(10; (45, 60))$	0.7738	0.0002	0.0001	0.0000	0.0000	-0.4377	-0.4048	-0.3463	-0.1747
$R(20; (45, 60))$	0.4874	0.0018	0.0016	0.0015	0.0016	-0.4191	-0.4030	-0.3690	-0.2202
$R(30; (45, 60))$	0.2791	0.0031	0.0029	0.0029	0.0031	-0.2672	-0.2625	-0.2506	-0.1699
$R(40; (45, 60))$	0.1512	0.0034	0.0033	0.0034	0.0036	-0.1493	-0.1482	-0.1448	-0.1079
$R(45; (45, 60))$	0.1097	0.0033	0.0033	0.0033	0.0035	-0.1090	-0.1085	-0.1068	-0.0826
$RMSE(\boldsymbol{\theta})$	-	0.0720	0.0729	0.0749	0.0768	1.3079	1.2220	1.0683	0.5879

Table 6.5.3: Lindley distribution at multiple stress levels: ALT plan, unbalanced data.

i	x_{i1}	x_{i2}	IT_i	K_i
1	55	70	1.5	90
2	55	100	1.5	90
3	85	70	1.5	75
4	85	100	1.5	75
5	55	70	4.5	75
6	55	100	4.5	75
7	85	70	4.5	75
8	85	100	4.5	75
9	55	70	7.5	75
10	55	100	7.5	60
11	85	70	7.5	60
12	85	100	7.5	30

A. Balanced data

The lifetimes of devices are simulated for different sample sizes, under 3 different stress conditions with 1 stress factor at 3 levels, $x \in \{30, 40, 50\}$. Then, all devices under each stress condition are inspected at 3 different inspection times, $IT \in \{12, 24, 36\}$. The corresponding data will then be collected under $I = 9$ test conditions. A balanced data with equal sample size for each group was

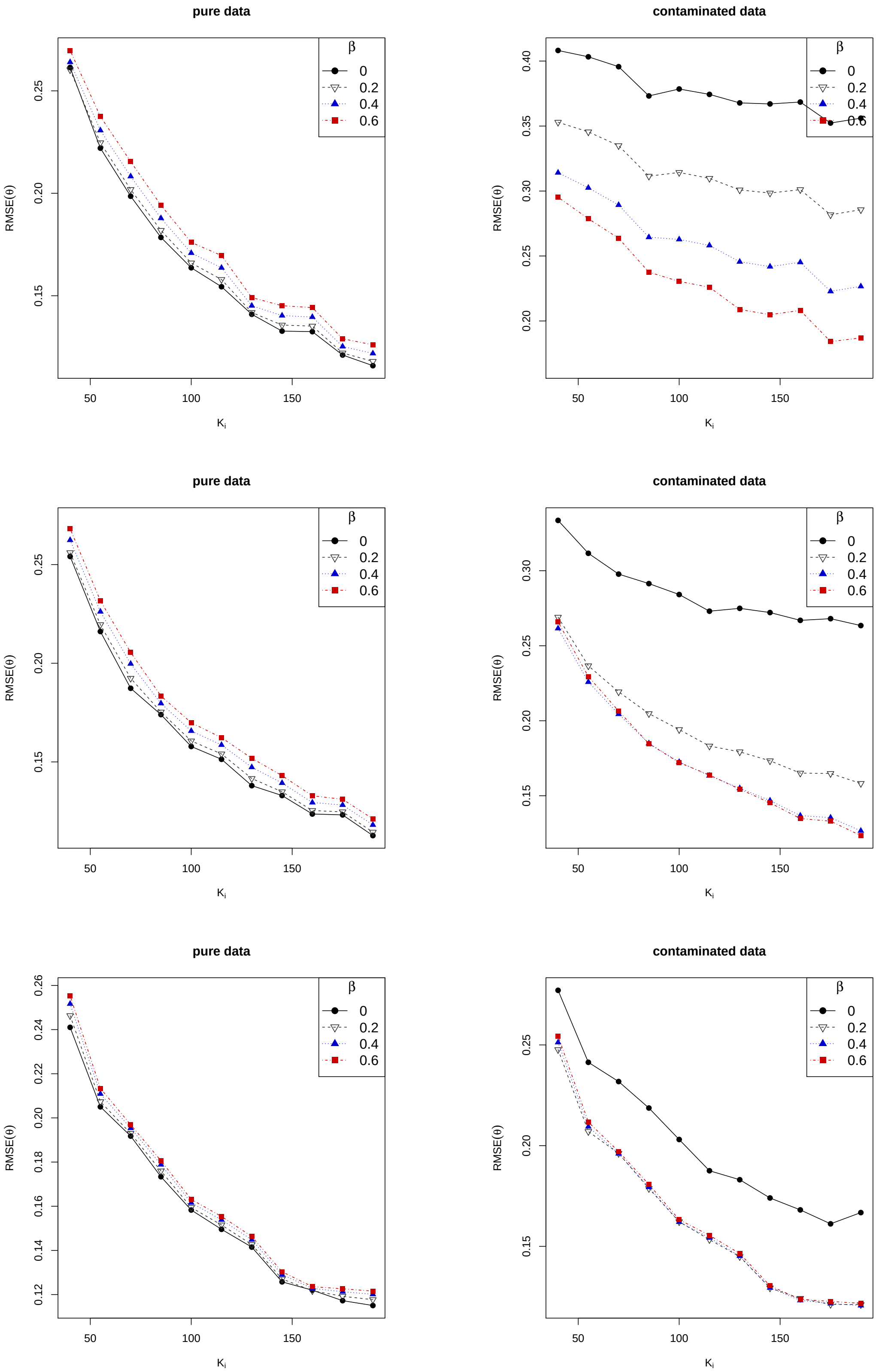

Figure 6.5.1: Lindley distribution at multiple stress levels: RMSEs of the vector of parameters for pure (left) and contaminated (right) data at high (top), moderate (medium) and low (bottom) reliability

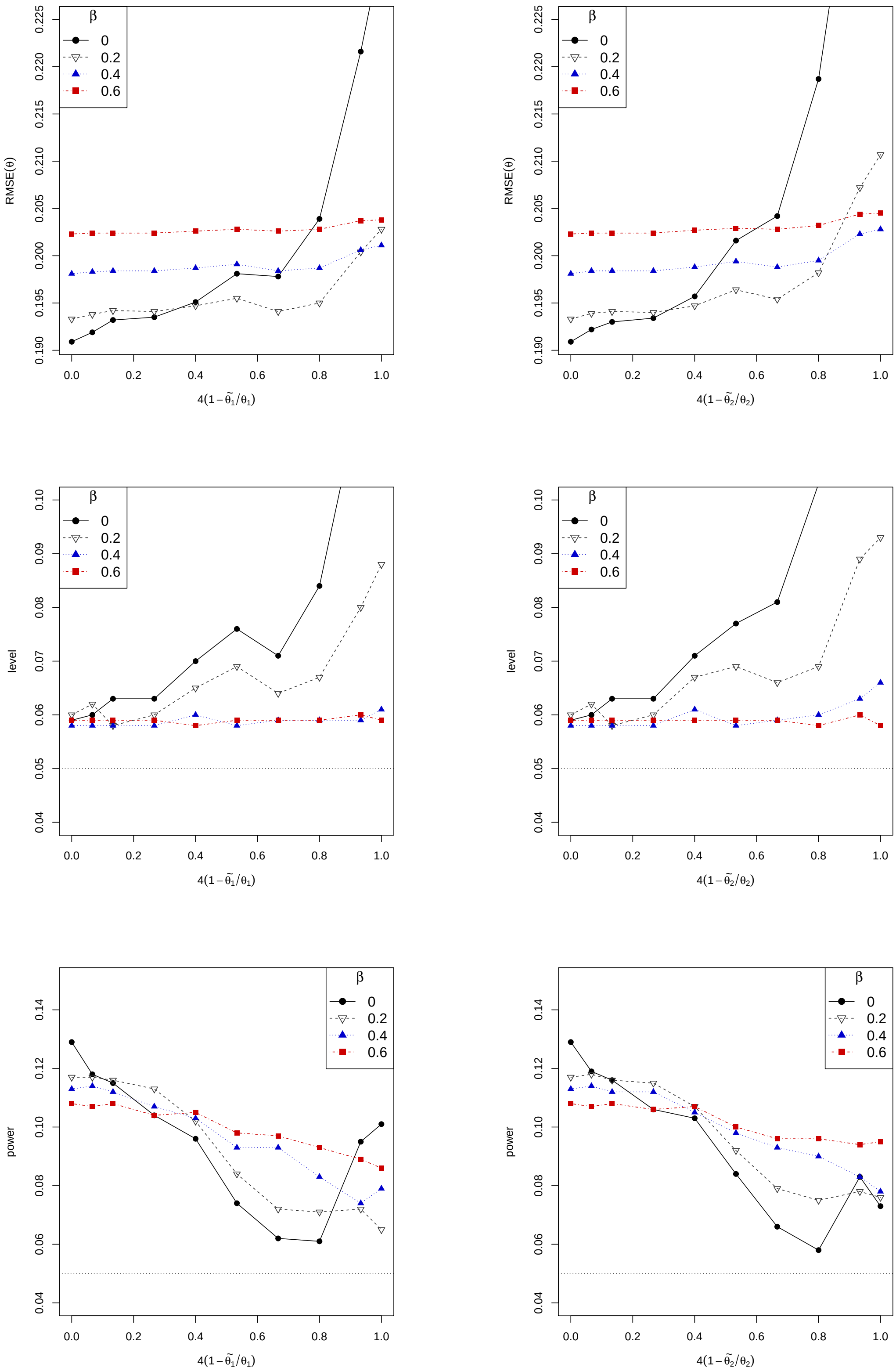

Figure 6.5.2: Lindley distribution at multiple stress levels: RMSE of the vector of parameters (top) empirical lavels (middle) and empirical powers (bottom) under different degreees of contamination

considered. K_i was taken to range from small to large sample sizes, and the model parameters were set to be $\boldsymbol{\theta}^T = (-6, 0.03, -0.6, 0.03)^T$. To evaluate the robustness of the weighted minimum DPD estimators, we studied their behavior in the presence of an outlying cell for the first testing condition in our table. This cell was generated under the parameters $\tilde{\boldsymbol{\theta}}^T = (-5.7, 0.03, -0.6, 0.03)^T$. Bias of model parameters, as well as bias for the reliability at normal testing conditions $(IT_0, x_0) = (60, 25)$, were then computed, for different tuning parameters, for the cases of both pure and contaminated data and are presented in Tables 6.6.1 and 6.6.2, respectively.

For the case of pure data, MLE presents the best behaviour and an increment in the tuning parameter β leads to a gradual loss in terms of efficiency. However, in the case of contaminated data, MLE turns to be the worst estimator, and weighted minimum DPD estimators with $\beta > 0$ present much more robust behaviour. Note that, as expected, an increase in the sample size improves the efficiency of the estimators, both for pure and contaminated data.

Table 6.6.1: Lognormal distribution at multiple stress levels: bias for the parameter vector.

	Pure data				Contaminated data			
K_i	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
40	1.0681	1.0455	1.0455	1.0600	1.0536	0.9811	0.9811	0.9789
50	0.8710	0.8647	0.8634	0.8891	0.9420	0.8932	0.8543	0.8593
60	0.8362	0.8509	0.8534	0.9021	0.9469	0.8869	0.8685	0.8533
70	0.7518	0.7617	0.7653	0.7802	0.8977	0.8210	0.7850	0.7741
80	0.7113	0.7121	0.7225	0.7338	0.8952	0.8135	0.7651	0.7478
90	0.6781	0.6830	0.7067	0.7118	0.8597	0.7780	0.7226	0.7100
100	0.6505	0.6458	0.6571	0.6700	0.8417	0.7520	0.7032	0.6871
110	0.6026	0.6011	0.6131	0.6297	0.8394	0.7469	0.6871	0.6613
120	0.5589	0.5603	0.5766	0.5865	0.8279	0.7360	0.6746	0.6524

Table 6.6.2: Lognormal distribution at multiple stress levels: bias for the reliability under normal testing conditions

	Pure data				Contaminated data			
K_i	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
40	0.1099	0.1061	0.1083	0.1084	0.1297	0.1215	0.1193	0.1157
50	0.0918	0.0907	0.0919	0.0911	0.1072	0.1030	0.0990	0.0967
60	0.0889	0.0881	0.0869	0.0874	0.1061	0.1009	0.0968	0.0934
70	0.0796	0.0788	0.0787	0.0780	0.0909	0.0864	0.0853	0.0833
80	0.0771	0.0780	0.0776	0.0767	0.0884	0.0864	0.0836	0.0832
90	0.0702	0.0700	0.0704	0.0698	0.0771	0.0754	0.0747	0.0738
100	0.0696	0.0691	0.0689	0.0682	0.0737	0.0726	0.0721	0.0714
110	0.0680	0.0670	0.0668	0.0665	0.0703	0.0705	0.0693	0.0690
120	0.0652	0.0648	0.0640	0.0636	0.0642	0.0641	0.0642	0.0642

B. Unbalanced data

In this scenario, an unbalanced data with different sample size in each condition is considered (see Table 6.6.3). The data are generated with $\boldsymbol{\theta} = (-6, 0.03, -0.6, 0.03)^T$. To contaminate the data, the first cell is generated by $\tilde{\boldsymbol{\theta}} = (-6, \tilde{a}_1, -0.6, 0.03)^T$, with $(1 - \tilde{a}_1/a_1) \in (0, ..., 1)$. Note that when pure data are considered this value is equal to 0.

Table 6.6.3: Lognormal distribution at multiple stress levels: ALT plan, unbalanced data.

i	x_{i1}	IT_i	K_i
1	30	30	60
2	40	30	40
3	50	30	20
4	30	60	60
5	40	60	20
6	50	60	20
7	30	90	40
8	40	90	20
9	50	90	20

Bias of the parameter vector $\boldsymbol{\theta}$ as well as bias of the reliability evaluated at $(IT_0, x_0) = (80, 20)$ are presented in top of Figure 6.6.1. While for low degrees of contamination the difference is very slight, this becomes important when considering a high degree of contamination.

6.6.2 The Wald-type tests

To evaluate the performance of the proposed Wald-type tests, we consider the scenario of unbalanced data proposed in the previous section. We consider the testing problem

$$H_0 : b_1 = 0.03 \quad \text{against} \quad H_1 : b_1 \neq 0.03. \tag{6.20}$$

We first evaluate the empirical levels, measured as the proportion of Wald-type test statistics exceeding the corresponding chi-square critical value for a nominal size $\alpha = 0.05$. The empirical powers are computed in a similar way, with $b_1^0 = 0.002$. Results are shown in bottom of Figure 6.6.1. Similar conclusions are obtained for the empirical levels, with an increment in the robustness for $\beta > 0$. The behaviour of the empirical powers is not so clear when a high value of β is taken.

6.7 Application of Lindley distribution to real data

6.7.1 The benzidine dihydrochloride experiment

Let us consider a simplified version of data given in Section 2.7.3. While the original study distinguished between mice sacrificed and died without tumor, we consider them in a same category, as our interest is focused on the carcinogenic effect of the drug. We then have an unbalanced data set with sample sizes $(K_1, K_2, K_3, K_4, K_5, K_6) = (72, 25, 49, 35, 46, 11)$ and observed failures $(n_1, n_2, n_3, n_4, n_5, n_6) = (0, 0, 0, 17, 7, 9)$. Figure 6.7.1 shows the estimated probabilities for each one of the 6 observed conditions as well as the estimated RMSE for the probabilities, both under the Exponential and Lindley distribution models and the MLE. The error under the Lindley distribution is clearly inferior to the exponential one. We apply our proposed estimators and evaluate their performance for different tuning parameters (top of Figure 6.7.2). It can be seen how the estimated probability error decrease for $\beta > 0$.

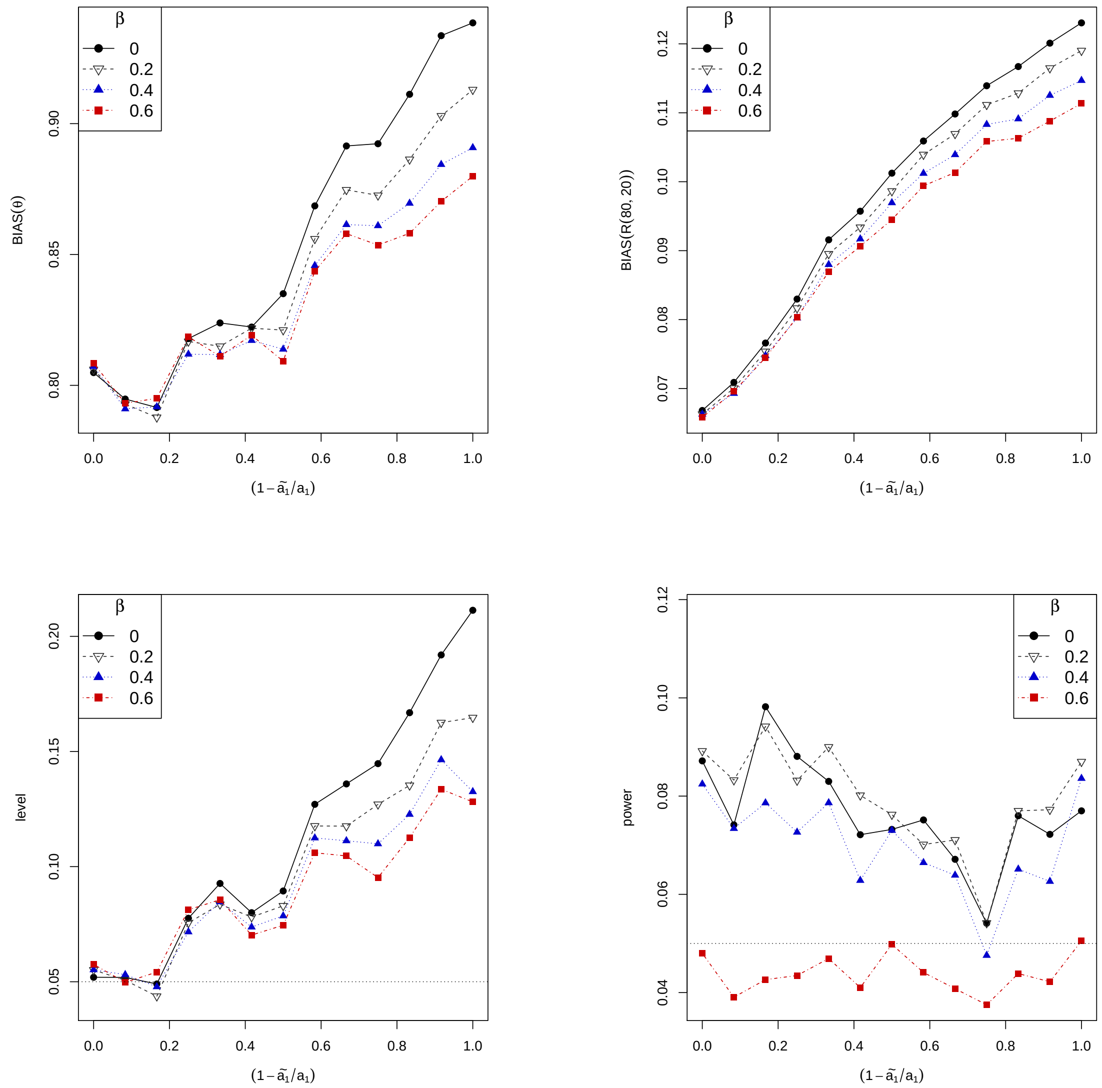

Figure 6.6.1: Lognormal distribution at multiple stress levels: bias for the parameter vector (top left), bias for the reliability under normal testing conditions (top right), emepirical levels (bottom left) and empirical powers (bottom right) for different degrees of contamination

6.7.2 Glass Capacitors

We now consider data from a life test of glass capacitors at higher than usual levels of temperature and voltage, which was already studied in Section 5.4.1. As in the previous example, we apply our proposed estimators and evaluate their performance for different tuning parameters (bottom of Figure 6.7.2), observing again how the estimated probability error decrease for $\beta > 0$.

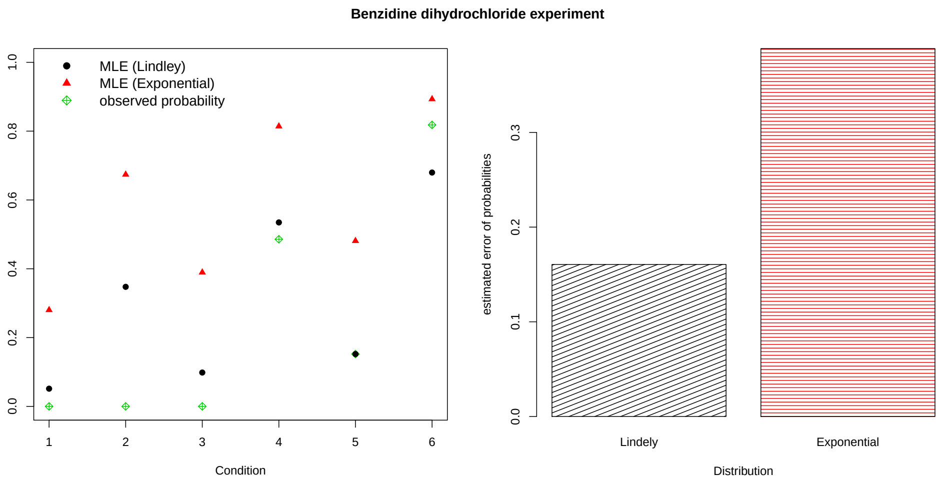

Figure 6.7.1: Lindley and exponential distributions: MLE approach for the benzidine dihydrochloride experiment

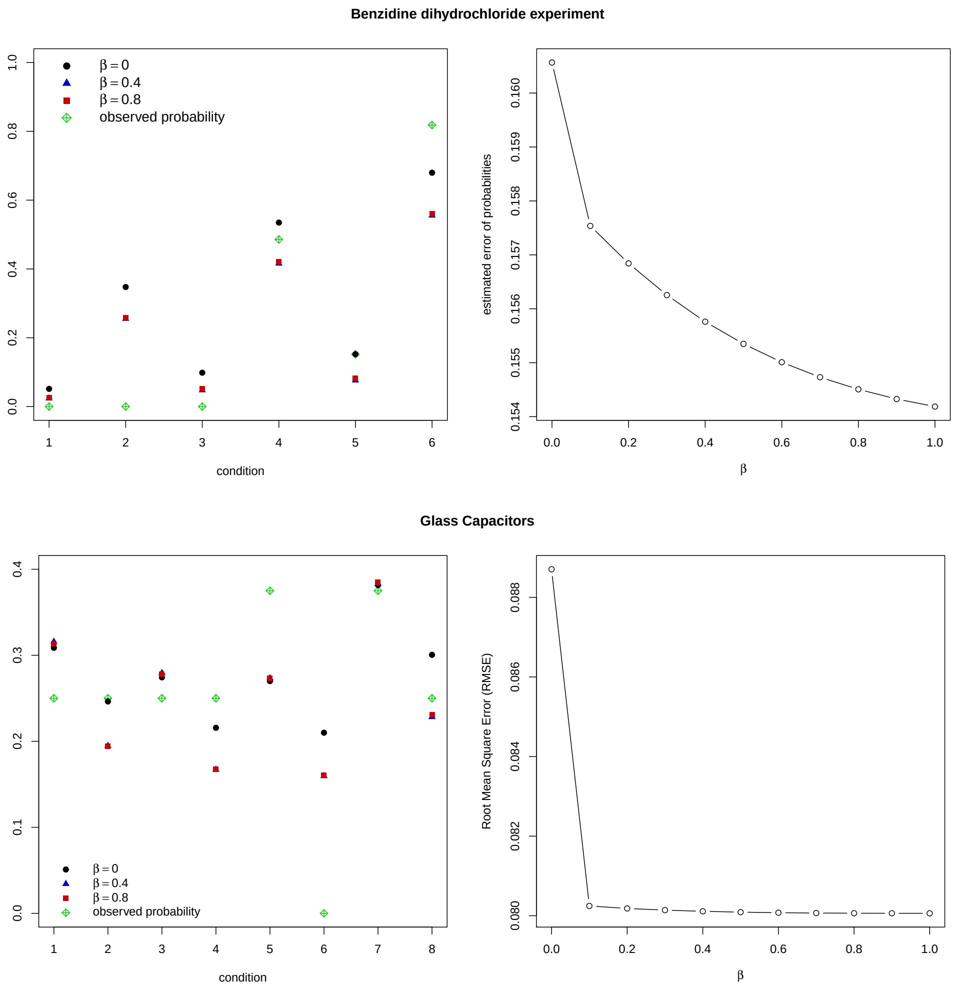

Figure 6.7.2: Lindley distribution: estimated probabilities and their corresponding empirical errors in the benzidine dihydrochloride experiment and glass capacitors examples

Chapter 7

Robust inference for one-shot device testing under proportional hazards model

7.1 Introduction

Under the classical parametric setup, product lifetimes are assumed to be fully described by a probability distribution involving some model parameters. This has been done with some common lifetime distributions such as exponential (Balakrishnan and Ling [2012b]), gamma or Weibull (Balakrishnan and Ling [2013]). However, as data from one-shot devices do not contain actual lifetimes, parametric inferential methods can be very sensitive to violations of the model assumption. Ling et al. [2015] proposed a semi-parametric model, in which, under the proportional hazards assumption, the hazard rate is allowed to change in a non-parametric way. The simulation study carried out by Ling et al. [2015] shows that their proposed method works very well. However, this method suffer again from lack of robustness, as it is based on the (non-robust) MLE of model parameters.

In this chapter, we extend the robust approach proposed in the above chapters and develop here robust estimators and tests for one-shot device testing based on divergence measures under proportional hazards model. Section 7.2 described the model and some basic concepts and results. The estimating equations and asymptotic properties of the proposed estimators are given in Section 7.3. Wald-type tests are then developed based on the proposed estimators, as a generalization of the classical Wald test. In Section 7.5, a simulation study is carried out to demonstrate the robustness of the proposed method. A numerical example is finally presented in Section 7.6.

The results of this Chapter have been published in the form of a paper (Balakrishnan et al. [2021]).

7.2 Model description and Maximum Likelihood Estimator

Consider S constant-stress accelerated life-tests and I inspection times. For the i-th life-test, K_s devices are placed under stress level combinations with J stress factors, $\boldsymbol{x}_s = (x_{s1}, \ldots, x_{sJ})$, of which K_{is} are tested at the i-th inspection time τ_i, where $K_s = \sum_{i=1}^{I} K_{is}$ and $0 < \tau_1 < \cdots < \tau_I$. Then, the numbers of devices that have failed by time τ_i at stress $\boldsymbol{x}_s$ are recorded as n_{is}. One-shot device testing data obtained from such a life-test can then be represented as $(n_{is}, K_{is}, \boldsymbol{x}_s, \tau_i)$, for $i = 1, 2, \ldots, I$ and $s = 1, 2, \ldots, S$.

Under the proportional hazards assumption, the cumulative hazard function is given by

$$H(t, \boldsymbol{x}; \boldsymbol{\eta}, \boldsymbol{\alpha}) = H_0(t; \boldsymbol{\eta})\lambda(\boldsymbol{x}; \boldsymbol{\alpha}), \tag{7.1}$$

where $H_0(t; \boldsymbol{\eta})$ is the baseline cumulative hazard function with $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_I)$, and $\boldsymbol{\alpha} = (\alpha_1 \ldots, \alpha_J)$ is a vector of coefficients for stress factors. We now assume a log-linear link function for relating the stress levels to the failure times of the units in the cumulative hazard function in (7.1), as

$$H(t, \boldsymbol{x}; \boldsymbol{\eta}, \boldsymbol{\alpha}) = H_0(t; \boldsymbol{\eta}) \exp\left(\sum_{j=1}^{J} \alpha_j x_{sj}\right).$$

The corresponding reliability function is given by

$$R(t, \boldsymbol{x}; \boldsymbol{\eta}, \boldsymbol{\alpha}) = \exp\left(-H(t, \boldsymbol{x}; \boldsymbol{\eta}, \boldsymbol{\alpha})\right) = R_0(t; \boldsymbol{\eta})^{\lambda(\boldsymbol{x}; \boldsymbol{\alpha})}, \tag{7.2}$$

where $R_0(t; \boldsymbol{\eta}) = \exp(-H_0(t; \boldsymbol{\eta}))$ is the baseline reliability function, with $0 < R_0(\tau_I; \boldsymbol{\eta}) < R_0(\tau_{I-1}; \boldsymbol{\eta}) < \cdots < R_0(\tau_1; \boldsymbol{\eta}) < 1$.

Instead of adopting here the parametric approach, wherein we assume a specific functional form for the baseline hazard $H_0(t; \boldsymbol{\eta})$, such as exponential, Weibull or gamma, we may adopt a semi-parametric approach in which we make mild assumptions about the baseline hazard. Specifically, we may subdivide time into the observed intervals and assume that the baseline hazard is constant in each interval, leading to a piece-wise exponential model. We let

$$\gamma(\eta_i) = \begin{cases} 1 - R_0(\tau_I; \boldsymbol{\eta}) = 1 - \exp(-\exp(\eta_I)), & i = I, \\ \dfrac{1 - R_0(\tau_i; \boldsymbol{\eta})}{1 - R_0(\tau_{i+1}; \boldsymbol{\eta})} = 1 - \exp(-\exp(\eta_i)), & i \neq I. \end{cases}$$

We then have

$$\begin{aligned} R_0(\tau_i; \boldsymbol{\eta}) &= 1 - \prod_{m=i}^{I} \{1 - \exp(-\exp(\eta_m))\} = 1 - G_i, \\ H_0(\tau_i; \boldsymbol{\eta}) &= -\log(1 - G_i), \end{aligned}$$

where $G_i = \prod_{m=i}^{I} \{1 - \exp(-\exp(\eta_m))\}$ for $i = 1, \ldots, I$.

Remark 7.1 *We now present a connection between the proportional hazards model and a parametric model with proportional hazard rates. The two-parameter Weibull distribution is commonly used as a lifetime distribution having proportional hazard rates. Suppose the lifetimes of one-shot devices under test follow the Weibull distribution with the same shape parameter* $\lambda = \exp(b)$ *and scale parameters related to the stress levels,* $a_s = \exp(\sum_{j=1}^{J} c_j x_{sj})$, $s = 1, \ldots, S$. *The cumulative distribution function of the Weibull distribution is then given by*

$$F_T(t; a_s, \lambda) = 1 - \exp\left(-\left(\frac{t}{a_s}\right)^{\lambda}\right), \quad t > 0.$$

If the proportional hazards assumption holds, then the baseline reliability and the coefficients of stress factors are given by

$$R_0(t; \lambda) = \exp(-t^{\lambda} \exp(-\lambda c_0))$$

and $\alpha_s = -\lambda c_s, \quad s = 1, \ldots, S$. *Furthermore, we have*

$$\begin{aligned} \eta_i &= \log\left(-\log\left(1 - \frac{1 - R_0(\tau_i)}{1 - R_0(\tau_{i+1})}\right)\right), \\ \eta_I &= \lambda(\log(\tau_I) - c_0). \end{aligned}$$

Consider the proportional hazards model for one-shot devices in (7.1). The log-likelihood function based on these data is then given by

$$
\begin{aligned}
&\ell(n_{11},\ldots,n_{IS};\boldsymbol{\eta},\boldsymbol{\alpha}) \qquad (7.3)\\
&=\sum_{i=1}^{I}\sum_{s=1}^{S} n_{is}\log\left[1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})\right]+(K_{is}-n_{is})\log\left[R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})\right]+C\\
&=\sum_{i=1}^{I}\sum_{s=1}^{S} n_{is}\log\left[1-(1-G_i)^{\exp\left(\sum_{j=1}^{J}\alpha_j x_{sj}\right)}\right]+(K_{is}-n_{is})\log(1-G_i)\exp\left(\sum_{j=1}^{J}\alpha_j x_{sj}\right)+C,
\end{aligned}
$$

where C is a constant not depending on $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$.

Definition 7.2 *Let $\boldsymbol{\theta}=(\boldsymbol{\eta},\boldsymbol{\alpha})$. The MLE, $\widehat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$, is obtained by maximization of (7.3), i.e.,*

$$
\widehat{\boldsymbol{\theta}}=\arg\max_{\boldsymbol{\theta}}\ \ell(n_{11},\ldots,n_{IS};\boldsymbol{\eta},\boldsymbol{\alpha}). \tag{7.4}
$$

In order to study the relation between the MLE, $\widehat{\boldsymbol{\theta}}$, in Definition 7.2, with the Kullback-Leibler divergence measure, we introduce the empirical and theoretical probability vectors, as follows:

$$
\widehat{\boldsymbol{p}}_{is}=(\widehat{p}_{is1},\widehat{p}_{is2})^T=\left(\frac{n_{is}}{K_{is}},\frac{K_{is}-n_{is}}{K_{is}}\right)^T, \tag{7.5}
$$

$$
\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})=\left(\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha}),\pi_{is2}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)^T, \tag{7.6}
$$

where $i=1,\ldots,I,\ s=1,\ldots,S,\ \ \pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})=1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})$ and $\pi_{is2}(\boldsymbol{\eta},\boldsymbol{\alpha})=R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})$.

Definition 7.3 *The Kullback-Leibler divergence measure between $\widehat{\boldsymbol{p}}_{is}$ and $\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})$ is given by*

$$
d_{KL}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}))=\widehat{p}_{is1}\log\left(\frac{\widehat{p}_{is1}}{\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})}\right)+\widehat{p}_{is2}\log\left(\frac{\widehat{p}_{is2}}{\pi_{is2}(\boldsymbol{\eta},\boldsymbol{\alpha})}\right)
$$

and similarly the weighted Kullback-Leibler divergence measure of all the units, where $K=\sum_{s=1}^{S}K_s$ is the total number of devices under the life-test, is given by

$$
\begin{aligned}
&\sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}d_{KL}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}))\\
&=\frac{1}{K}\sum_{i=1}^{I}\sum_{s=1}^{S}K_{is}\left[\widehat{p}_{is1}\log\left(\frac{\widehat{p}_{is1}}{\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})}\right)+\widehat{p}_{is2}\log\left(\frac{\widehat{p}_{is2}}{\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})}\right)\right]\\
&=\frac{1}{K}\sum_{i=1}^{I}\sum_{s=1}^{S}\left[n_{is}\log\left(\frac{\frac{n_{is}}{K_{is}}}{1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})}\right)+(K_{is}-n_{is})\log\left(\frac{\frac{K_{is}-n_{is}}{K_{is}}}{R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})}\right)\right].
\end{aligned}
$$

For more details, one may refer to Pardo [2005]. The relation between the MLE and the estimator obtained by minimizing the weighted Kullback-Leibler divergence measure is obtained on the basis of the following result.

Theorem 7.4 *The log-likelihood function $\ell(n_{11},\ldots,n_{IS};\boldsymbol{\eta},\boldsymbol{\alpha})$, given in (7.3), is related to the weighted Kullback-Leibler divergence measure through*

$$
\sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}d_{KL}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}))=c-\frac{1}{K}\ell(n_{11},\ldots,n_{IS};\boldsymbol{\eta},\boldsymbol{\alpha}),
$$

with c being a constant not dependent on $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$.

Definition 7.5 *The MLE, $\widehat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$, can then be defined as*

$$
\widehat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}\sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}d_{KL}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})). \tag{7.7}
$$

7.3 Weighted minimum DPD estimator

7.3.1 Definition

Given the probability vectors $\widehat{\boldsymbol{p}}_{is}$ and $\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})$ in (7.5) and (7.6), respectively, the DPD between them, as a function of a single tuning parameter $\beta \geq 0$, is given by

$$\begin{aligned} d_{\beta}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) &= \left(\pi_{is1}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha})+\pi_{is2}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)-\frac{\beta+1}{\beta}\left(\widehat{p}_{is1}\pi_{is1}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha})+\widehat{p}_{is2}\pi_{is2}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha})\right) \\ &\quad +\frac{1}{\beta}\left(\widehat{p}_{is1}^{\beta+1}+\widehat{p}_{is2}^{\beta+1}\right), \quad \text{if } \beta>0, \end{aligned} \tag{7.8}$$

and $d_{\beta=0}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = \lim_{\beta\to 0^+} d_{\beta}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = d_{KL}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}))$, for $\beta=0$.

As the term $\frac{1}{\beta}\left(\widehat{p}_{is1}^{\beta+1}+\widehat{p}_{is2}^{\beta+1}\right)$ in (7.8) has no role in the minimization with respect to $\boldsymbol{\theta}$, we can consider the equivalent measure

$$d_{\beta}^{*}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = \left(\pi_{is1}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha})+\pi_{is2}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)-\frac{\beta+1}{\beta}\left(\widehat{p}_{is1}\pi_{is1}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha})+\widehat{p}_{is2}\pi_{is2}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha})\right),$$

and then can redefine the weighted minimum DPD estimator as follows.

Definition 7.6 *The weighted minimum DPD estimator for $\boldsymbol{\theta}$ is given by*

$$\widehat{\boldsymbol{\theta}}_{\beta} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K} d_{\beta}^{*}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})), \quad \text{for } \beta>0,$$

and for $\beta=0$, we have the MLE, $\widehat{\boldsymbol{\theta}}$, as defined in (7.7).

7.3.2 Estimation and asymptotic distribution

The estimating equations for the weighted minimum DPD estimator are as given in the following result.

Theorem 7.7 *For $\beta \geq 0$, the estimating equations are given by*

$$\sum_{i=1}^{I}\sum_{s=1}^{S}\delta_{is}(\boldsymbol{\eta})\left(K_{is}(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))-n_{is}\right)\times\left[(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))^{\beta-1}+R^{\beta-1}(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})\right]=\boldsymbol{0}_I,$$

$$\sum_{i=1}^{I}\sum_{s=1}^{S}\delta_{is}(\boldsymbol{\alpha})\left(K_{is}(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))-n_{is}\right)\times\left[(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))^{\beta-1}+R^{\beta-1}(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})\right]=\boldsymbol{0}_J,$$

where

$$\delta_{is}(\boldsymbol{\eta}) = \frac{\partial R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})}{\partial\boldsymbol{\eta}} = -(1-G_i)^{\lambda(\boldsymbol{x}_s;\boldsymbol{\alpha})-1}\lambda(\boldsymbol{x}_s;\boldsymbol{\alpha})\frac{\partial G_i}{\partial\boldsymbol{\eta}}, \tag{7.9}$$

$$\delta_{is}(\boldsymbol{\alpha}) = \frac{\partial R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}} = (1-G_i)^{\lambda(\boldsymbol{x}_s;\boldsymbol{\alpha})}\log(1-G_i)\lambda(\boldsymbol{x}_s;\boldsymbol{\alpha})\boldsymbol{x}_s, \tag{7.10}$$

with

$$\frac{\partial G_i}{\partial \eta_u} = \begin{cases} \exp(\eta_u)\exp(-\exp(\eta_u))G_i/\gamma(\eta_u) & ,i\leq u, \\ 0 & ,i>u. \end{cases} \tag{7.11}$$

Proof. The estimating equations are given by

$$\frac{\partial}{\partial\boldsymbol{\eta}}\sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}d_{\beta}^{*}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\frac{\partial}{\partial\boldsymbol{\eta}}d_{\beta}^{*}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = \boldsymbol{0}_I,$$

$$\frac{\partial}{\partial\boldsymbol{\alpha}}\sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}d_{\beta}^{*}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\frac{\partial}{\partial\boldsymbol{\alpha}}d_{\beta}^{*}(\widehat{\boldsymbol{p}}_{is},\boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})) = \boldsymbol{0}_J,$$

with

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\eta}} d_\beta^*(\widehat{\boldsymbol{p}}_{is}, \boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}))\\
&= \left(\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is1}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha}) + \frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is2}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right) - \frac{\beta+1}{\beta}\left(\widehat{p}_{is1}\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{is2}\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is2}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)\\
&= (\beta+1)\left(\pi_{is1}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \pi_{is2}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{is1}\pi_{is1}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha}) + \widehat{p}_{is2}\pi_{is2}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})\\
&= (\beta+1)\left((\pi_{i1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{i1})\,\pi_{is1}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - (\pi_{is2}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{is2})\,\pi_{is2}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})\\
&= (\beta+1)\left((\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{i1})\,\pi_{is1}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha}) + (\pi_{i1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{i1})\,\pi_{is2}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})\\
&= (\beta+1)\,(\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{is1})\left(\pi_{is1}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha}) + \pi_{is2}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha}) \qquad (7.12)
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\alpha}} d_\beta^*(\widehat{\boldsymbol{p}}_{is}, \boldsymbol{\pi}_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}))\\
&= (\beta+1)\,(\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \widehat{p}_{is1})\left(\pi_{i1}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha}) + \pi_{is2}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right)\frac{\partial}{\partial \boldsymbol{\alpha}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha}). \qquad (7.13)
\end{aligned}
$$

But, $\frac{\partial}{\partial \boldsymbol{\eta}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})$ and $\frac{\partial}{\partial \boldsymbol{\alpha}}\pi_{is1}(\boldsymbol{\eta},\boldsymbol{\alpha})$ are as given in (7.9) and (7.10), respectively. See equations (25) and (26) of Ling et al. [2015] for details. ■

Theorem 7.8 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter $\boldsymbol{\theta}$. Then, the asymptotic distribution of the weighted minimum DPD estimator, $\widehat{\boldsymbol{\theta}}_\beta$, is given by*

$$
\sqrt{K}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\mathbf{0}_{I+J}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^*)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),
$$

where $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ and $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are given by

$$
\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})\left[(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))^{\beta-1} + R^{\beta-1}(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})\right], \qquad (7.14)
$$

$$
\begin{aligned}
\boldsymbol{K}_\beta(\boldsymbol{\theta}) &= \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))\\
&\quad\times\left[(1-R(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha}))^{\beta-1} + R^{\beta-1}(\tau_i,\boldsymbol{x}_s;\boldsymbol{\eta},\boldsymbol{\alpha})\right]^2, \qquad (7.15)
\end{aligned}
$$

with

$$
\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}) = \begin{pmatrix} \delta_{is}(\boldsymbol{\eta})\delta_{is}^T(\boldsymbol{\eta}) & \delta_{is}(\boldsymbol{\eta})\delta_{is}^T(\boldsymbol{\alpha}) \\ \delta_{is}(\boldsymbol{\alpha})\delta_{is}^T(\boldsymbol{\eta}) & \delta_{is}(\boldsymbol{\alpha})\delta_{is}^T(\boldsymbol{\alpha}) \end{pmatrix},
$$

and $\delta_{is}(\boldsymbol{\eta})$ and $\delta_{is}(\boldsymbol{\alpha})$ are as given in (7.9) and (7.10), respectively.

Proof. We denote

$$
\begin{aligned}
\boldsymbol{u}_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha}) &= \left(\frac{\partial \log \pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}{\partial \boldsymbol{\eta}}, \frac{\partial \log \pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right)^T\\
&= \left(\frac{1}{\pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}\frac{\partial \pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}{\partial \boldsymbol{\eta}}, \frac{1}{\pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}\frac{\partial \pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right)^T\\
&= \left(\frac{(-1)^{j+1}}{\pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}\delta_{is}(\boldsymbol{\eta}), \frac{(-1)^{j+1}}{\pi_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})}\delta_{is}(\boldsymbol{\alpha})\right)^T,
\end{aligned}
$$

with $\delta_{is}(\boldsymbol{\eta})$ and $\delta_{is}(\boldsymbol{\alpha})$ as given in (7.9) and (7.10), respectively.

Now, upon using Result 3.1 of Ghosh et al. [2013], we have

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\mathbf{0}_{I+J}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{j=1}^{2}\frac{K_{is}}{K}\boldsymbol{u}_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})\boldsymbol{u}_{isj}^T(\boldsymbol{\eta},\boldsymbol{\alpha})\pi_{isj}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha}),$$

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{j=1}^{2}\frac{K_{is}}{K}\boldsymbol{u}_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})\boldsymbol{u}_{isj}^T(\boldsymbol{\eta},\boldsymbol{\alpha})\pi_{isj}^{2\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha}) - \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\boldsymbol{\xi}_{is,\beta}(\boldsymbol{\eta},\boldsymbol{\alpha})\boldsymbol{\xi}_{is,\beta}^T(\boldsymbol{\eta},\boldsymbol{\alpha}),$$

with

$$\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\eta},\boldsymbol{\alpha}) = \sum_{j=1}^{2}\boldsymbol{u}_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})\pi_{isj}^{\beta+1}(\boldsymbol{\eta},\boldsymbol{\alpha}) = (\delta_{is}(\boldsymbol{\eta}),\delta_{is}(\boldsymbol{\alpha}))^T\sum_{j=1}^{2}(-1)^{j+1}\pi_{isj}^{\beta}(\boldsymbol{\eta},\boldsymbol{\alpha}).$$

Now, for $\boldsymbol{u}_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})\boldsymbol{u}_{isj}^T(\boldsymbol{\eta},\boldsymbol{\alpha})$, we have

$$\boldsymbol{u}_{isj}(\boldsymbol{\eta},\boldsymbol{\alpha})\boldsymbol{u}_{isj}^T(\boldsymbol{\eta},\boldsymbol{\alpha}) = \frac{1}{\pi_{isj}^2(\boldsymbol{\eta},\boldsymbol{\alpha})}\begin{pmatrix}\delta_{is}(\boldsymbol{\eta})\delta_{is}^T(\boldsymbol{\eta}) & \delta_{is}(\boldsymbol{\eta})\delta_{is}^T(\boldsymbol{\alpha})\\ \delta_{is}(\boldsymbol{\alpha})\delta_{is}^T(\boldsymbol{\eta}) & \delta_{is}(\boldsymbol{\alpha})\delta_{is}^T(\boldsymbol{\alpha})\end{pmatrix} = \frac{1}{\pi_{ij}^2(\boldsymbol{\theta})}\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}),$$

with

$$\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha}) = \begin{pmatrix}\delta_{is}(\boldsymbol{\eta})\delta_{is}^T(\boldsymbol{\eta}) & \delta_{is}(\boldsymbol{\eta})\delta_{is}^T(\boldsymbol{\alpha})\\ \delta_{is}(\boldsymbol{\alpha})\delta_{is}^T(\boldsymbol{\eta}) & \delta_{is}(\boldsymbol{\alpha})\delta_{is}^T(\boldsymbol{\alpha})\end{pmatrix}.$$

It then follows that

$$\begin{aligned}\boldsymbol{J}_\beta(\boldsymbol{\theta}) &= \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})\sum_{j=1}^{2}\pi_{isj}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\\ &= \sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\Delta_{is}(\boldsymbol{\eta},\boldsymbol{\alpha})\left(\pi_{is1}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha}) + \pi_{is2}^{\beta-1}(\boldsymbol{\eta},\boldsymbol{\alpha})\right).\end{aligned}$$

■

From here on, and for simplicity, we will denote $R(\tau_i, \boldsymbol{x}_0; \boldsymbol{\eta}, \boldsymbol{\alpha})$ simply by $R(\tau_i, \boldsymbol{x}_0; \boldsymbol{\theta}))$. Based on Result 7.8, the asymptotic variance of the weighted minimum DPD estimator of the reliability at inspection time τ_i under normal operating condition $\boldsymbol{x}_0$ is given by

$$Var(R(\tau_i, \boldsymbol{x}_0; \widehat{\boldsymbol{\theta}}_\beta)) \equiv Var(R(\widehat{\boldsymbol{\theta}}_\beta)) = \boldsymbol{P}^T\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{P},$$

where

$$\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta) = \boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta), \tag{7.16}$$

$\boldsymbol{J}_\beta(\boldsymbol{\theta})$, $\boldsymbol{K}_\beta(\boldsymbol{\theta})$ are as given in (7.14) and (7.15), respectively, and $\boldsymbol{P}$ is a vector of the first-order derivatives of $R(\tau_i, \boldsymbol{x}_0; \boldsymbol{\theta}))$ with respect to the model parameters (see (7.9) and (7.10)). Consequently, the $100(1-\alpha)\%$ asymptotic confidence interval for the reliability function $R(\boldsymbol{\theta})$ is given by

$$\left(R(\widehat{\boldsymbol{\theta}}_\beta) - z_{1-\alpha/2}se(R(\widehat{\boldsymbol{\theta}}_\beta)),\ R(\widehat{\boldsymbol{\theta}}_\beta) + z_{1-\alpha/2}se(R(\widehat{\boldsymbol{\theta}}_\beta))\right),$$

where $se(R(\widehat{\boldsymbol{\theta}}_\beta)) = \sqrt{\widehat{Var(R(\widehat{\boldsymbol{\theta}}_\beta))}}$ and z_γ is the uppper γ percentage point of the standard normal distribution.

However, an asymptotic confidence interval may be satisfactory only for large sample sizes as it is based on the asymptotic properties of the estimators. Balakrishnan and Ling [2013] found that, in the case of small sample sizes, the distribution of the MLE of reliability is quite skewed, and so proposed a logit-transformation for obtaining a confidence interval for the reliability function, which can be extended to the case of the weighted minimum DPD estimators of the reliabilities as well to obtain a confidence interval of the form

$$\left(\frac{R(\widehat{\boldsymbol{\theta}}_\beta)}{R(\widehat{\boldsymbol{\theta}}_\beta)+(1-R(\widehat{\boldsymbol{\theta}}_\beta))T}, \frac{R(\widehat{\boldsymbol{\theta}}_\beta)}{R(\widehat{\boldsymbol{\theta}}_\beta)+(1-R(\widehat{\boldsymbol{\theta}}_\beta))/T}\right), \tag{7.17}$$

where $T = \exp\left(z_{1-\alpha/2}\frac{se(R(\widehat{\boldsymbol{\theta}}_\beta))}{R(\widehat{\boldsymbol{\theta}}_\beta)(1-R(\widehat{\boldsymbol{\theta}}_\beta))}\right)$.

7.3.3 Study of the Influence Function

Theorem 7.9 *Let us consider the one-shot device testing under proportional hazards model defined in (7.2) and let us define the statistical functional $\boldsymbol{U}_\beta(\cdot)$ corresponding to the weighted minimum DPD estimator as the minimizer of the weighted sum of DPDs between the true and model densities. The IF with respect to the $k-th$ observation of the i_0s_0-th group is given by*

$$\begin{aligned}IF(t_{i_0s_0,k},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^*}) =& \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\frac{K_{i_0s_0}}{K}\left(F^{\beta-1}(\tau_{i_0};\boldsymbol{x}_{s_0},\boldsymbol{\theta}^*)+R^{\beta-1}(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta}^*)\right)\\ &\times(\delta_{i_0s_0}^T(\boldsymbol{\eta}^*),\delta_{i_0s_0}^T(\boldsymbol{\alpha}^*))^T\left(F(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta}^*)-\Delta^{(1)}_{t_{i_0s_0,k}}\right),\end{aligned} \tag{7.18}$$

where $\Delta^{(1)}_{t_{i_0s_0,k}}$ is the degenerating function at point $t_{i_0s_0,k}$.

The IF with respect to all the observations is given by

$$\begin{aligned}IF(\boldsymbol{t},\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}^*}) =& \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^*)\sum_{i=1}^{I}\sum_{s=1}^{S}\frac{K_{is}}{K}\left[\left(F^{\beta-1}(\tau_i,\boldsymbol{x}_s;\boldsymbol{\theta}^*)+R^{\beta-1}(\tau_i,\boldsymbol{x}_s;\boldsymbol{\theta}^*)\right)\right.\\ &\left.\times(\delta_{is}^T(\boldsymbol{\eta}^*),\delta_{is}^T(\boldsymbol{\alpha}^*))^T\left(F(\tau_i,\boldsymbol{x}_s;\boldsymbol{\theta}^*)-\Delta^{(1)}_{t_{is}}\right)\right],\end{aligned} \tag{7.19}$$

where $\Delta^{(1)}_{t_is_i}=\sum_{k=1}^{K_i}\Delta^{(1)}_{t_{is,k}}$.

Derivations of equations (7.18) and (7.19) require some heavy computations that are quite similar to those developed in Chapter 2.

Remark 7.10 *Let*

$$\begin{aligned}h_{1,i}(\tau_{i_0},\boldsymbol{x}_{s_0},\boldsymbol{\theta}) &= \frac{1}{R_0(\tau_{i_0},\boldsymbol{\eta})}\frac{\partial G_{i_0}}{\partial \eta_i}\lambda(\boldsymbol{x}_{s_0};\boldsymbol{\alpha})\left[R(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta})(1-R(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta}))^{\beta-1}+R^\beta(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta})\right]\\ h_{2,j}(\tau_{i_0},\boldsymbol{x}_{s_0},\boldsymbol{\theta}) &= \log(R_0(\tau_{i_0},\boldsymbol{\eta}))\lambda(\boldsymbol{x}_{s_0};\boldsymbol{\alpha})\boldsymbol{x}_{s_0j}\left[R(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta})(1-R(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta}))^{\beta-1}+R^\beta(\tau_{i_0},\boldsymbol{x}_{s_0};\boldsymbol{\theta})\right]\end{aligned}$$

be the factors of the influence function of $\boldsymbol{\theta}$ given in (7.18) and (7.19). Based on this, it may be mentioned that conditions for boundedness of the influence functions presented in this paper, either with respect to an observation or with respect to all the observations, are bounded on $t_{i_0s_0,k}$ or $\boldsymbol{t}$, but if $\beta = 0$ the norm of the IFs can be very large, in comparison to $\beta > 0$, since it can be deduced that

$$\lim_{x_{s_0j}\to+\infty} h_{1,i}(\tau_{i_0},\boldsymbol{x}_{s_0},\boldsymbol{\theta}) = \lim_{x_{s_0j}\to+\infty} h_{2,j}(\tau_{i_0},\boldsymbol{x}_{s_0},\boldsymbol{\theta}) = \left\{\begin{array}{ll} =\infty, & \text{if } \beta=0\\ <\infty, & \text{if } \beta>0\end{array}\right. . \tag{7.20}$$

This implies that the proposed weighted minimum DPD estimators with $\beta > 0$ are robust against leverage points, but the classical MLE is clearly non-robust.

7.4 Wald-type tests

Let us consider the function $\boldsymbol{m} : \mathbb{R}^{I+J} \longrightarrow \mathbb{R}^r$, where $r \leq (I+J)$ and

$$\boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{0}_r, \tag{7.21}$$

which corresponds to a composite null hypothesis. We assume that the $(I+J)\times r$ matrix $\boldsymbol{M}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{m}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exists and is continuous in $\boldsymbol{\theta}$ and rank $\boldsymbol{M}(\boldsymbol{\theta}) = r$. Then, for testing

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \text{ against } H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0, \tag{7.22}$$

where $\boldsymbol{\Theta}_0 = \left\{\boldsymbol{\theta} \in \mathbb{R}^{(I+J)} : \boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{0}_r\right\}$, we can consider the following Wald-type test statistics:

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) = K\boldsymbol{m}^T(\widehat{\boldsymbol{\theta}}_\beta)\left(\boldsymbol{M}^T(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{M}(\widehat{\boldsymbol{\theta}}_\beta)\right)^{-1}\boldsymbol{m}(\widehat{\boldsymbol{\theta}}_\beta), \tag{7.23}$$

where $\boldsymbol{\Sigma}_\beta(\widehat{\boldsymbol{\theta}}_\beta)$ is as given in (7.16).

Theorem 7.11 *Under (7.21), we have*

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \chi^2_r,$$

where χ^2_r *denotes a central chi-square distribution with* r *degrees of freedom.*

Proof. Let $\boldsymbol{\theta}^0 \in \Theta_0$ be the true value of the parameter $\boldsymbol{\theta}$. It is clear that

$$\begin{aligned}\boldsymbol{m}\left(\widehat{\boldsymbol{\theta}}_\beta\right) &= \boldsymbol{m}\left(\boldsymbol{\theta}^0\right) + \boldsymbol{M}^T\left(\widehat{\boldsymbol{\theta}}_\beta\right)\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) + o_p\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right\|\right)\\ &= \boldsymbol{M}^T\left(\widehat{\boldsymbol{\theta}}_\beta\right)\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) + o_p\left(K^{-1/2}\right).\end{aligned}$$

But, under H_0, $\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\mathbf{0}_{(I+J)}, \boldsymbol{\Sigma}_\beta\left(\boldsymbol{\theta}^0\right)\right)$. Therefore, under H_0,

$$\sqrt{K}\boldsymbol{m}\left(\widehat{\boldsymbol{\theta}}_\beta\right) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{M}^T\left(\boldsymbol{\theta}^0\right)\boldsymbol{\Sigma}_\beta\left(\boldsymbol{\theta}^0\right)\boldsymbol{M}\left(\boldsymbol{\theta}^0\right)\right)$$

and taking into account that $rank(\boldsymbol{M}\left(\boldsymbol{\theta}^0\right)) = r$, we obtain

$$K\boldsymbol{m}^T\left(\widehat{\boldsymbol{\theta}}_\beta\right)\left(\boldsymbol{M}^T\left(\boldsymbol{\theta}^0\right)\boldsymbol{\Sigma}_\beta\left(\boldsymbol{\theta}^0\right)\boldsymbol{M}\left(\boldsymbol{\theta}^0\right)\right)^{-1}\boldsymbol{m}\left(\widehat{\boldsymbol{\theta}}_\beta\right) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \chi^2_r.$$

Because $\left(\boldsymbol{M}^T\left(\widehat{\boldsymbol{\theta}}_\beta\right)\boldsymbol{\Sigma}_\beta\left(\widehat{\boldsymbol{\theta}}_\beta\right)\boldsymbol{M}\left(\widehat{\boldsymbol{\theta}}_\beta\right)\right)^{-1}$ is a consistent estimator of $\left(\boldsymbol{M}^T\left(\boldsymbol{\theta}^0\right)\boldsymbol{\Sigma}_\beta\left(\boldsymbol{\theta}^0\right)\boldsymbol{M}\left(\boldsymbol{\theta}^0\right)\right)^{-1}$, we get

$$W_K\left(\widehat{\boldsymbol{\theta}}_\beta\right) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \chi^2_r.$$

■

Based on Theorem 7.11, we shall reject the null hypothesis in (7.22) if

$$W_K(\widehat{\boldsymbol{\theta}}_\beta) > \chi^2_{r,\alpha}, \tag{7.24}$$

where $\chi^2_{r,\alpha}$ is the upper α percentage point of χ^2_r distribution.

7.5 Simulation Study

In this section, an extensive simulation study is carried out for evaluating the proposed weighted minimum DPD estimators and Wald-type tests. The simulations results are computed based on 1,000 simulated samples in the R statistical software. Mean square error (MSE) and bias are computed for evaluating the estimators in both balanced and unbalanced data sets, while empirical levels and powers are computed for evaluating the tests.

7.5.1 The weighted minimum DPD estimators

Suppose the lifetimes of test units follow a Weibull distribution (see Remark 7.1). All the test units were divided into $S = 4$ groups, subject to different acceleration conditions with $J = 2$ stress factors at two elevated stress levels each, that is, $(x_1, x_2) = \{(55, 70), (55, 100), (85, 70), (85, 100)\}$, and were inspected at $I = 3$ different times,
$(\tau_1, \tau_2, \tau_3) = (2, 5, 8)$.

Balanced data

We assume $(c_1, c_2) = (-0.03, -0.03)$, $c_0 \in \{6, 6.5\}$ for different degrees of reliability and $b \in \{0, 0.5\}$. Note that the exponential distribution will be included as a special case when we take $b = 0$. In this framework, we consider "outlying cells" rather than "outlying observations". A cell which does not follow the one-shot device model will be called an outlying cell or outlier. In this cell, the number of devices failed will be different than what is expected. This is in the spirit of principle of inflated models in distribution theory. This outlying cell (taken to be $i = 3$, $s = 4$), is generated under the parameters $(\tilde{c}_1, \tilde{c}_2) = (-0.027, -0.027)$ and $\tilde{b} \in \{0.05, 0.45\}$.

Bias of estimates are then computed for different (equal) samples sizes $K_{is} \in \{50, 70, 100\}$ and tuning parameters $\beta \in \{0, 0.2, 0.4, 0.6\}$ for both pure and contaminated data. The obtained results are presented in Tables 7.5.1, 7.5.2, 7.5.3 and 7.5.4. As expected, when the sample size increases, errors tend to decrease, while in the contaminated data set, these errors are generally greater than in the case of uncontaminated data. Weighted minimum DPD estimators with $\beta > 0$ present a better behaviour than the MLE in terms of robustness. Note that reliabilities are underestimated and that the estimates are quite precise in all the cases.

Unbalanced data

In this setting, we consider an unbalanced data set, in which at each inspection time i,
$(K_{i1}, K_{i2}, K_{i3}, K_{i4}) = (10r, 15r, 20r, 30r)$ for different values of the factor $r \in \{1, 2, \ldots, 10\}$. We then assume $(c_0, c_1, c_2) = (6, 5, -0.03, -0.03)$, $b = 0.5$, and $\tilde{c}_2 = -0.027$. MSEs of the parameter $\boldsymbol{\theta}$ are then computed and the obtained results are presented in Figure 7.5.1.

As expected, when the sample size increases, the MSE decreases, but lack of robustness of the MLE ($\beta = 0$) as compared to the weighted minimum DPD estimators with $\beta > 0$ becomes quite evident.

7.5.2 Confidence Intervals

We now study the performance of the proposed methods for the estimation of reliabilities and their confidence intervals. Let us consider the scenario of balanced data with $(c_0, b) = (6, 0.5)$ described previously. We estimate the bias for the reliability at the inspection times under the normal operating conditions $\boldsymbol{x}_0 = (25, 35)$ for different values of the tuning parameter $\beta \in \{0, 0.2, 0.4, 0.6\}$. Coverage Probabilities (CP) and Average Widths (AW), both in their basic form and based on the logit-transformation, are also computed and presented in Table 7.5.5, Table 7.5.6 and Table 7.5.7 for $K_{is} \in \{50, 70, 100\}$, respectively.

It is clear that each estimate tends to the true value accurately, and the coverage probability is close to the nominal level with a larger sample size resulting in a smaller width. The tuning parameter is not very significant when an uncontaminated data-set is considered, while in case of contaminated data, estimates and confidence intervals based on MLE are improved by those based on $\beta > 0$. Confidence intervals obtained through the logit transformation are generally more satisfactory.

Table 7.5.1: Proportional hazards model: Bias for the semi-parametric model with $b = 0$ and $c_0 = 6$.

		Pure data				Contaminated data			
$K_{is} = 50$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-0.66688	-0.00494	-0.00276	-0.00053	-0.00372	0.09898	0.06722	0.03547	0.01708
η_2	-0.01304	-0.00228	-0.00078	0.00109	-0.00131	0.06902	0.04716	0.02531	0.01286
η_3	-3.92056	-0.02788	-0.01810	-0.01389	-0.01982	0.34916	0.23252	0.12087	0.05402
α_1	0.03000	0.00010	0.00002	-0.00001	0.00002	-0.00281	-0.00193	-0.00107	-0.00056
α_2	0.03000	0.00033	0.00027	0.00025	0.00030	-0.00259	-0.00167	-0.00081	-0.00028
		Pure data				Contaminated data			
$K_{is} = 70$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-0.66688	-0.00780	-0.00675	-0.00716	-0.00802	0.09876	0.06233	0.03209	0.01278
η_2	-0.01304	-0.00459	-0.00386	-0.00410	-0.00465	0.06810	0.04315	0.02257	0.00948
η_3	-3.92056	-0.03954	-0.03763	-0.04073	-0.04458	0.35084	0.21624	0.10254	0.03023
α_1	0.03000	0.00027	0.00025	0.00026	0.00029	-0.00276	-0.00173	-0.00086	-0.00030
α_2	0.03000	0.00035	0.00034	0.00038	0.00041	-0.00267	-0.00163	-0.00075	-0.00017
		Pure data				Contaminated data			
$K_{is} = 100$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-0.66688	-0.00778	-0.00682	-0.00711	-0.00785	0.09857	0.06207	0.03231	0.01320
η_2	-0.01304	-0.00477	-0.00412	-0.00429	-0.00477	0.06776	0.04275	0.02248	0.00952
η_3	-3.92056	-0.02739	-0.02332	-0.02387	-0.02586	0.36315	0.23019	0.12013	0.04993
α_1	0.03000	0.00031	0.00028	0.00029	0.00031	-0.00271	-0.00169	-0.00084	-0.00029
α_2	0.03000	0.00016	0.00013	0.00013	0.00015	-0.00287	-0.00185	-0.00099	-0.00045

Table 7.5.2: Proportional hazards model: Bias for the semi-parametric model with $b = 0.5$ and $c_0 = 6$.

		Pure data				Contaminated data			
$K_{is} = 50$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-1.38827	-0.01224	-0.00988	-0.03700	-0.07648	0.03590	-0.00540	-0.03356	-0.08611
η_2	-0.48138	-0.00687	-0.00537	-0.02153	-0.04394	0.02362	-0.00239	-0.01962	-0.04962
η_3	-6.46391	-0.05973	-0.05148	-0.19693	-0.40632	0.14538	-0.03531	-0.17304	-0.47276
α_1	0.04946	0.00032	0.00023	0.00124	0.00274	-0.00125	0.00010	0.00106	0.00325
α_2	0.04946	0.00061	0.00056	0.00174	0.00361	-0.00097	0.00043	0.00155	0.00410
		Pure data				Contaminated data			
$K_{is} = 70$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-1.38827	-0.02188	-0.01955	-0.05972	-0.10800	0.03391	-0.01254	-0.06567	-0.12770
η_2	-0.48138	-0.01318	-0.01171	-0.03481	-0.06298	0.02199	-0.00731	-0.03817	-0.07423
η_3	-6.46391	-0.06923	-0.06287	-0.27869	-0.55900	0.16868	-0.03493	-0.30595	-0.67033
α_1	0.04946	0.00062	0.00055	0.00202	0.00446	-0.00121	0.00033	0.00217	0.00530
α_2	0.04946	0.00054	0.00051	0.00235	0.00433	-0.00128	0.00029	0.00260	0.00520
		Pure data				Contaminated data			
$K_{is} = 100$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-1.38827	-0.01771	-0.01652	-0.06256	-0.08467	0.04334	-0.01518	-0.06228	-0.08025
η_2	-0.48138	-0.01071	-0.00996	-0.03610	-0.04904	0.02774	-0.00875	-0.03612	-0.04659
η_3	-6.46391	-0.05209	-0.04718	-0.27304	-0.41072	0.21133	-0.04194	-0.27462	-0.38078
α_1	0.04946	0.00057	0.00053	0.00204	0.00340	-0.00145	0.00048	0.00226	0.00316
α_2	0.04946	0.00034	0.00030	0.00217	0.00315	-0.00168	0.00026	0.00204	0.00295

Table 7.5.3: Proportional hazards model: Bias for the semi-parametric model with $b = 0.5$ and $c_0 = 6.5$.

		Pure data				Contaminated data			
$K_{is} = 50$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-0.66879	0.00223	0.00097	0.00020	-0.00900	0.17046	0.14959	0.12180	0.11436
η_2	-0.01553	0.00320	0.00227	0.00156	-0.00465	0.11845	0.10401	0.08415	0.08029
η_3	-4.42056	-0.01196	-0.00948	-0.01543	-0.03474	0.56481	0.50586	0.43705	0.36957
α_1	0.03000	0.00004	0.00001	0.00004	0.00021	-0.00437	-0.00389	-0.00336	-0.00286
α_2	0.03000	0.00014	0.00013	0.00018	0.00030	-0.00426	-0.00379	-0.00324	-0.00274
		Pure data				Contaminated data			
$K_{is} = 70$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-0.66879	0.00050	-0.00537	-0.00598	-0.00468	0.16516	0.14230	0.12082	0.10748
η_2	-0.01553	0.00166	-0.00294	-0.00333	-0.00217	0.11371	0.09772	0.08279	0.07497
η_3	-4.42056	-0.03260	-0.03492	-0.03697	-0.04026	0.55639	0.49446	0.42414	0.36377
α_1	0.03000	0.00016	0.00017	0.00019	0.00021	-0.00434	-0.00386	-0.00329	-0.00322
α_2	0.03000	0.00029	0.00032	0.00034	0.00036	-0.00418	-0.00367	-0.00314	-0.00318
		Pure data				Contaminated data			
$K_{is} = 100$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-0.66879	-0.00302	-0.00428	-0.00423	-0.00453	0.15788	0.14182	0.12319	0.10443
η_2	-0.01553	-0.00136	-0.00242	-0.00237	-0.00256	0.10788	0.09716	0.08443	0.07154
η_3	-4.42056	-0.02671	-0.02485	-0.02528	-0.02720	0.55329	0.49534	0.43019	0.36647
α_1	0.03000	0.00019	0.00018	0.00019	0.00020	-0.00422	-0.00376	-0.00325	-0.00276
α_2	0.03000	0.00014	0.00014	0.00014	0.00015	-0.00427	-0.00381	-0.00331	-0.00281

Table 7.5.4: Proportional hazards model: Bias for the semi-parametric model with $b = 0.5$ and $c_0 = 6.5$.

		Pure data				Contaminated data			
$K_{is} = 50$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-1.38845	-0.00292	-0.02634	-0.06961	-0.10441	0.28565	0.19158	0.13289	0.07420
η_2	-0.48171	-0.00007	-0.01454	-0.03906	-0.08819	0.18184	0.12322	0.12394	0.12467
η_3	-7.28827	-0.08460	-0.15057	-0.34018	-0.97897	1.21433	0.81382	0.14850	0.33555
α_1	0.04946	0.00058	0.00102	0.00237	-0.08206	-0.00889	-0.00592	-0.00116	-0.32679
α_2	0.04946	0.00057	0.00108	0.00246	-0.10152	-0.00891	-0.00594	-0.00112	-0.39906
		Pure data				Contaminated data			
$K_{is} = 70$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-1.38845	-0.01510	-0.03564	-0.05629	-0.08443	0.28682	0.18935	0.03294	0.02196
η_2	-0.48171	-0.00878	-0.02069	-0.03247	-0.11914	0.18302	0.12036	0.02563	0.04542
η_3	-7.28827	-0.06124	-0.15643	-0.22439	-0.99978	1.24650	0.87564	0.14490	-0.01407
α_1	0.04946	0.00041	0.00110	0.00155	-0.01596	-0.00916	-0.00637	-0.00104	-0.15771
α_2	0.04946	0.00047	0.00111	0.00168	-0.02075	-0.00910	-0.00631	-0.00113	-0.19334
		Pure data				Contaminated data			
$K_{is} = 100$	True value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
η_1	-1.38845	-0.00904	-0.01105	-0.05924	-0.23063	0.28616	0.19928	0.05644	0.03762
η_2	-0.48171	-0.00531	-0.00635	-0.03368	-0.12308	0.18172	0.12619	0.03911	-0.01015
η_3	-7.28827	-0.06888	-0.07584	-0.29436	-0.96654	1.22401	0.87425	0.21922	-0.38512
α_1	0.04946	0.00048	0.00053	0.00202	-0.00879	-0.00897	-0.00635	-0.00168	-0.09403
α_2	0.04946	0.00041	0.00047	0.00207	-0.01198	-0.00904	-0.00642	-0.00164	-0.11596

Table 7.5.5: Proportional hazards model: Bias, Coverage Probabilities (CP) and the Average Widths (AW) of 95% confidence intervals for the reliability with $b = 0.5$, $c_0 = 6$ and $K_{is} = 50$

	Uncontaminated data				Contaminated data			
$R(2, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00028	0.00032	0.00035	0.00037	0.00116	0.00043	0.00036	0.00038
CP	0.90400	0.90500	0.89718	0.89898	0.90800	0.90891	0.89627	0.89867
AW	0.00646	0.00666	0.00691	0.00716	0.00780	0.00684	0.00693	0.00718
CP_{logit}	0.93800	0.93500	0.93347	0.93571	0.85100	0.92993	0.93152	0.93552
AW_{logit}	0.00749	0.00776	0.00814	0.00852	0.00895	0.00796	0.00815	0.00853
	Uncontaminated data				Contaminated data			
$R(5, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00081	0.00095	0.00103	0.00107	0.00343	0.00129	0.00105	0.00110
CP	0.92100	0.91700	0.91734	0.91224	0.92100	0.91592	0.91541	0.91198
AW	0.02385	0.02436	0.02508	0.02585	0.02725	0.02483	0.02511	0.02589
CP_{logit}	0.93900	0.93600	0.93649	0.93980	0.87900	0.93293	0.93555	0.93961
AW_{logit}	0.02638	0.02700	0.02794	0.02898	0.02993	0.02750	0.02797	0.02901
	Uncontaminated data				Contaminated data			
$R(8, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00147	0.00169	0.00183	0.00191	0.00585	0.00228	0.00186	0.00196
CP	0.92600	0.92200	0.92137	0.91939	0.92300	0.91992	0.91944	0.91914
AW	0.04710	0.04814	0.04930	0.05017	0.05207	0.04886	0.04934	0.05023
CP_{logit}	0.93900	0.93800	0.93952	0.94490	0.88400	0.93894	0.93756	0.94473
AW_{logit}	0.05098	0.05222	0.05362	0.05469	0.05601	0.05295	0.05366	0.05475

Table 7.5.6: Proportional hazards model: Bias, Coverage Probabilities (CP) and the Average Widths (AW) of 95% confidence intervals for the reliability with $b = 0.5$, $c_0 = 6$ and $K_{is} = 70$

	Uncontaminated data				Contaminated data			
$R(2, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00016	0.00019	0.00022	0.00022	0.00108	0.00029	0.00023	0.00022
CP	0.89800	0.90090	0.90389	0.90071	0.90800	0.90891	0.90276	0.90051
AW	0.00531	0.00546	0.00568	0.00586	0.00654	0.00560	0.00570	0.00585
CP_{logit}	0.95000	0.94795	0.94785	0.94630	0.84500	0.93894	0.94882	0.94619
AW_{logit}	0.00593	0.00612	0.00640	0.00665	0.00724	0.00627	0.00642	0.00664
	Uncontaminated data				Contaminated data			
$R(5, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00046	0.00057	0.00070	0.00066	0.00333	0.00088	0.00073	0.00065
CP	0.91500	0.91992	0.92638	0.92097	0.91200	0.92392	0.92426	0.91980
AW	0.01981	0.02021	0.02087	0.02139	0.02302	0.02057	0.02090	0.02137
CP_{logit}	0.94600	0.94695	0.95399	0.95339	0.86600	0.93794	0.95496	0.95330
AW_{logit}	0.02132	0.02178	0.02258	0.02325	0.02465	0.02217	0.02261	0.02323
	Uncontaminated data				Contaminated data			
$R(8, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00086	0.00106	0.00133	0.00122	0.00576	0.00160	0.00138	0.00119
CP	0.92400	0.93093	0.93047	0.92908	0.91800	0.93293	0.92835	0.92792
AW	0.03933	0.04018	0.04129	0.04175	0.04415	0.04076	0.04135	0.04172
CP_{logit}	0.95400	0.95295	0.95297	0.94934	0.88900	0.94695	0.95292	0.94924
AW_{logit}	0.04169	0.04266	0.04392	0.04448	0.04657	0.04324	0.04398	0.04445

Table 7.5.7: Proportional hazards model: Bias, Coverage Probabilities (CP) and the Average Widths (AW) of 95% confidence intervals for the reliability with $b = 0.5$, $c_0 = 6$ and $K_{is} = 100$

	Uncontaminated data				Contaminated data			
$R(2, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00008	0.00009	0.00010	0.00010	0.00105	0.00019	0.00011	0.00010
CP	0.93100	0.93594	0.92995	0.93394	0.91200	0.93800	0.93219	0.93516
AW	0.00437	0.00447	0.00461	0.00475	0.00546	0.00458	0.00462	0.00475
CP_{logit}	0.96200	0.95996	0.95939	0.96037	0.80900	0.95600	0.96053	0.96150
AW_{logit}	0.00472	0.00484	0.00502	0.00520	0.00587	0.00496	0.00503	0.00520
	Uncontaminated data				Contaminated data			
$R(5, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00022	0.00026	0.00029	0.00025	0.00324	0.00056	0.00031	0.00025
CP	0.93300	0.93193	0.93604	0.93496	0.91200	0.93700	0.93725	0.93617
AW	0.01636	0.01661	0.01704	0.01747	0.01923	0.01692	0.01706	0.01747
CP_{logit}	0.95300	0.95996	0.96041	0.96240	0.83900	0.95700	0.95951	0.96454
AW_{logit}	0.01724	0.01753	0.01802	0.01854	0.02018	0.01784	0.01805	0.01854
	Uncontaminated data				Contaminated data			
$R(8, \boldsymbol{x}_0)$	0	0.2	0.4	0.6	0	0.2	0.4	0.6
Bias	0.00041	0.00047	0.00053	0.00047	0.00559	0.00099	0.00057	0.00046
CP	0.94200	0.93994	0.94010	0.93598	0.92100	0.94200	0.94130	0.93718
AW	0.03259	0.03318	0.03387	0.03428	0.03691	0.03365	0.03391	0.03428
CP_{logit}	0.95000	0.95696	0.96142	0.95833	0.86000	0.95400	0.96255	0.96049
AW_{logit}	0.03397	0.03463	0.03541	0.03588	0.03834	0.03511	0.03545	0.03588

7.5.3 Wald-type tests

To evaluate the performance of the proposed Wald-type tests, we consider the scenario of unbalanced data proposed discussed above. We consider the testing problem

$$H_0 : \alpha_1 = 0.04946 \quad \text{against} \quad H_1 : \alpha_1 \neq 0.04946, \tag{7.25}$$

Under the same simulation scheme as used above in Section 7.5.1, we first evaluate the empirical levels, measured as the proportion of Wald-type test statistics exceeding the corresponding chi-square critical value for a nominal size of 0.05. The empirical powers are computed in a similar manner, with $\alpha_1^0 = 0.05276$ ($c_1 = -0.032$, $c_2 = -0.028$). The obtained results are shown in Figure 7.5.1. In the case of uncontaminated data, the conventional Wald test has level to be close to nominal value and also has good power performance. The robust tests, however, has a slightly inflated level values (as compared to the nominal value), but possesses similar power as the conventional Wald test (which is evident from the Figure 7.5.1). But, when the data is contaminated, the level of the conventional Wald test turns out to be quite non-robust and takes on very high values as compared to the nominal level. This, in turn, results in higher power (see Figure 7.5.1). However, the proposed robust tests maintain levels close to the nominal value and also possesses good power values (as can be seen in the Figure 7.5.1). Thus, taking both level and power into account, the robust tests, though is slightly inferior to the conventional Wald test in the case of uncontaminated data, turn out to be considerably more efficient than the conventional Walk test in the case of contaminated data

7.6 Application to Real Data

7.6.1 Testing on proportional Hazard rates

Based on Balakrishnan and Ling [2012b], we suggest a distance-based statistic on the form

$$M_\beta = max_{i,s} \left| n_{is} - K_{is}(1 - R(IT_i, \boldsymbol{x}_s; \widehat{\boldsymbol{\theta}}_\beta)) \right| \tag{7.26}$$

as a discrepancy measure for evaluating the fit of the assumed model to the observed data. If the assumed model is not a good fit to the data, we will obtain a large value of M_β. In fact, under the assumed model, we have

$$n_{is} \sim \text{Binomial}(K_{is}, 1 - R(IT_i, \boldsymbol{x}_s; \boldsymbol{\theta})),$$

and so, by denoting $\Phi_{is} = \lceil K_{is}(1-R(IT_i, \boldsymbol{x}_s; \widehat{\boldsymbol{\theta}}_\beta)) - M_\beta \rceil$ and $\Psi_{is} = \lfloor K_{is}(1-R(IT_i, \boldsymbol{x}_s; \widehat{\boldsymbol{\theta}}_\beta)) + M_\beta \rfloor$, the corresponding exact p-value is given by

$$\begin{aligned}
p - value &= Pr\left(max_{i,s} \left| n_{is} - K_{is}(1 - R(IT_i, \boldsymbol{x}_s; \widehat{\boldsymbol{\theta}}_\beta)) \right| > M_\beta \right) \\
&= 1 - Pr\left(max_{i,s} \left| n_{is} - K_{is}(1 - R(IT_i, \boldsymbol{x}_s; \widehat{\boldsymbol{\theta}}_\beta)) \right| \leq M_\beta \right) \\
&= 1 - \prod_{i=1}^{I} \prod_{s=1}^{S} Pr\left(\left| n_{is} - K_{is}(1 - R(IT_i, \boldsymbol{x}_s; \widehat{\boldsymbol{\theta}}_\beta)) \right| \leq M_\beta \right) \\
&= 1 - \prod_{i=1}^{I} \prod_{s=1}^{S} Pr\left(\Phi_{is} \leq n_{is} \leq \Psi_{is} \right).
\end{aligned} \tag{7.27}$$

From (7.27), we can readily validate the proportional hazards assumption if the p-value is sufficiently large.

7.6.2 Choice of the tuning parameter

In the preceding discussion, we have seen how weighted minimum DPD estimators with $\beta > 0$ tend to be more robust than the classical MLE overall when contamination is present in the data.

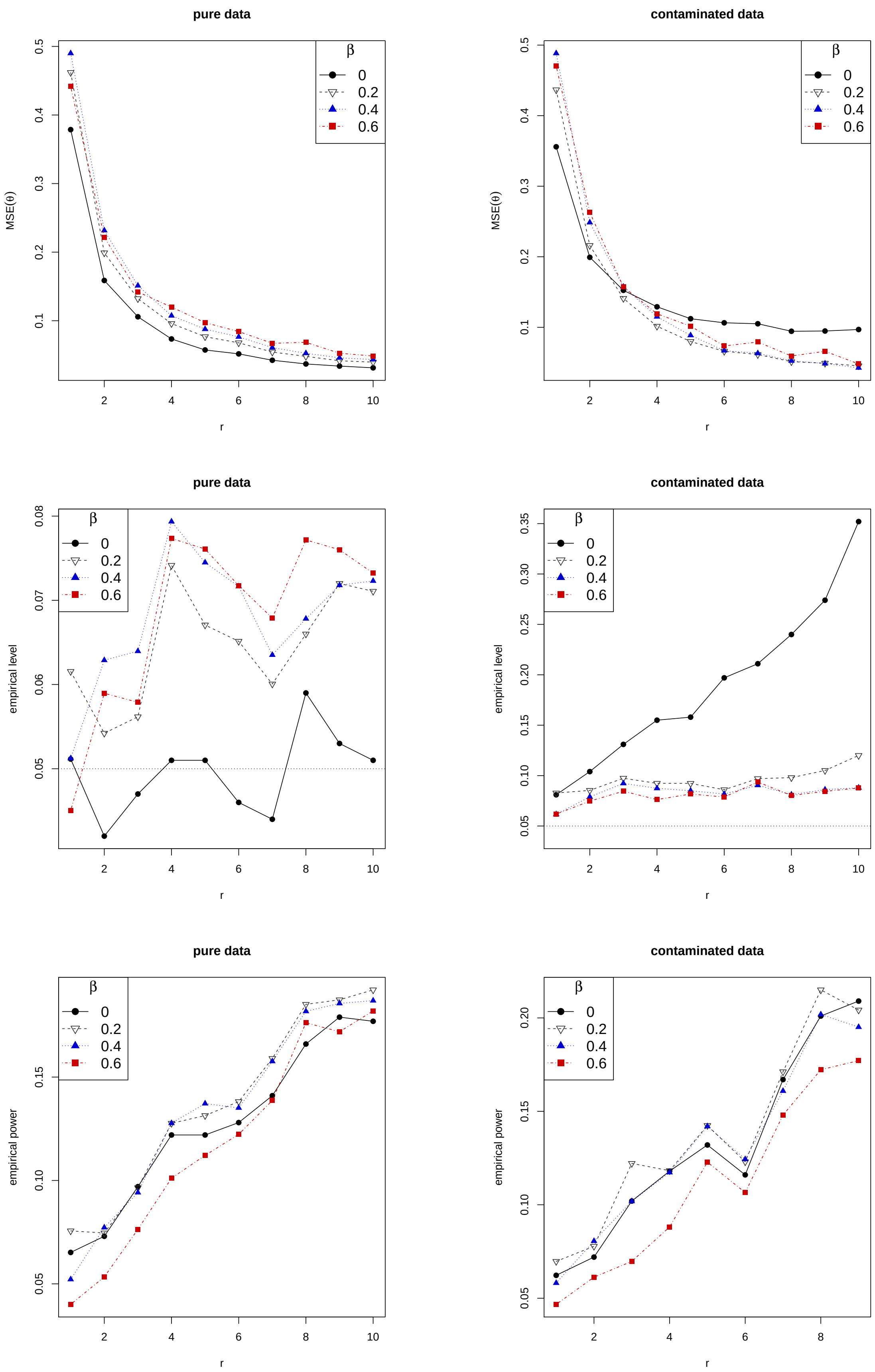

Figure 7.5.1: Proportional hazards model: MSEs and estimated levels and powers for unbalanced data

MLE has been shown to be more efficient when there is no contamination in the data. It is then necessary to provide a data-driven procedure for the determination of the optimal choice of the tuning parameter that would provide a trade-off between efficiency and robustness. One way to do this is as follows: In a grid of possible tuning parameters, apply a measure of discrepancy to the data. Then, the tuning parameter that leads to the minimum discrepancy-statistic can be chosen as the "optimal" one.

A possible choice of the discrepancy measure could be M_β, given in (7.26). Another idea may be by minimizing the estimated mean square error, as suggested in Warwick and Jones [2005] and in previous chapters. The need for a pilot estimator became the major drawback of this procedure, as will be seen in the next section.

7.6.3 Electric Current data

We now consider the Electric Current data (Ling et al. [2015]), in which 120 one-shot devices were divided into four accelerated conditions with higher-than-normal temperature and electric current, and inspected at three different times (see Table 7.6.1). In Table 7.6.3, estimates of the model parameters by the use of the proportional hazards model and the Weibull distribution are provided, for different values of the tuning parameter. Estimates of reliabilities and confidence intervals under the proportional hazards assumption are given in Table 7.6.2. Table 7.6.3 also presents the dvalues of the distance-statistic M_β and the corresponding p-values. From these values, it seems that the proportional hazards assumption fits the data at least as well as the Weibull model. The best fit is obtained for $\beta = 0.5$. To complete the study, Warwick and Jones [2005] approach is achieved for different values of the pilot estimator in a grid of width 100. However, as pointed out before, the final choice of the optimal tuning parameter depends too much on the pilot estimator used (see Figure 7.6.1).

Table 7.6.1: Electric Current data

Inspection Time τ_i	2	2	2	2	5	5	5	5	8	8	8	8
Temperature x_{s1}	55	80	55	80	55	80	55	80	55	80	55	80
Electric current x_{s2}	70	70	100	100	70	70	100	100	70	70	100	100
Number of failures n_{is}	4	8	9	8	7	9	9	9	6	10	9	10
Number of tested items K_{is}	10	10	10	10	10	10	10	10	10	10	10	10

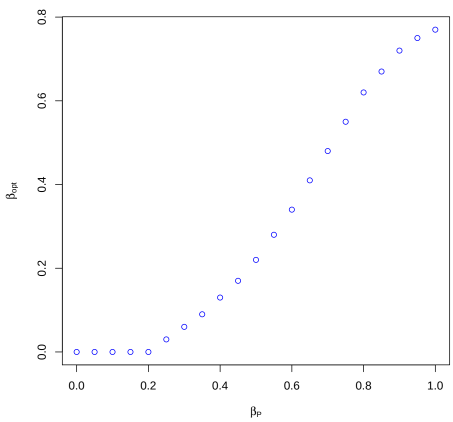

Figure 7.6.1: Electric Current data: estimation of the optimal tuning parameter depending on a pilot estimator by Warwick and Jones procedure

Table 7.6.2: Electric Current data: estimates of reliabilities and corresponding confidence intervals

β	$R(2, 25, 35; \widehat{\boldsymbol{\theta}}_\beta)$	$R(5, 25, 35; \widehat{\boldsymbol{\theta}}_\beta)$	$R(8, 25, 35; \widehat{\boldsymbol{\theta}}_\beta)$
0	0.817 (0.516, 0.949)	0.739 (0.397, 0.924)	0.689 (0.336, 0.907)
0.1	0.824 (0.526, 0.952)	0.751 (0.412, 0.928)	0.704 (0.353, 0.912)
0.2	0.833 (0.535, 0.956)	0.765 (0.427, 0.934)	0.721 (0.370, 0.919)
0.3	0.843 (0.545, 0.960)	0.780 (0.442, 0.941)	0.740 (0.387, 0.927)
0.4	0.855 (0.555, 0.965)	0.797 (0.457, 0.948)	0.760 (0.405, 0.936)
0.5	0.868 (0.566, 0.971)	0.816 (0.474, 0.956)	0.782 (0.423, 0.946)
0.6	0.884 (0.581, 0.976)	0.837 (0.493, 0.965)	0.807 (0.445, 0.956)
0.7	0.901 (0.601, 0.982)	0.861 (0.516, 0.973)	0.836 (0.471, 0.967)
0.8	0.918 (0.626, 0.987)	0.885 (0.544, 0.980)	0.863 (0.503, 0.975)
0.9	0.931 (0.649, 0.990)	0.902 (0.570, 0.985)	0.884 (0.533, 0.981)

Table 7.6.3: Electric Current data: one-shot device testing data analysis by using the proportional hazards model and the Weibull distribution

	Proportional Hazards model							Weibull distribution					
β	M_β	p-value	T°	current	η_1	η_2	η_3	M_β	p-value	intercept	T°	current	shape
0	1.80	0.695	0.023	0.018	0.123	0.543	-2.182	1.80	0.695	7.022	-0.053	-0.040	-0.817
0.1	1.72	0.745	0.024	0.018	0.141	0.555	-2.283	1.72	0.745	7.398	-0.055	-0.043	-0.845
0.2	1.65	0.796	0.024	0.019	0.156	0.565	-2.399	1.65	0.796	7.803	-0.057	-0.046	-0.869
0.3	1.58	0.833	0.025	0.020	0.167	0.572	-2.534	1.57	0.833	8.254	-0.060	-0.050	-0.890
0.4	1.49	0.931	0.026	0.022	0.177	0.579	-2.695	1.49	0.931	8.747	-0.064	-0.054	-0.906
0.5	1.40	0.942	0.027	0.023	0.183	0.582	-2.887	1.40	0.942	9.324	-0.068	-0.058	-0.920
0.6	1.51	0.892	0.029	0.025	0.187	0.585	-3.130	1.51	0.892	10.026	-0.073	-0.063	-0.931
0.7	1.64	0.876	0.031	0.027	0.190	0.586	-3.438	1.64	0.876	10.868	-0.079	-0.069	-0.938
0.8	1.76	0.861	0.033	0.030	0.189	0.586	-3.798	1.76	0.861	11.827	-0.086	-0.076	-0.942
0.9	1.84	0.750	0.036	0.032	0.185	0.582	-4.106	1.84	0.750	12.575	-0.091	-0.082	-0.938

Chapter 8

Robust inference for one-shot device testing under exponential distribution and competing risks

8.1 Introduction

In lifetime data analysis, it is often the case that the products under study can experience one of different types of failure. For example, in the context of survival analysis, we can have several different types of failure (death, relapse, opportunistic infection, etc.) that are of interest to us, leading to the so-called "competing risks" scenario. A competing risk is an event whose occurrence precludes the occurrence of the primary event of interest. In a study examining time to death attributable, for instance, to cardiovascular causes, death attributable to noncardiovascular causes would be a competing risk. Crowder [2006] has presented review of this competing risks problem for which one needs to estimate the failure rates for each cause. Balakrishnan et al. [2015a,b] and So [2016] have discussed the problem of one-shot devices under competing risk for the first time. However, in previous chapters, it was assumed that there is only one survival endpoint of interest, and that censoring is independent of the event in interest. The main purpose of this chapter is to develop weighted minimum DPD estimators as well as Wald-type test statistics under competing risk models for one-shot device testing assuming exponential lifetimes.

In Section 8.2, we present the model formulation as well as the notation to be used the rest of the chapter. The weighted minimum DPD estimators for one-shot device testing exponential model under competing risks are then developed in Section 8.3. Their asymptotic distribution and a new family of Wald-type test statistics based on them are also presented in this section. In Section 8.4, an extensive Monte Carlo simulation study is carried out for demonstrating the robust behaviour of the proposed estimators as well as the testing procedures. The developed methods are then applied to a pharmacology data for illustrative purposes.

The results of this Chapter have been writen in the form of a paper (Balakrishnan et al. [2020d]).

8.2 Model description and MLE

In this section, we shall introduce the notation necessary for the developments in this chapter, paying special attention to the MLE of the model, as well as its relation with the minimization of Kullback-Leibler divergence.

The setting for an accelerate life-test for one-shot devices under competing risks considered here is stratified in I testing conditions as follows:

1. The tests are checked at inspection times IT_i, for $i = 1, \ldots, I$;
2. The devices are tested under J different stress levels, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iJ})^T$, for $i = 1, \ldots, I$;
3. K_i devices are tested in the ith test condition, for $i = 1, \ldots, I$;

4. The number of devices failed due to the r-th cause under the i-th test condition is denoted by n_{ir}, for $i = 1, \ldots I,\ \ r = 1, \ldots, R$;
5. The number of devices that survive under the i-th test condition is denoted by $n_{i0} = K_i - \sum_{r=1}^{R} n_{ir}$.

Table 8.2.1: One-shot device testing under competing risks.

				Failures			Stress levels		
Condition	Times	Devices	Survivals	Cause 1	$\cdots$	Cause R	Stress 1	$\cdots$	Stress J
1	IT_1	K_1	n_{10}	n_{11}	$\cdots$	n_{1R}	x_{11}	$\cdots$	x_{1J}
2	IT_2	K_2	n_{20}	n_{21}	$\cdots$	n_{2R}	x_{21}	$\cdots$	x_{2J}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$		$\vdots$
I	IT_I	K_I	n_{I0}	n_{I1}	$\cdots$	n_{IR}	x_{I1}	$\cdots$	x_{IJ}

This setting is summarized in Table 8.2.1. For simplicity, and as considered in Balakrishnan et al. [2015a], we will limit in this chapter, the number of stress levels to $J = 1$ and the number of competing causes to $R = 2$, even though inference for the general case when $J > 1$ and $R > 2$ can be presented in an analogous manner.

Let us denote the random variable for the failure time due to causes 1 and 2 as T_{irk}, for $r = 1, 2,\ i = 1, \ldots, I$, and $k = 1, \ldots, K_i$, respectively. We now assume that T_{irk} follows an exponential distribution with failure rate parameter $\lambda_{ir}(\boldsymbol{\theta})$ and its probability density function

$$\begin{aligned} f_r(t; x_i, \boldsymbol{\theta}) &= \lambda_{ir}(\boldsymbol{\theta}) e^{-\lambda_{ir}(\boldsymbol{\theta})t}, \quad t > 0, \\ \lambda_{ir}(\boldsymbol{\theta}) &= \theta_{r0} \exp(\theta_{r1} x_i), \\ \boldsymbol{\theta} &= (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21})^T, \quad \theta_{r0}, \theta_{r1} > 0, \quad r = 1, 2, \end{aligned}$$

where x_i is the stress factor of the condition i and $\boldsymbol{\theta}$ is the model parameter vector, with $\boldsymbol{\theta} \in \mathbb{R}^4$.

We shall use $\pi_{i0}(\boldsymbol{\theta})$, $\pi_{i1}(\boldsymbol{\theta})$ and $\pi_{i2}(\boldsymbol{\theta})$ for the survival probability, failure probability due to cause 1 and failure probability due to cause 2, respectively. Their expressions are

$$\begin{aligned} \pi_{i0}(\boldsymbol{\theta}) &= (1 - F_1(IT_i; x_i, \boldsymbol{\theta}))(1 - F_2(IT_i; x_i, \boldsymbol{\theta})) = \exp(-(\lambda_{i1} + \lambda_{i2})IT_i), \\ \pi_{i1}(\boldsymbol{\theta}) &= \frac{\lambda_{i1}}{\lambda_{i1} + \lambda_{i2}}(1 - \exp(-(\lambda_{i1} + \lambda_{i2})IT_i)), \\ \pi_{i2}(\boldsymbol{\theta}) &= \frac{\lambda_{i2}}{\lambda_{i1} + \lambda_{i2}}(1 - \exp(-(\lambda_{i1} + \lambda_{i2})IT_i)), \end{aligned}$$

where $\lambda_{ir} = \lambda_{ir}(\boldsymbol{\theta}),\ r = 1, 2$. Derivations of these expressions can be found in So [2016] (pp. 151). Now, the likelihood function is given by

$$\mathcal{L}(n_{01}, \ldots, n_{I2}; \boldsymbol{\theta}) \propto \prod_{i=1}^{I} \pi_{i0}(\boldsymbol{\theta})^{n_{i0}} \pi_{i1}(\boldsymbol{\theta})^{n_{i1}} \pi_{i2}(\boldsymbol{\theta})^{n_{i2}}, \tag{8.1}$$

where $n_{0i} + n_{1i} + n_{2i} = K_i,\ i = 1, \ldots, I$.

Now, we present the classical definition of the MLE.

Definition 8.1 *The MLE of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, is obtained by maximizing the likelihood function in (8.1) or, equivalently, its logarithm.*

We will present an alternative definition of the MLE later on (see Definition 8.3). Let us introduce the following probability vectors:

$$\widehat{\boldsymbol{p}}_i = (\widehat{p}_{i0}, \widehat{p}_{i1}, \widehat{p}_{i2})^T = \frac{1}{K_i}(n_{i0}, n_{i1}, n_{i2})^T, \quad i = 1, \ldots, I, \tag{8.2}$$

$$\boldsymbol{\pi}_i(\boldsymbol{\theta}) = (\pi_{i0}(\boldsymbol{\theta}), \pi_{i1}(\boldsymbol{\theta}), \pi_{i2}(\boldsymbol{\theta}))^T, \quad i = 1, \ldots, I. \tag{8.3}$$

The Kullback-Leibler divergence measure (see, for instance, Pardo [2005]), between $\widehat{\boldsymbol{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\theta})$, is given by

$$\begin{aligned}
d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \sum_{r=0}^{2} \widehat{p}_{ir} \log\left(\frac{\widehat{p}_{ir}}{\pi_{ir}(\boldsymbol{\theta})}\right) \\
&= \widehat{p}_{i0} \log\left(\frac{\widehat{p}_{i0}}{\pi_{i0}(\boldsymbol{\theta})}\right) + \widehat{p}_{i1} \log\left(\frac{\widehat{p}_{i1}}{\pi_{i1}(\boldsymbol{\theta})}\right) + \widehat{p}_{i2} \log\left(\frac{\widehat{p}_{i2}}{\pi_{i2}(\boldsymbol{\theta})}\right) \\
&= \frac{n_{i0}}{K_i} \log\left(\frac{n_{i0}/K_i}{\pi_{i0}(\boldsymbol{\theta})}\right) + \frac{n_{i1}}{K_i} \log\left(\frac{n_{i1}/K_i}{\pi_{i1}(\boldsymbol{\theta})}\right) + \frac{n_{i2}}{K_i} \log\left(\frac{n_{i2}/K_i}{\pi_{i2}(\boldsymbol{\theta})}\right) \\
&= \frac{1}{K_i}\left\{ n_{i0} \log\left(\frac{n_{i0}/K_i}{\pi_{i0}(\boldsymbol{\theta})}\right) + n_{i1} \log\left(\frac{n_{i1}/K_i}{\pi_{i1}(\boldsymbol{\theta})}\right) + n_{i2} \log\left(\frac{n_{i2}/K_i}{\pi_{i2}(\boldsymbol{\theta})}\right)\right\},
\end{aligned}$$

and the weighted Kullback-Leibler divergence measure is given by

$$\begin{aligned}
d_{KL}^{W}(\boldsymbol{\theta}) &= \sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) \\
&= \frac{1}{K} \sum_{i=1}^{I} \left\{ n_{i0} \log\left(\frac{n_{i0}/K_i}{\pi_{i0}(\boldsymbol{\theta})}\right) + n_{i1} \log\left(\frac{n_{i1}/K_i}{\pi_{i1}(\boldsymbol{\theta})}\right) + n_{i2} \log\left(\frac{n_{i2}/K_i}{\pi_{i2}(\boldsymbol{\theta})}\right)\right\},
\end{aligned}$$

with $K = K_1 + \cdots + K_I$.

Theorem 8.2 *The likelihood function $\mathcal{L}(n_{01}, \ldots, n_{I2}; \boldsymbol{\theta})$, given in (8.1), is related to the weighted Kullback-Leibler divergence measure through*

$$d_{KL}^{W}(\boldsymbol{\theta}) = \sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = c - \frac{1}{K} \log \mathcal{L}(n_{01}, \ldots, n_{I2}; \boldsymbol{\theta}), \tag{8.4}$$

with c being a constant, not dependent on $\boldsymbol{\theta}$.

Proof. We have

$$\begin{aligned}
\sum_{i=1}^{I} \frac{K_i}{K} d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) &= \frac{1}{K} \sum_{i=1}^{I} \left\{ n_{i0} \log\left(\frac{n_{i0}/K_i}{\pi_{i0}(\boldsymbol{\theta})}\right) + n_{i1} \log\left(\frac{n_{i1}/K_i}{\pi_{i1}(\boldsymbol{\theta})}\right) + n_{i2} \log\left(\frac{n_{i2}/K_i}{\pi_{i2}(\boldsymbol{\theta})}\right)\right\} \\
&= \frac{1}{K} \sum_{i=1}^{I} \left\{ n_{i0} \log\left(\frac{n_{i0}}{K_i}\right) + n_{i1} \log\left(\frac{n_{i1}}{K_i}\right) + n_{i2} \log\left(\frac{n_{i2}}{K_i}\right)\right\} \\
&- \frac{1}{K} \sum_{i=1}^{I} \left\{ n_{i0} \log\left(\pi_{i0}(\boldsymbol{\theta})\right) + n_{i1} \log\left(\pi_{i1}(\boldsymbol{\theta})\right) + n_{i2} \log\left(\pi_{i2}(\boldsymbol{\theta})\right)\right\} \\
&= c - \frac{1}{K} \log\left(\prod_{i=1}^{I} \pi_{i0}(\boldsymbol{\theta})^{n_{i0}} \pi_{i1}(\boldsymbol{\theta})^{n_{i1}} \pi_{i2}(\boldsymbol{\theta})^{n_{i2}}\right) \\
&= c - \frac{1}{K} \log\left(\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\delta}, \boldsymbol{IT}, \boldsymbol{x})\right),
\end{aligned}$$

where $c = \frac{1}{K} \sum_{i=1}^{I} \sum_{r=0}^{2} \left\{ n_{ir} \log\left(\frac{n_{ir}}{K_i}\right)\right\}$ and it does not depend on the parameter vector $\boldsymbol{\theta}$. ■

Based on Theorem 8.2 we can give the following alternative definition for the MLE.

Definition 8.3 *The MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, can be obtained by the minimization of the weighted Kullback-Leibler divergence measure given in (8.4).*

8.3 Weighted minimum DPD estimator

8.3.1 Definition

Given the probability vectors $\widehat{\boldsymbol{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\theta})$, defined in (8.2) and (8.3), respectively, the DPD between both probability vectors is given by

$$\begin{aligned} d_\beta(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = & \left(\pi_{i0}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta+1}(\boldsymbol{\theta})\right) \\ & - \frac{\beta+1}{\beta}\left(\widehat{p}_{i0}\pi_{i0}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i1}\pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2}\pi_{i2}^{\beta}(\boldsymbol{\theta})\right) \\ & + \frac{1}{\beta}\left(\widehat{p}_{i0}^{\beta+1} + \widehat{p}_{i1}^{\beta+1} + \widehat{p}_{i2}^{\beta+1}\right), \quad \text{if } \beta > 0, \end{aligned}$$

and $d_{\beta=0}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = \lim_{\beta\to 0^+} d_\beta(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta})) = d_{KL}(\widehat{\boldsymbol{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\theta}))$, for $\beta = 0$.

The weighted DPD is given by

$$\begin{aligned} d_\beta^W(\boldsymbol{\theta}) = & \sum_{i=1}^{I}\frac{K_i}{K}\left[\left(\pi_{i0}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta+1}(\boldsymbol{\theta})\right)\right. \\ & \left. - \frac{\beta+1}{\beta}\left(\widehat{p}_{i0}\pi_{i0}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i1}\pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2}\pi_{i2}^{\beta}(\boldsymbol{\theta})\right) + \frac{1}{\beta}\left(\widehat{p}_{i0}^{\beta+1} + \widehat{p}_{i1}^{\beta+1} + \widehat{p}_{i2}^{\beta+1}\right)\right] \end{aligned}$$

but the term $\frac{1}{\beta}\left(\widehat{p}_{i0}^{\beta+1} + \widehat{p}_{i1}^{\beta+1} + \widehat{p}_{i2}^{\beta+1}\right)$, $i = 1, ..., I$, does not have any role in its minimization with respect to $\boldsymbol{\theta}$. Therefore, in order to minimize $d_\beta^W(\boldsymbol{\theta})$, we can consider the equivalent measure

$$\begin{aligned} {}^*d_\beta^W(\boldsymbol{\theta}) = & \sum_{i=1}^{I}\frac{K_i}{K}\left[\left(\pi_{i0}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i1}^{\beta+1}(\boldsymbol{\theta}) + \pi_{i2}^{\beta+1}(\boldsymbol{\theta})\right)\right. \\ & \left. - \frac{\beta+1}{\beta}\left(\widehat{p}_{i0}\pi_{i0}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i1}\pi_{i1}^{\beta}(\boldsymbol{\theta}) + \widehat{p}_{i2}\pi_{i2}^{\beta}(\boldsymbol{\theta})\right)\right]. \end{aligned} \tag{8.5}$$

Definition 8.4 *We can define the weighted minimum DPD estimator of $\boldsymbol{\theta}$ as*

$$\widehat{\boldsymbol{\theta}}_\beta = \arg\min_{\boldsymbol{\theta}\in\Theta} {}^*d_\beta^W(\boldsymbol{\theta}), \quad \textit{for } \beta > 0$$

and for $\beta = 0$ we get the weighted maximum likelihood estimator.

8.3.2 Estimation and asymptotic distribution

Theorem 8.5 *The weighted minimum DPD estimator of $\boldsymbol{\theta}$, with tuning parameter $\beta \geq 0$, $\widehat{\boldsymbol{\theta}}_\beta$, can be obtained as the solution of the following system of four equations:*

$$\sum_{i=1}^{I} K_i\left\{-\pi_{i0}(\boldsymbol{\theta})IT_i\left[\pi_{i0}(\boldsymbol{\theta})^{\beta-1}(\pi_{i0}(\boldsymbol{\theta}) - p_{i0}) - (1-\pi_{i0}(\boldsymbol{\theta}))^{\beta-1}\Gamma_{i,\beta}\right]\boldsymbol{l}_i + (1-\pi_{i0}(\boldsymbol{\theta}))^{\beta}\Gamma_{i,\beta}^*\right\} = \boldsymbol{0}_4,$$

where

$$\Gamma_{i,\beta} = \frac{\lambda_{i1}^{\beta}\left[\frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}}(1-\pi_{i0}(\boldsymbol{\theta})) - p_{i1}\right] + \lambda_{i2}^{\beta}\left[\frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}(1-\pi_{i0}(\boldsymbol{\theta})) - p_{i2}\right]}{(\lambda_{i1}+\lambda_{i2})^{\beta}},$$

$$\Gamma_{i,\beta}^* = \frac{\lambda_{i1}^{\beta-1}\left[\frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}}(1-\pi_{i0}(\boldsymbol{\theta})) - p_{i1}\right] - \lambda_{i2}^{\beta-1}\left[\frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}(1-\pi_{i0}(\boldsymbol{\theta})) - p_{i2}\right]}{(\lambda_{i1}+\lambda_{i2})^{\beta-1}},$$

$\boldsymbol{l}_i = (\lambda_{i1}/\theta_{10}, \lambda_{i1}x_i, \lambda_{i2}/\theta_{20}, \lambda_{i2}x_i)^T$ *and* $\boldsymbol{r}_i = \frac{\lambda_{i1}\lambda_{i2}}{(\lambda_{i1}+\lambda_{i2})^2}(1/\theta_{10}, x_i, -1/\theta_{20}, -x_i)^T$.

Proof. The estimating equations are given by

$$\frac{\partial}{\partial \boldsymbol{\theta}} {}^{*}d_{\beta}^{W}(\boldsymbol{\theta}) = \mathbf{0}_4, \tag{8.6}$$

where ${}^{*}d_{\beta}^{Weighted}(\boldsymbol{\theta})$ is as given in (8.5). Equation (8.6) is equivalent to

$$\frac{1}{\beta+1}\frac{\partial}{\partial \boldsymbol{\theta}}\sum_{i=1}^{I}\sum_{r=0}^{2} K_i \pi_{ir}^{\beta+1}(\boldsymbol{\theta}) - \frac{1}{\beta}\frac{\partial}{\partial \boldsymbol{\theta}}\sum_{i=1}^{I}\sum_{r=0}^{2} K_i p_{ir}\pi_{ir}^{\beta}(\boldsymbol{\theta}) = \mathbf{0}_4; \tag{8.7}$$

that is,

$$\frac{1}{\beta+1}\sum_{i=1}^{I}\sum_{r=0}^{2} K_i(\beta+1)\pi_{ir}^{\beta}(\boldsymbol{\theta})\frac{\partial \pi_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{1}{\beta}\sum_{i=1}^{I}\sum_{r=0}^{2} K_i p_{ir}\beta\pi_{ir}^{\beta-1}(\boldsymbol{\theta})\frac{\partial \pi_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}_4,$$

or, equivalently

$$\sum_{i=1}^{I}\sum_{r=0}^{2} K_i \pi_{ir}^{\beta-1}(\boldsymbol{\theta})\frac{\partial \pi_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}[\pi_{ir}(\boldsymbol{\theta}) - p_{ir}] = \mathbf{0}_4.$$

But,

$$\begin{aligned}
\pi_{i0}(\boldsymbol{\theta}) &= \exp(-(\lambda_{i1}+\lambda_{i2})IT_i), \\
\pi_{i1}(\boldsymbol{\theta}) &= \frac{\lambda_{i1}}{\lambda_{i2}+\lambda_{2i}}(1-\exp(-(\lambda_{i1}+\lambda_{i2})IT_i)) = \frac{\lambda_{i1}}{\lambda_{i2}+\lambda_{2i}}(1-\pi_{i0}(\boldsymbol{\theta})), \\
\pi_{i2}(\boldsymbol{\theta}) &= \frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}(1-\exp(-(\lambda_{i1}+\lambda_{i2})IT_i)) = \frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}(1-\pi_{i0}(\boldsymbol{\theta})),
\end{aligned}$$

and so

$$\frac{\partial \pi_{i0}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -IT_i\pi_{i0}(\boldsymbol{\theta})\frac{\partial}{\partial \boldsymbol{\theta}}[\lambda_{i1}+\lambda_{i2}] = -IT_i\pi_{i0}(\boldsymbol{\theta})(\lambda_{i1}/\theta_{10}, \lambda_{i1}x_i, \lambda_{i2}/\theta_{20}, \lambda_{i2}x_i)^T = -IT_i\pi_{i0}(\boldsymbol{\theta})\boldsymbol{l}_i,$$

$$\begin{aligned}
\frac{\partial \pi_{i1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= (1-\pi_{i0}(\boldsymbol{\theta}))\left[\frac{\partial}{\partial \boldsymbol{\theta}}\frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}}\right] - \frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}}\frac{\partial \pi_{i0}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\
\frac{\partial \pi_{i2}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= (1-\pi_{i0}(\boldsymbol{\theta}))\left[\frac{\partial}{\partial \boldsymbol{\theta}}\frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}\right] - \frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}\frac{\partial \pi_{i0}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},
\end{aligned}$$

where

$$\left[\frac{\partial}{\partial \boldsymbol{\theta}}\frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}}\right] = -\left[\frac{\partial}{\partial \boldsymbol{\theta}}\frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}\right] = \frac{\lambda_{i1}\lambda_{i2}}{(\lambda_{i1}+\lambda_{i2})^2}(1/\alpha_{10}, x_i, -1/\alpha_{20}, -x_i)^T.$$

We then obtain the desired result. ■

Now, by using (1.11), we can obtain the asymptotic distribution of the above weighted minimum DPD estimator.

Theorem 8.6 *Let $\boldsymbol{\theta}^0$ be the true value of the parameter $\boldsymbol{\theta}$. The asymptotic distribution of the weighted minimum DPD estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_\beta$, is given by*

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^0\right) \xrightarrow[K\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_4, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\boldsymbol{K}_\beta(\boldsymbol{\theta}^0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}^0)\right),$$

where

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{r=0}^{2}\frac{K_i}{K}\boldsymbol{u}_{ir}^{*}(\boldsymbol{\theta})\boldsymbol{u}_{ir}^{*T}(\boldsymbol{\theta})\pi_{ir}^{\beta-1}(\boldsymbol{\theta}), \tag{8.8}$$

$$\boldsymbol{K}_{\beta}(\boldsymbol{\theta})=\sum_{i=1}^{I}\sum_{r=0}^{2}\frac{K_i}{K}\boldsymbol{u}_{ir}^{*}(\boldsymbol{\theta})\boldsymbol{u}_{ir}^{*T}(\boldsymbol{\theta})\pi_{ir}^{2\beta-1}(\boldsymbol{\theta})-\sum_{i=1}^{I}\frac{K_i}{K}\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})\boldsymbol{\xi}_{i,\beta}^{T}(\boldsymbol{\theta}), \tag{8.9}$$

with $\boldsymbol{\xi}_{i,\beta}(\boldsymbol{\theta})=\sum_{r=0}^{2}\boldsymbol{u}_{ir}^{*}(\boldsymbol{\theta})\pi_{ir}^{\beta}(\boldsymbol{\theta})$ *and* $\boldsymbol{u}_{ir}^{*}(\boldsymbol{\theta})=\frac{\partial\pi_{ir}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{T}}$, *where*

$$\begin{aligned}
\frac{\partial\pi_{i0}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= -IT_i\pi_{i0}(\boldsymbol{\theta})\boldsymbol{l}_i,\\
\frac{\partial\pi_{i1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}}IT_i\pi_{i0}(\boldsymbol{\theta})\boldsymbol{l}_i+(1-\pi_{i0}(\boldsymbol{\theta}))\boldsymbol{r}_i,\\
\frac{\partial\pi_{i2}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}}IT_i\pi_{i0}(\boldsymbol{\theta})\boldsymbol{l}_i-(1-\pi_{i0}(\boldsymbol{\theta}))\boldsymbol{r}_i,
\end{aligned}$$

$\boldsymbol{l}_i=(\lambda_{i1}/\theta_{10},\lambda_{i1}x_i,\lambda_{i2}/\theta_{20},\lambda_{i2}x_i)^T$ *and* $\boldsymbol{r}_i=\frac{\lambda_{i1}\lambda_{i2}}{(\lambda_{i1}+\lambda_{i2})^2}(1/\theta_{10},x_i,-1/\theta_{20},-x_i)^T$.

Proof. Directly from (1.11) and proof of Theorem 8.5. ■

8.3.3 Wald-type tests

Let us consider the function $\boldsymbol{m}:\mathbb{R}^{J+1}\longrightarrow\mathbb{R}^{r}$, where $r\leq 4$. Then

$$\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{0}_r, \tag{8.10}$$

with $\boldsymbol{0}_r$ being the null column vector of dimension r, which represents the null hypothesis. We assume that the $4\times r$ matrix

$$\boldsymbol{M}(\boldsymbol{\theta})=\frac{\partial\boldsymbol{m}^{T}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

exists and is continuous in "$\boldsymbol{\theta}$ " and that rank$(\boldsymbol{M}(\boldsymbol{\theta}))=r$. For testing

$$H_0:\boldsymbol{\theta}\in\boldsymbol{\Theta}_0 \text{ against } H_1:\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0, \tag{8.11}$$

where

$$\boldsymbol{\Theta}_0=\{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0:\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{0}_r\},$$

we can consider the following Wald-type test statistics:

$$W_K\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)=K\boldsymbol{m}^{T}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\left(\boldsymbol{M}^{T}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\boldsymbol{\Sigma}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\boldsymbol{M}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\right)^{-1}\boldsymbol{m}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right),$$

where

$$\boldsymbol{\Sigma}_{\beta}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)=\boldsymbol{J}_{\beta}^{-1}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\boldsymbol{K}_{\beta}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\boldsymbol{J}_{\beta}^{-1}\left(\widehat{\boldsymbol{\theta}}_{\beta}\right),$$

and $\boldsymbol{J}_{\beta}(\boldsymbol{\theta})$ and $\boldsymbol{K}_{\beta}(\boldsymbol{\theta})$ are as given in (8.8) and (8.9), respectively.

Theorem 8.7 *Under the null hypothesis, we have*

$$W_K\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)\underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}}\chi_r^2.$$

Based on Theorem 8.7 , we can reject the null hypothesis, in (8.11), if

$$W_K\left(\widehat{\boldsymbol{\theta}}_{\beta}\right)>\chi_{r,\alpha}^2, \tag{8.12}$$

where $\chi_{r,\alpha}^2$ is the upper α percentage point of χ_r^2 distribution.

Remark 8.8 (Robustness properties) *In Chapters 2 and 3, the robustness of the weighted minimum DPD estimators and Wald-type tests, for $\beta>0$, was theoretically derived through local dependence under the exponential assumption but in a non-competing risk framework, for large leverages $\boldsymbol{x}_i$s. Analogous computations would result in the same conclusion for the competing risks scenario. However, we could not directly infer about the robustness against outliers in the response variable which are, in fact, the misspecification errors. In the next section, a simulation study is carried out in order to empirically illustrate the robustness of the proposed statistics with $\beta>0$, and the non-robustness when $\beta=0$, also against such misspecification errors.*

8.4 Simulation Study

In this section, a Monte Carlo simulation study that examines the accuracy of the proposed weighted minimum DPD estimators is presented. Section 8.4.1 focuses on the efficiency, measured in terms of root of mean square error (RMSE), mean bias error (MBE) and mean absolute error (MAE), of the estimators of model parameters, while Section 8.4.2 examines the behavior of Wald-type tests developed in preceding sections. Every step of simulation was tested under S = 5,000 replications with R statistical software. The main purpose of this study is to show that within the family of weighted minimum DPD estimators, developed in the preceding sections, there are estimators with better robustness properties than the MLE, and the Wald-type tests constructed based on them are at the same time more robust than the classical Wald test constructed based on the MLE.

8.4.1 The weighted minimum DPD estimators

The lifetimes of devices are simulated for different levels of reliability and different sample sizes, under 4 different stress conditions with 1 stress factor at 4 levels. Then, all devices under each stress condition are inspected at 3 different inspection times, depending on the level of reliability. The corresponding data will then be collected under $I = 12$ test conditions.

A. Balanced data: Effect of the sample size

Firstly, a balanced data with equal sample size for each group was considered. K_i was taken to range from small to large sample sizes, two causes of failure were considered, and the model parameters were set to be $\boldsymbol{\theta} = (\theta_{10}, 0.05, \theta_{20}, 0.08)^T$ with $\theta_{10} \in \{0.008, 0.004, 0.001\}$ and $\theta_{20} \in \{0.0008, 0.0004, 0.0001\}$ for devices with low, moderate and high reliability, respectively. To prevent many zero-observations in test groups, the inspection times were set as $IT \in \{5, 10, 20\}$ for the case of low reliability, $IT \in \{7, 15, 25\}$ for the case of moderate reliability, and $IT \in \{10, 20, 30\}$ for the case of high reliability. To evaluate the robustness of the weighted minimum DPD estimators, we studied their behavior in the presence of an outlying cell for the first testing condition in our table. This cell was generated under the parameters $\tilde{\boldsymbol{\theta}} = (\theta_{10}, 0.05, \theta_{20}, 0.15)^T$. See Table 8.4.1 for a summary of these scenarios. RMSEs, MAEs and MBEs of model parameters were then computed for the cases of both pure and contaminated data and are plotted in Figures 8.5.1, 8.5.2 and 8.5.3, respectively, with similar conclusions for the three error measures.

For the case of pure data, MLE presents the best behaviour (overall in the model with high reliability) and an increment in the tuning parameter β leads to a gradual loss in terms of efficiency. However, in the case of contaminated data, MLE turns to be the worst estimator, and weighted minimum DPD estimators with $\beta > 0$ present much more robust behaviour. Note that, as expected, an increase in the sample size improves the efficiency of the estimators, both for pure and contaminated data.

B. Unbalanced data: Effect of the degree of contamination

Now, we consider an unbalanced data with unequal sample sizes for the test conditions. This data set, which consists a total of $K = 300$ devices, is presented in Table 8.4.2. A competing risks model, with two different causes of failure, was generated with parameters $\boldsymbol{\theta} = (0.001, 0.05, 0.0001, 0.08)^T$. To examine the robustness in this accelerated life test (ALT) plan (in which the devices are tested under high stress levels, so that more failures can be observed), we increased each of the parameters of the outlying first cell (Figure 8.4.1). The contaminated parameters are expressed by $\tilde{\theta}_{10}, \tilde{\theta}_{11}, \tilde{\theta}_{20}$ and $\tilde{\theta}_{21}$, respectively.

When there is no contamination in the cell or the degree of contamination is very low, and in concordance with results obtained in the previous scenario, MLE is observed to be the most efficient estimator. However, when the degree of contamination increases, there is an increase in

Table 8.4.1: Parameter values used in the simulation. Study of efficiency.

Reliability	**Parameters**	**Symbols**	**Values**
	Risk 1	$\theta_{10},\, \theta_{11}$	$0.008, 0.05$
Low reliability	Risk 2	$\theta_{20},\, \theta_{21}$	$0.0008, 0.08$
	Contamination	$\tilde{\theta}_{21}$	0.15
	Temperature (°C)	$\boldsymbol{x}^T = (x_1, x_2, x_3, x_4)$	$(35, 45, 55, 65)$
	Inspection Time (days)	$IT = \{IT_1, IT_2, IT_3\}$	$\{5, 10, 20\}$
	Risk 1	$\theta_{10},\, \theta_{11}$	$0.004, 0.05$
Moderate reliability	Risk 2	$\theta_{20},\, \theta_{21}$	$0.0004, 0.08$
	Contamination	$\tilde{\theta}_{21}$	0.15
	Temperature (°C)	$\boldsymbol{x}^T = (x_1, x_2, x_3, x_4)$	$(35, 45, 55, 65)$
	Inspection Time (days)	$IT = (IT_1, IT_2, IT_3)$	$\{7, 15, 25\}$
	Risk 1	$\theta_{10},\, \theta_{11}$	$0.001, 0.05$
High reliability	Risk 2	$\theta_{20},\, \theta_{21}$	$0.0001, 0.08$
	Contamination	$\tilde{\theta}_{21}$	0.15
	Temperature (°C)	$\boldsymbol{x}^T = (x_1, x_2, x_3, x_4)$	$(35, 45, 55, 65)$
	Inspection Time (days)	$IT = (IT_1, IT_2, IT_3)$	$\{10, 20, 30\}$

Table 8.4.2: ALT plan, unbalanced data.

i	x_i	IT_i	K_i
1	35	10	50
2	45	10	40
3	55	10	20
4	65	10	40
5	35	20	20
6	45	20	20
7	55	20	30
8	65	20	20
9	35	30	20
10	45	30	20
11	55	30	10
12	65	30	10

the error for all the estimators, but weighted minimum DPD estimators are shown to be much more robust. This is also the case for whatever choice of the contamination parameters we considered.

8.4.2 Wald-type tests

Let us consider the balanced data under moderate reliability defined in the previous section. To compute the accuracy in terms of contrast, we consider the testing problem

$$H_0 : \theta_{21} = 0.08 \quad \text{vs.} \quad H_1 : \theta_{21} \neq 0.08. \tag{8.13}$$

For computing the empirical test level, we measured the proportion of test statistics exceeding the corresponding chi-square critical value. The simulated test powers were also obtained under H_1 in (8.13) in a similar manner. We used a nominal level of 0.05. Table 8.4.3 summarizes the model considered for this purpose. As in the previous section, an outlying cell with $\tilde{\theta}_{21} = 0.15$ is considered to illustrate the robustness of the proposed Wald-type tests (Figure 8.4.2).

In the case of pure data, we see how a big sample size is needed to obtain empirical tests close to the nominal level. In the case of contaminated data, empirical test levels are far away from the nominal level, with the MLE again presenting the least robust behaviour.

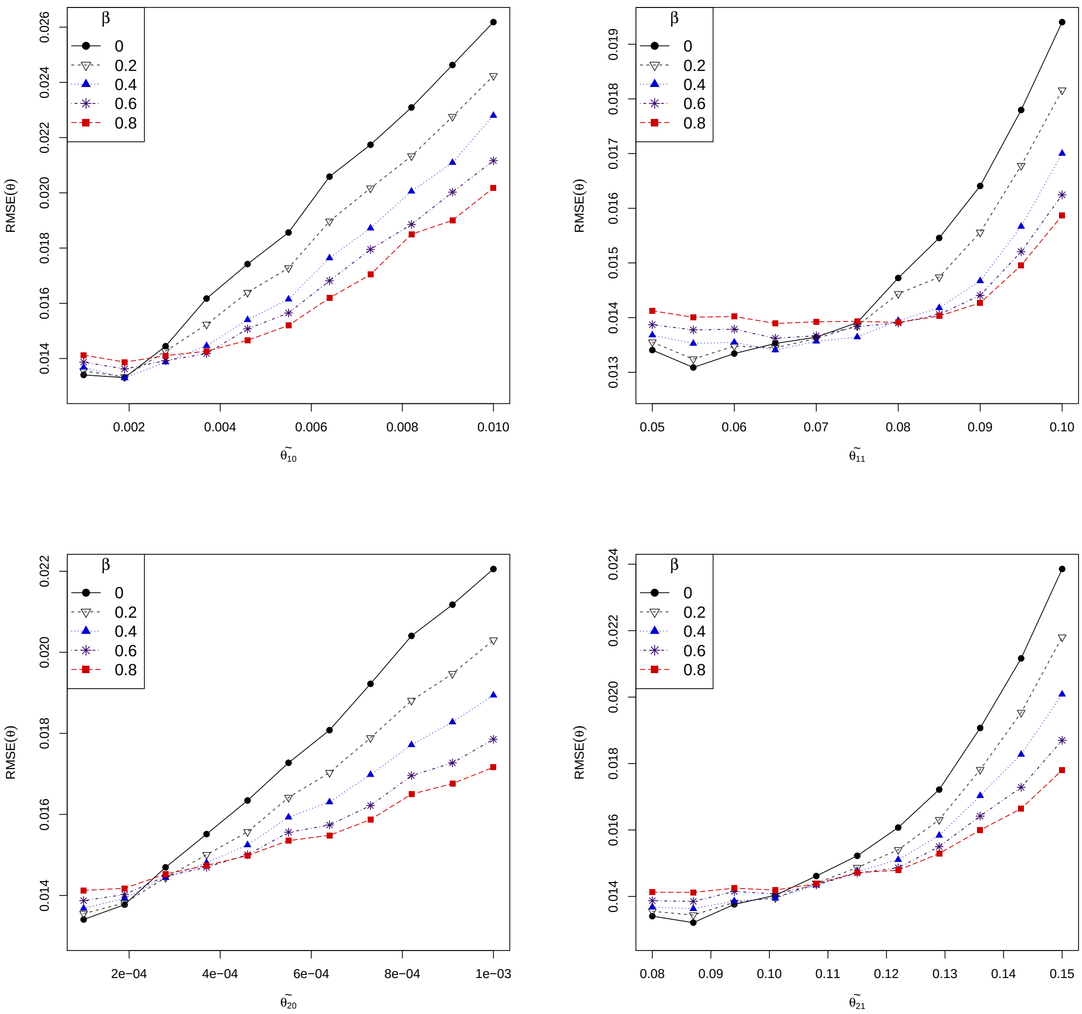

Figure 8.4.1: RMSEs of the weighted minimum DPD estimators of $\boldsymbol{\theta}$ for different contamination parameter values. Unbalanced data.

Table 8.4.3: Parameter values used in the simulation study of Wald-type tests.

Study	Parameters	Symbols	Values
Levels	Model True Parameters	$\boldsymbol{\theta}^T = (\theta_{10}, \theta_{11}, \theta 20, \theta_{21})$	$(0.004, 0.05, 0.0004, 0.08)$
Powers	Model True Parameters	$\boldsymbol{\theta}^T = (\theta_{10}, \theta_{11}, \theta 20, \theta_{21})$	$(0.004, 0.05, 0.0004, 0.09)$

This simulation study has illustrated well the robust properties of the weighted minimum DPD estimators for $\beta > 0$, which is inevitably accompanied with a loss of efficiency in a the case of pure data. It seems that a moderate low value of the tuning parameter can be a good choice when applying the estimators to a real data set. However, when dealing with specific data sets, especially when we have small data sets, a data driven procedure for the choice of tuning parameter will become necessary.

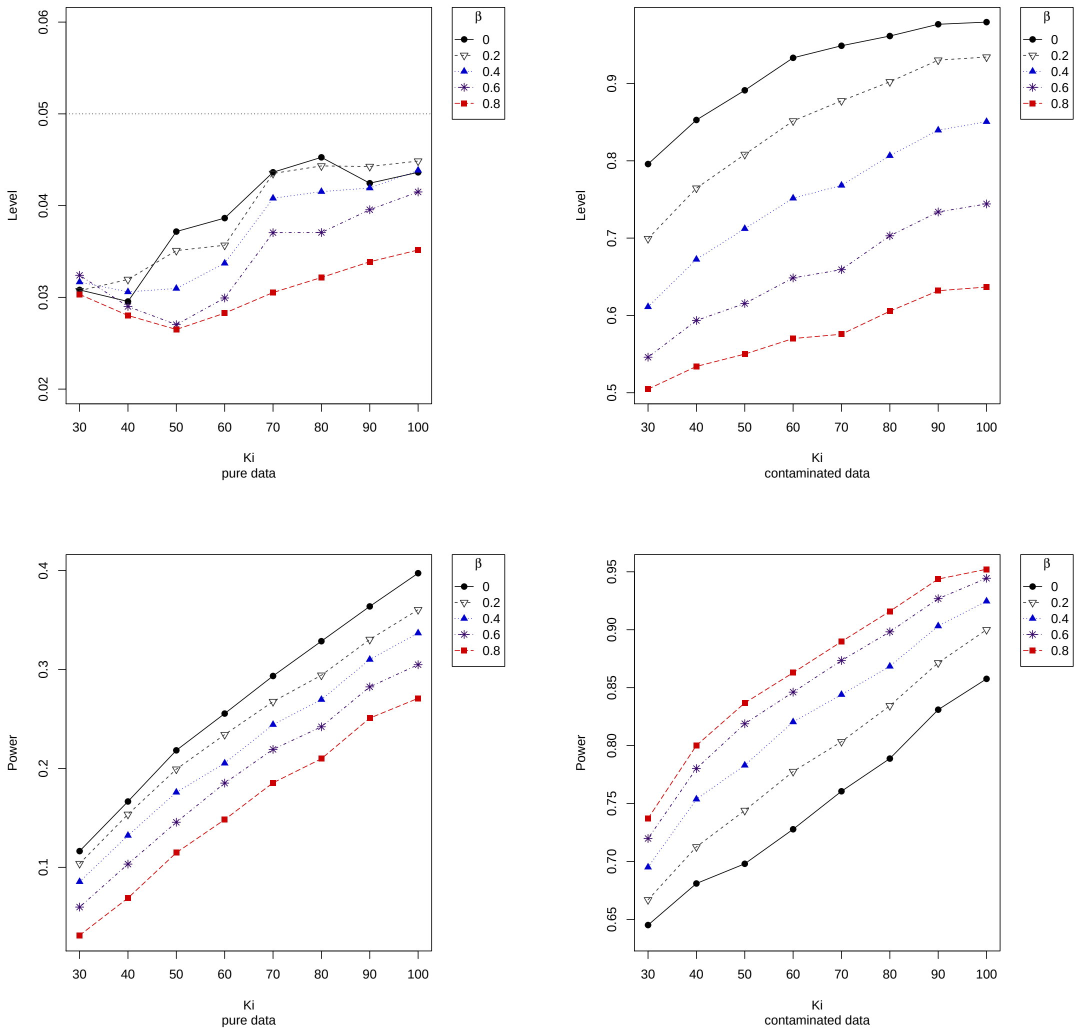

Figure 8.4.2: Levels and Powers of the weighted minimum DPD estimators-based Wald-type tests for different values of K_i with pure (left) and contaminated data (right), for the case of moderate reliability.

8.4.3 Choice of tuning parameter

Let us consider again the problem of choosing the optimal tuning parameter in a DPD-based family of estimators, which has been extensively discussed in the literature; see, for example, Hong and Kim [2001], Warwick and Jones [2005], and Ghosh and Basu [2015], which consisted on minimizing the estimated mean square error of the estimators, computed as the sum of estimated squared bias and variance; that is,

$$\widehat{MSE}_\beta = (\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_P)^T(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_P) + \frac{1}{K}\text{trace}\left[\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{K}_\beta(\widehat{\boldsymbol{\theta}}_\beta)\boldsymbol{J}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_\beta)\right],$$

where $\boldsymbol{\theta}_P$ is a pilot estimator, whose choice will be empirically discussed, since the overall procedure depends on this choice. We consider again the balanced scenario under moderate reliability discussed earlier. For different pilot estimators and a grid of 100 points, optimal tuning parameters and their corresponding RMSEs are computed (see Figure 8.4.3). The optimal tuning parameter increases when the contamination level increases in the data, and it seems that a moderate value of β is the best choice for the pilot estimator, as suggested also in previous chapters.

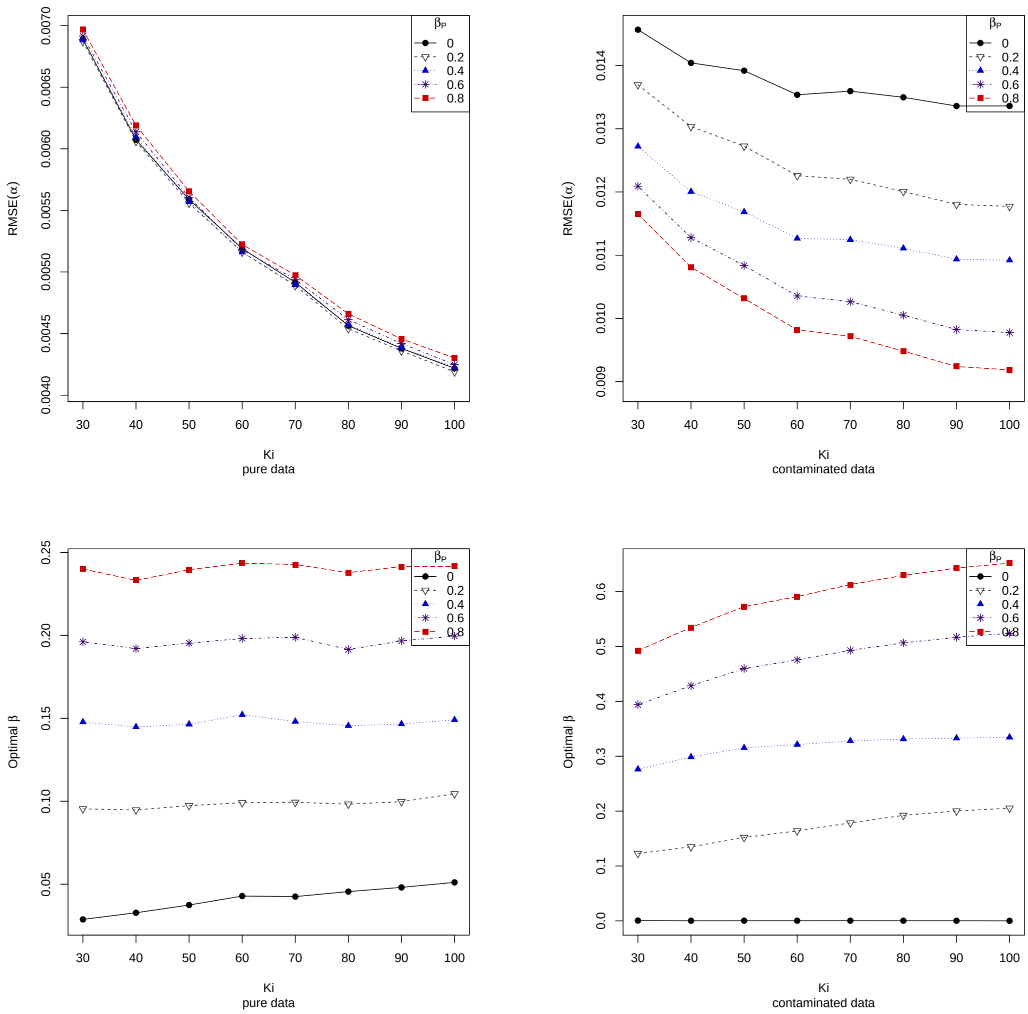

Figure 8.4.3: Estimated optimal β and the corresponding RMSEs for different pilot estimators in the proposed ad-hoc approach for the case of moderate reliabiulity.

8.5 Benzidine dihydrochloride experiment

Let us reconsider our example in Section 2.7.3, the Benzidine dihydrochloride experiment, to study the performance of the proposed procedures. As noted in Section 2.7.3, this experiment considers two different doses of drug induced in the mice: 60 parts per million ($x = 1$) and 400 parts per million ($x = 2$) and two causes of death are recorded: died without tumor ($r = 1$) and died with tumor ($r = 2$). The data are presented in Table 2.7.6.

Estimators of parameters were obtained for different choices of tuning parameters. We then computed the expected mean lifetime of the devices under the two doses of drug, both for the whole population ($E_{x=1}$ and $E_{x=2}$) and particularly for the mice that died without tumor ($E^1_{x=1}$ and $E^1_{x=2}$). We have also computed the probability of failure due to cause 1 (die without tumor) given failure, for both doses of drug ($P^1_{x=1}$ and $P^1_{x=2}$). We applied the procedure described in Section 8.4.3 to determine the optimal tuning parameter for this data set, over a grid of 100 points. The resulting optimal tuning parameter, 0.37, and its corresponding estimators are presented in Table 8.5.1. Finally, we estimate the errors, as given by

$$\frac{1}{3I}\sum_{i=1}^{I}\sum_{r=0}^{2}\left|\frac{n_{ir}-K_i\pi_{ir}(\widehat{\boldsymbol{\theta}})}{K_i}\right|, \tag{8.14}$$

for different tuning parameters β, and the corresponding results in Table 8.5.2. The minimum is obtained for $\beta = 0.8$, while $\beta = 0.37$ also presents a lower estimated error, which is in concordance with the estimate obtained earlier.

Table 8.5.1: Estimations for the BDC experiment for different hoices of tuning parameters

β	θ_{10}	θ_{11}	θ_{20}	θ_{21}	$E^1_{x=1}$	$E^1_{x=2}$	$E_{x=1}$	$E_{x=2}$	$P^1_{x=1}$	$P^1_{x=2}$
0	0.00089	1.3191	0.00028	2.493	300.545	80.355	150.203	18.952	0.4997	0.2358
0.1	0.00091	1.3072	0.00029	2.465	297.876	80.593	146.984	18.872	0.4934	0.2341
0.2	0.00094	1.2844	0.00031	2.441	295.010	81.658	144.138	18.869	0.4885	0.2310
0.3	0.00097	1.2627	0.00033	2.408	291.902	82.572	140.528	18.818	0.4814	0.2279
0.4	0.00281	0.5329	0.00027	2.531	208.917	122.608	122.893	19.891	0.5882	0.1622
0.5	0.00104	1.2150	0.00036	2.367	285.233	84.626	135.755	18.859	0.4759	0.2228
0.6	0.00285	0.5253	0.00028	2.511	207.847	122.908	121.491	19.884	0.5845	0.1617
0.7	0.00282	0.5277	0.00028	2.503	209.051	123.322	121.277	19.824	0.5801	0.1607
0.8	0.00112	1.1412	0.00041	2.313	284.037	90.723	130.889	18.988	0.4608	0.2093
0.9	0.00271	0.5458	0.00029	2.496	213.458	123.669	122.077	19.741	0.5719	0.1596
1	0.00263	0.5514	0.00030	2.488	219.303	126.339	123.241	19.715	0.5619	0.1560
0.37	0.00279	0.5378	0.00026	2.537	209.275	122.221	123.529	19.946	0.5902	0.1632

Table 8.5.2: Estimated errors for the BDC experiment

β	0	0.1	0.2	0.3	0.37	0.4	0.6	0.7	0.8	0.9
est. error	0.1051	0.1049	0.1047	0.1044	0.1043	0.1051	0.1052	0.1050	0.1040	0.1048

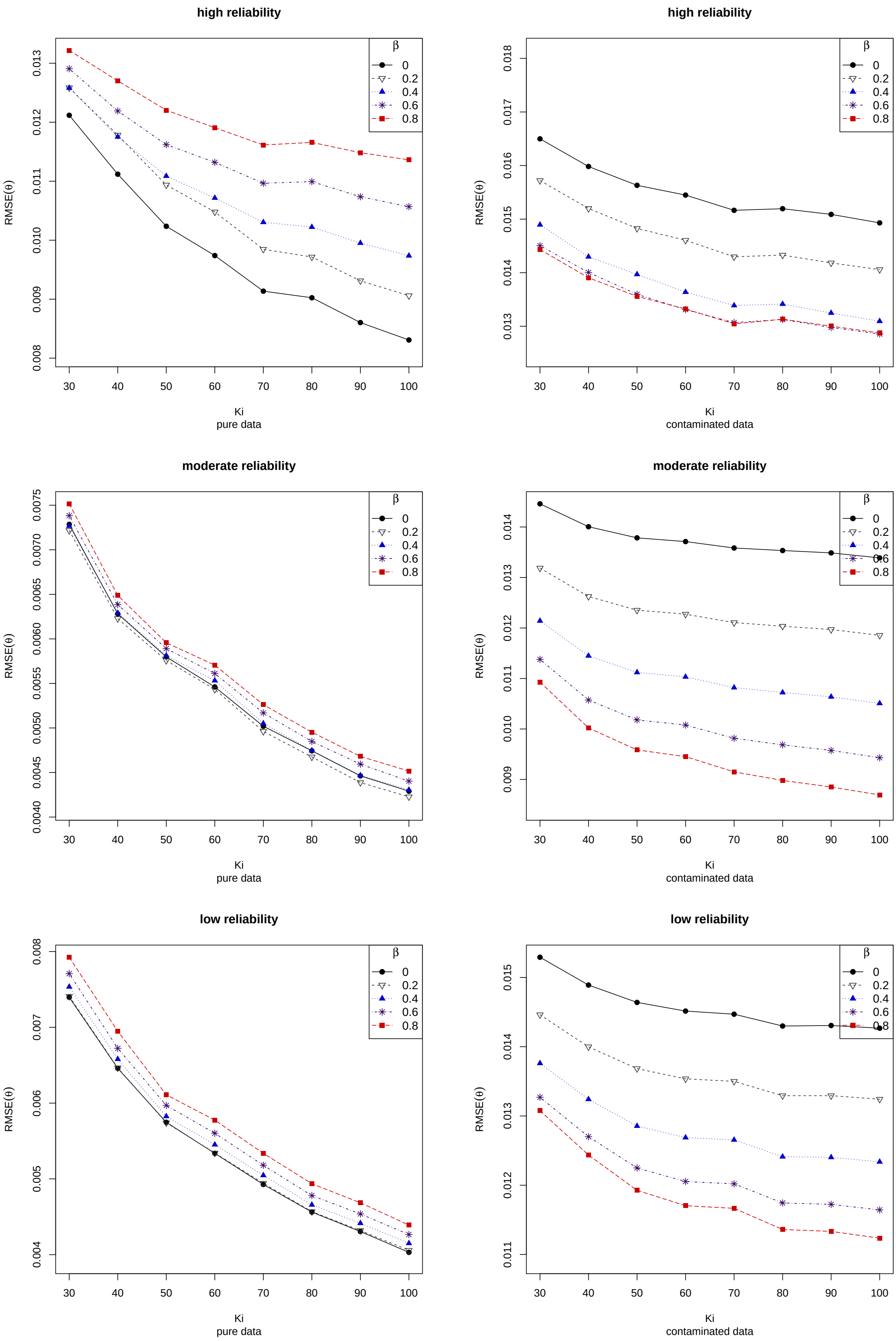

Figure 8.5.1: RMSEs of the weighted minimum DPD estimators of $\boldsymbol{\theta}$ for different values of reliability with pure (left) and contaminated data (right)

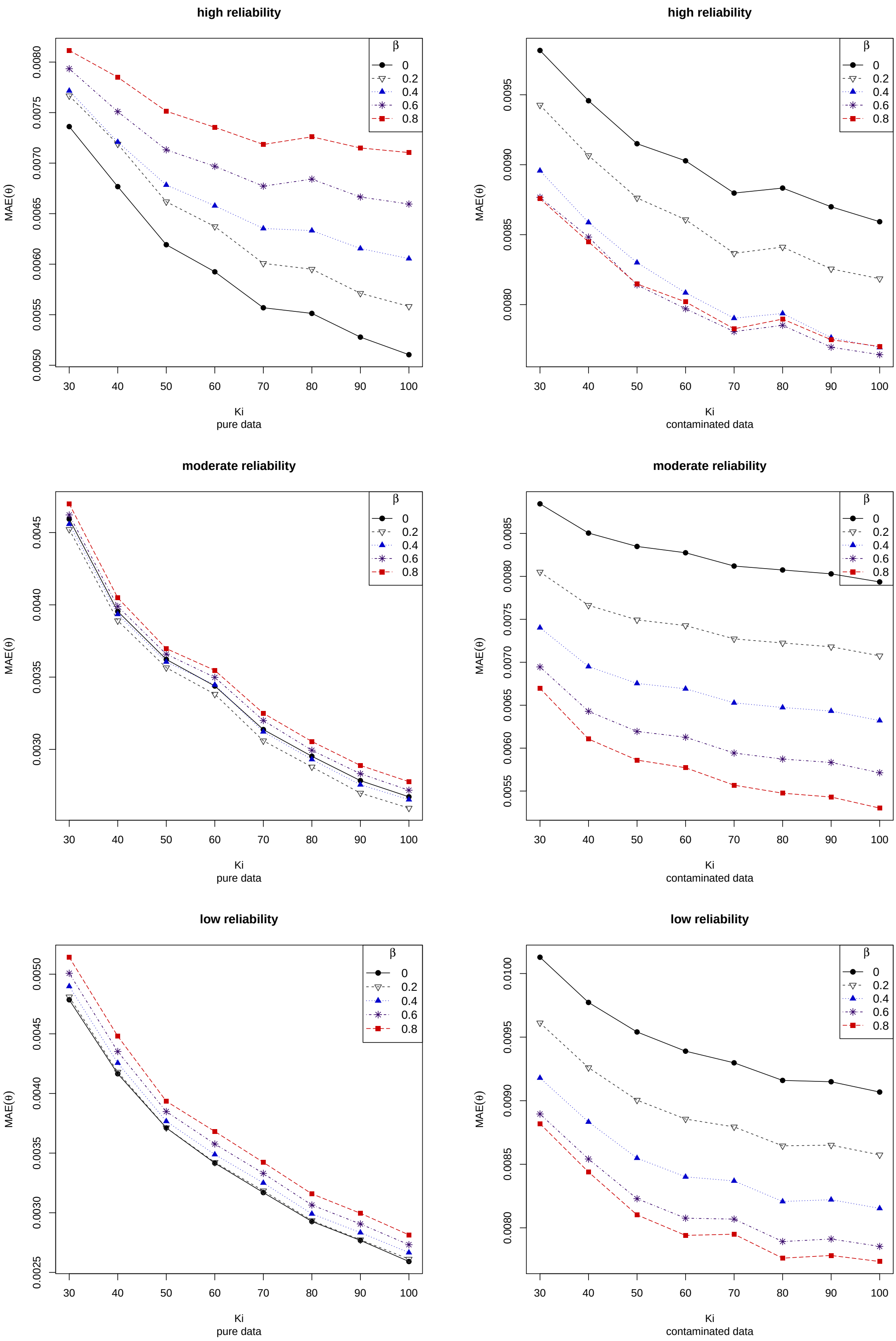

Figure 8.5.2: MAEs of the weighted minimum DPD estimators of $\boldsymbol{\theta}$ for different values of reliability with pure (left) and contaminated data (right).

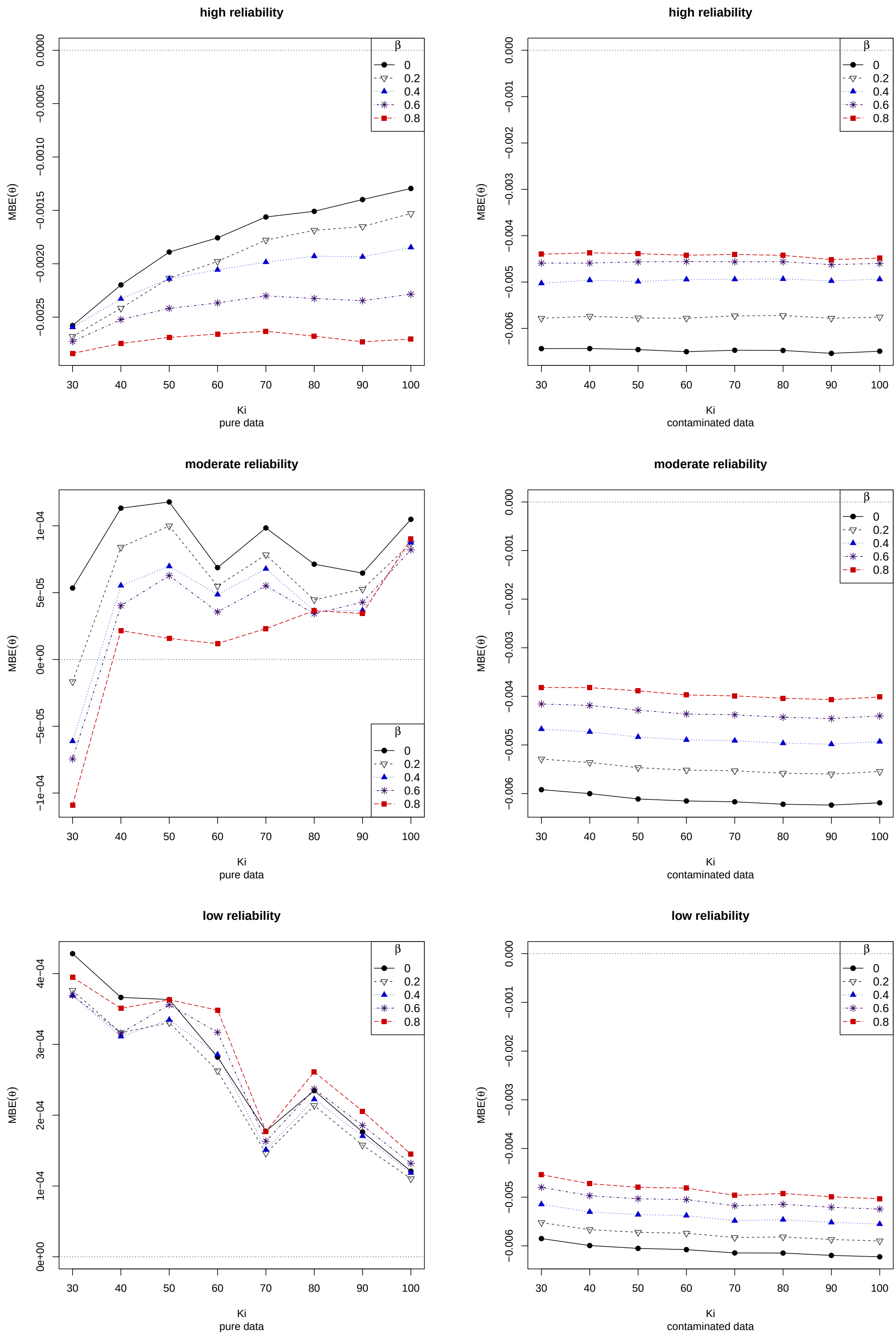

Figure 8.5.3: MBEs of the weighted minimum DPD estimators of $\boldsymbol{\theta}$ for different values of reliability with pure (left) and contaminated data (right).

Chapter 9

Conclusions and further work

9.1 Notes and Comments

In this Thesis, an overview on divergence-based robust methodology for one-shot device testing is done. The development of this work followed different phases.

First of all, we developed robust inference for one-shot device testing under exponential lifetimes and one single stress factor in a non-competing risks setting (see Chapter 2). Despite the apparent simplicity of this model, this was the first and necessary step in order to develop a complete divergence-based robust theory on one-shot device testing. The choice of DPD as our divergence candidate, the computation of the asymptotic distribution of the resulted estimators, or the study of the influence function in this non-homogeneous setup were some of the major challenges we faced. Boundedness of the IF, accompanied with an illustrative simulation study, were not only excellent results, but also a motivation to continue with this research. In this point, it may be important to note that in the original study, presented in the corresponding paper of Balakrishnan et al. [2019b], a quite different notation to that used in Chapter 2 was used. We considered a balanced setting, with same number of observations in each condition, let say K. On the other hand, as only one stress factor was considered, we could see our data as a $I \times J$ contingency table, in which at each time, IT_i, $i, j = 1, 2, ..., I$, K devices are placed under temperatures x_j, $j = 1, \ldots, J$. At each combination of temperature and inspection time, n_{ij} failures were observed. Then, the likelihood function, based on the observed data was given by

$$\mathcal{L}(n_{11}, \ldots, n_{IJ}; \boldsymbol{\theta}) = \prod_{i=1}^{I} \prod_{j=1}^{J} F^{n_{ij}} (IT_i; x_j, \boldsymbol{\theta}) \, R^{K-n_{ij}} (IT_i; x_j, \boldsymbol{\theta}) \, .$$

In this case, we defined the minimum DPD estimator as the minimizer of

$$\begin{aligned} d_\beta(\widehat{\boldsymbol{p}}, \boldsymbol{\pi}(\boldsymbol{\theta})) = \frac{1}{(IJ)^{\beta+1}} \sum_{i=1}^{I} \sum_{j=1}^{J} \Bigg[& F^{\beta+1} (IT_i; x_j, \boldsymbol{\theta}) + R^{\beta+1} (IT_i; x_j, \boldsymbol{\theta}) \\ & - \frac{\beta+1}{\beta} \left(\frac{n_{ij}}{K} F^{\beta} (IT_i; x_j, \boldsymbol{\theta}) + \frac{K - n_{ij}}{K} R^{\beta} (IT_i; x_j, \boldsymbol{\theta}) \right) + \frac{1}{\beta} \left[\left(\frac{n_{ij}}{K} \right)^{\beta+1} + \left(\frac{K - n_{ij}}{K} \right)^{\beta+1} \right] \Bigg]. \end{aligned} \tag{9.1}$$

The following step was to extend the model considered in Chapter 2 to the case of multiple stress factors. This was done in Chapter 3. The first difficulty found here was the formulation of the problem, as the introduction of more stress factors and the possibility of unbalanced data did not allow the original formulation in (9.1). Once this problem was solved with the introduction of the weighted minimum DPD estimators more general results, which includes the single-stress setup as a particular case, were developed. In this case, we could not talk any more about Z-type tests, but Wald-type tests with asymptotically chi-square distribution instead of normal distribution.

Once this extension to the multiple-stress factors setting was completed, it was necessary to consider other more realistic distributions for the lifetimes, although computation of estimating equations and asymptotic variances became more complicated. For example, gamma and Weibull distributions, presented in Chapter 4 and Chapter 5, respectively, or Lindley and Lognormal distributions, studied in Chapter 6. Finally, we extended our robust methods to develop robust estimators and tests for one-shot device testing based on divergence measures under proportional hazards model and competing risks model, in Chapter 7 and Chapter 8, respectively. However, many problems remain still open. We present some of them in the following section.

9.2 Some challenges

9.2.1 On the choice of the tuning parameter

Along this Thesis, we have discussed the problem of choosing the optimal tuning parameter given a data set. Different procedures are discussed for this purpose, all of them based on the following idea: in a grid of possible tuning parameters, apply a measure of discrepancy to the data. Then, the tuning parameter that leads to the minimum discrepancy-statistic can be chosen as the "optimal" one. A possible choice of the discrepancy measure could be M_β, given in (7.26). Another idea may be by minimizing the estimated mean square error, as suggested in Warwick and Jones [2005]. However, as noted in Section 7.6.3, the need for a pilot estimator became the major drawback of this procedure, as the final result will depend excessively on this choice. This problem was also highlighted recently in Basak et al. [2020], where an "iterative Warwick and Jones algorithm" (IWJ algorithm) is proposed. Application of the IJW algorithm and other possible approaches is an interesting issue to be faced in the future.

9.2.2 Robust inference for one-shot devices with competing risks under gamma or Weibull distribution

In Chapter 8, robust inference for one-shot device testing under competing risks is developed, under the assumption of exponential lifetime distribution. However, the competing risks model has been also considered in literature under other distributions. For example, in Balakrishnan et al. [2015c], an expectation maximization (EM) algorithm is developed for the estimation of model parameters with Weibull lifetime distribution. For further work, we can also develop robust inference for one-shot devices with competing risks under gamma or Weibull distributions.

For example, let us consider the setting described in Table 8.2.1, limiting, for simplicity, the number of competing causes to $R = 2$. Let us denote the random variable for the failure time due to causes 1 and 2 as T_{irk}, for $r = 1, 2$, $i = 1, \dots, I$, and $k = 1, \dots, K_i$, respectively. We now assume that T_{irk} follows a Weibull distribution with scale parameter α_{ir} and shape parameter η_{ir}, with probability and cumulative density functions as

$$
\begin{aligned}
f_{T_r}(t; \boldsymbol{x}_i, \boldsymbol{\theta}) &= \left(\frac{\eta_{ir}}{\alpha_{ir}(\boldsymbol{\theta})}\right)\left(\frac{t}{\alpha_{ir}}\right)^{\eta_{ir}-1} \exp\left(-\left(\frac{1}{\alpha_{ir}}\right)^{\eta_{ir}(\boldsymbol{\theta})}\right), \quad t > 0, \\
F_{T_r}(t; \boldsymbol{x}_i, \boldsymbol{\theta}) &= 1 - \exp\left(-\left(\frac{1}{\alpha_{ir}}\right)^{\eta_{ir}}\right), \quad t > 0,
\end{aligned}
$$

respectively, where $\boldsymbol{x}_i$ is the stress factor of the condition i, related to shape and scale parameters by a log-link function

$$
\alpha_{ir} \equiv \alpha_{ir}(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_r \boldsymbol{x}_i),
$$

$$\eta_{ir} \equiv \eta_{ir}(\boldsymbol{\theta}) = \exp(\boldsymbol{b}_r \boldsymbol{x}_i),$$

where $\boldsymbol{a}_r = (a_{r0}, a_{r1}, \ldots, a_{rJ})$, $\boldsymbol{b}_r = (b_{r0}, b_{r1}, \ldots, b_{rJ})$ and $\boldsymbol{\theta} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{b}_1, \boldsymbol{b}_2)^T \in \mathbb{R}^{4(J+1)}$ is the model parameter vector. As it happened in the non-competing-risks model (Chapter 5), instead of working with Weibull lifetimes, it is more convenient to work with the log-transformed lifetime $W_{irk} = log(T_{irk})$, which follows an extreme value (Gumbel) distribution (see Meeker et al. [1998]). The corresponding probability and cumulative density functions of the extreme value distribution are

$$\begin{aligned}
f_{W_r}(t; \boldsymbol{x}_i, \boldsymbol{\theta}) &= \left(\frac{\eta_{ir}}{\alpha_{ir}(\boldsymbol{\theta})}\right)\left(\frac{t}{\alpha_{ir}(\boldsymbol{\theta})}\right)^{\eta_{ir}(\boldsymbol{\theta})-1} \exp\left(-\left(\frac{1}{\alpha_{ir}(\boldsymbol{\theta})}\right)^{\eta_{ir}(\boldsymbol{\theta})}\right), \quad t > 0, \\
F_{W_r}(t; \boldsymbol{x}_i, \boldsymbol{\theta}) &= 1 - \exp\left(-\exp\left(\frac{\omega - \mu_{ir}}{\sigma_{ir}}\right)\right), \quad t > 0,
\end{aligned}$$

respectively, where $-\infty < \omega < \infty$, $\mu_{ir} = log(\alpha_{ir})$ and $\sigma_{ir} = \eta_{ir}^{-1}$. We define correspondingly the log-transformed inspection times $lIT_i = log(IT_i)$. We shall use $\pi_{i0}(\boldsymbol{\theta})$, $\pi_{i1}(\boldsymbol{\theta})$ and $\pi_{i2}(\boldsymbol{\theta})$ for the survival probability, failure probability due to cause 1 and failure probability due to cause 2, respectively. Their expressions are

$$\begin{aligned}
\pi_{i0}(\boldsymbol{\theta}) &= \exp\left(-\exp\left(\frac{lIT_i - \mu_{i1}}{\sigma_{i1}}\right) - \exp\left(\frac{lIT_i - \mu_{i2}}{\sigma_{i2}}\right)\right), \\
\pi_{i1}(\boldsymbol{\theta}) &= \int_{-\infty}^{lIT_i} \exp\left\{-\exp\left(\frac{\omega - \mu_{i1}}{\sigma_{i1}}\right) - \exp\left(\frac{\omega - \mu_{i2}}{\sigma_{i2}}\right)\right\} \exp\left(\frac{\omega - \mu_{i1}}{\sigma_{i1}}\right) \frac{1}{\sigma_{i1}} d\omega, \\
\pi_{i2}(\boldsymbol{\theta}) &= \int_{-\infty}^{lIT_i} \exp\left\{-\exp\left(\frac{\omega - \mu_{i1}}{\sigma_{i1}}\right) - \exp\left(\frac{\omega - \mu_{i2}}{\sigma_{i2}}\right)\right\} \exp\left(\frac{\omega - \mu_{i2}}{\sigma_{i2}}\right) \frac{1}{\sigma_{i2}} d\omega,
\end{aligned}$$

Then, the likelihood function is given by

$$\mathcal{L}(n_{01}, \ldots, n_{I2}; \boldsymbol{\theta}) \propto \prod_{i=1}^{I} \pi_{i0}(\boldsymbol{\theta})^{n_{i0}} \pi_{i1}(\boldsymbol{\theta})^{n_{i1}} \pi_{i2}(\boldsymbol{\theta})^{n_{i2}}, \tag{9.2}$$

where $n_{0i} + n_{1i} + n_{2i} = K_i,\ i = 1, \ldots, I$; and MLE is obtained by minimizing 9.2 on $\boldsymbol{\theta}$. Following the same spirit of previous chapters, we can define the weighted minimum DPD estimator of $\boldsymbol{\theta}$ as

$$\widehat{\boldsymbol{\theta}}_\beta = \arg\min_{\boldsymbol{\theta} \in \Theta} {}^* d_\beta^W(\boldsymbol{\theta}), \quad \text{for } \beta > 0,$$

where ${}^* d_\beta^W(\boldsymbol{\theta})$ is as in (8.5). For $\beta = 0$, we have the MLE. Estimating equations and asymptotic distribution of proposed estimators may need to be obtained. We are currently working on this problem and hope to report the findings in a future paper.

9.2.3 EM algorithm for one-shot device testing under the lognormal distribution

Chapter 6 deals with the problem of one-shot device testing under lognormal distribution, in particular, new estimators and tests are proposed based on divergence measures and are shown to present a better behaviour than classical MLE in terms of robustness. However, up to our knowledge, no previous literature was done in relation of one-shot device testing under log-normal distribution. It would be of interest to develop an Expectation Maximization (EM) algorithm for the estimation of the MLE in this context.

EM algorithm, (Dempster et al. [1977]), is a very popular tool to handle any missing or incomplete data situation. This iterative method has two steps. In the E-step, it replaces any missing data by its expected value and in the M-step the log-likelihood function is maximized with the observed data and expected value of the incomplete data, producing an update of the parameter estimates. The MLEs of the parameters are obtained by repeating the E- and M-steps until convergence occurs. Note that, in the case of lognormal lifetimes, it would be very helpful to work with the logarithm of lifetimes, so in the E-step, which will need of the conditional expectation of the log-likelihood of complete data, we could use the left-truncated and right-truncated normal distributions (see Basak et al. [2009] for progressively censored data).

9.2.4 Model selection in one-shot devices by means of the generalized gamma distribution

Let us assume, without loss of generality, the multiple stress factors setting presented in Table 2.2.1. We may assume that the lifetimes follow a generalized gamma distribution, $f(t, x_i; q, \lambda_i, \sigma_i)$, where $t > 0$ is the lifetime, $-\infty < q < \infty$ and $\sigma_i > 0$ are shape parameters and $\lambda_i > 0$ is a scale parameter, related to the stress level x_i in a log-linear form. The generalized gamma distribution has been widely studied in recent years because its flexibility. It contains many distributions as special cases. For example, is the lognormal distribution when $q = 0$, the Weibull distribution when $q = 1$ and the gamma distribution when $q/\sigma_i = 1$. Other well known probability distributions, such as the half-normal or spherical normal distributions, can be obtained as special cases. For more details, see Stacy and Mihram [1965] and Balakrishnan and Peng [2006].

The advantage of using the generalized gamma distribution as the lifetime distribution is also demonstrate by the flexible tails of its density function, which can determine the type of dependence among the correlated observations (see Hougaard [1986]). However, the generalized gamma distribution has not been considered so far for one-shot device models. Model discrimination within the generalized gamma distribution, by means of information-based criteria, likelihood ratio tests or Wald tests, will be a challenging and interesting problem for further consideration.

9.3 Productions

During the PhD a total of 16 manuscripts have been produced, 13 of which are already accepted in JCR impact factor journals, while the others are currently under revision. Furthermore, 2 chapters of book have been also published. These are numerated as follows, following the order they appear in the Thesis.[1]

Articles published in JCR journals

1. Balakrishnan, N., Castilla, E., Martín N. and Pardo, L. (2019). *Robust estimators and test-statistics for one-shot device testing under the exponential distribution.* IEEE transactions on Information Theory. 65(5), pp. 3080-3096.

2. Balakrishnan, N., Castilla, E., Martín N. and Pardo, L. (2020). *Robust inference for one-shot device testing data under exponential lifetime model with multiple stresses.* Quality and Reliability Engineering International. 36, pp. 1916-1930.

3. Balakrishnan, N., Castilla, E., Martín N. and Pardo, L. (2019).*Robust estimators for one-shot device testing data under gamma lifetime model with an application to a tumor toxicological data.* Metrika. 82(8), pp. 991–1019.

4. Balakrishnan, N., Castilla, E., Martín N. and Pardo, L. (2019). *Robust inference for one-shot device testing data under Weibull lifetime model.* IEEE transactions on Reliability. 69(3), pp. 937-953.

[1] Last updated version: May 2021

5. Balakrishnan, N., Castilla, E., Martín N. and Pardo, L. (2021). *Divergence-based robust inference under proportional hazards model for one-shot device testing.* IEEE transactions on Reliability. DOI: 10.1109/TR.2021.3062289.

6. Castilla, E., Martín N., Muñoz S. and Pardo, L. (2020). *Robust Wald-type tests based on minimum Rényi pseudodistance estimators for the multiple regression model.* Journal of Statistical Computation and Simulation. 90(14), pp. 2655-2680.

7. Castilla, E., Martin, N. and Pardo, L. (2018). *Pseudo minimum phi-divergence estimator for the multinomial logistic regression model with complex sample design.* AStA Adv. Stat. Anal., 102(3), pp. 381-411.

8. Castilla, E., Ghosh, A., Martín, N. and Pardo, L. (2018). *New statistical robust procedures for polytomous logistic regression models.* Biometrics, 74(4), pp. 1282-1291.

9. Castilla, E., Martín, N. and Pardo, L. (2020). *Testing linear hypotheses in logistic regression analysis with complex sample survey data based on phi-divergence measures.* Communications in Statistics-Theory and Methods. DOI: 10.1080/03610926.2020.1746342.

10. Castilla, E., Ghosh, A., Martín, N. and Pardo, L. (2020). *Robust semiparametric inference for polytomous logistic regression with complex survey design.* Advances in Data Analysis and Classification . DOI: 10.1007/s11634-020-00430-7.

11. Castilla, E., Martín, N., Pardo, L. and Zografos, K. (2018). *Composite likelihood methods based on minimum density power divergence estimator.* Entropy 20(1), 18.

12. Castilla, E., Martín, N., Pardo, L. and Zografos, K. (2019). *Composite likelihood methods: Rao-type tests based on composite minimum density power divergence estimator.* Statistical Papers. 62, pp. 1003-1041.

13. Castilla, E., Martín, N., Pardo, L. and Zografos, K. (2020). *Model Selection in a composite likelihood framework based on density power divergence.* Entropy. 22(3), 270.

Book Chapters

1. Balakrishnan, N., Castilla, E. and Pardo, L. (2021). *Robust statistical inference for one-shot devices based on density power divergences: An overview.* In Arnold, B.C, Balakrishnan, N. and Coelho, C. (eds) Contributions to Statistical Distribution Theory and Inference. Festschrift in Honor of C. R. Rao on the Occasion of His 100th Birthday. Springer, New York. (Accepted).

2. Castilla, E., Martin, N. and Pardo, L. (2018). *A Logistic Regression Analysis approach for sample survey data based on phi-divergence measures.* In: Gil E., Gil E., Gil J., Gil M. (eds) The Mathematics of the Uncertain. Studies in Systems, Decision and Control, vol 142. Springer, Cham, pp 465-474.

Other articles submitted for publication

1. Balakrishnan, N., Castilla, E. and Ling, M.H. *Optimal designs of constant-stress accelerated life-tests for one-shot devices with model mis-specification analysis.*

2. Balakrishnan, N., Castilla, E., Martín N. and Pardo, L. *Power divergence approach for one-shot device testing under competing risks.* arxiv:2004.13372.

3. Castilla, E. and Chocano, P.J. *A new robust approach for multinomial logistic regression with complex design model.* arxiv:2102.03073.

Appendix A

Optimal design of CSALTs for one-shot devices and the effect of model misspecification

Along this Thesis, we have focused our efforts on developing robust inference for one-shot device testing by means of divergence measures. So far, however, we have little discussed about optimal design, which is another problem of great importance in reliability, as it would result in great savings in both time and cost. Note that there are many types of ALTs. For example, constant-stress ALTs (CSALTs) assume that each device is subject to only one pre-specified stress level, while step-stress ALTs (SSALTs) apply stress to devices in such a way that stress levels will get changed at prespecified times. To design efficient CSALTs for one-shot devices under Weibull lifetime distribution, subject to a prespecified budget and a termination time, Balakrishnan and Ling [2014b] considered the minimization of the asymptotic variance of the MLE of reliability at a mission time under normal operating conditions. In a similar manner, Ling [2019] and Ling and Hu [2020] designed optimal SSALTs for one-shot devices under exponential and Weibull lifetime distributions, respectively. In this Appendix, we briefly present the problem of optimal design of CSALTs in one-shot device testing. This problem, as well as the effect of model misspecification, have been extensively studied in Balakrishnan et al. [2020a].

Let us suppose, that the data are stratified into I testing conditions $S_1, \dots, S_I$, and that in testing condition S_i, N_i individuals are tested with J types of stress factors being maintained at certain levels, and inspected at K_i equally-spaced time points. Specifically, N_{ik} items are drawn and inspected at a specific time T_{ik} with $\sum_{k=1}^{K_i} N_{ik} = N_i$. Then, n_{ik} failure items are collected from the test at inspection time T_{ik}. Let $\boldsymbol{x}_i = (1, x_{i1}, \dots, x_{iJ})^T$ be the vector of stress factors associated to testing condition S_i $(i = 1, \dots, I)$. The MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, is then determined by maximizing the log-likelihood function of the data, with respect to the model parameter $\boldsymbol{\theta}$.

We want to describe an algorithm for the determination of the best ALT plan, by minimizing the asymptotic variance of the MLE of reliability at a specific mission time under normal operating conditions. Then, we would need the Fisher information matrix for model parameters. Let us consider the inspection plan $\boldsymbol{\zeta} = (f, K_i, N_{ik})$, consisting of inspection frequency, number of inspections at each condition, and allocation of the products. The Fisher information matrix under $\boldsymbol{\zeta}$, is given by

$$\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{\zeta}) = \sum_{i=1}^{I} \sum_{k=1}^{K_i} N_{ik} \left(\frac{1}{R(T_{ik}; S_i)} + \frac{1}{1 - R(T_{ik}; S_i)} \right) \left(\frac{\partial R(T_{ik}; S_i)}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial R(T_{ik}; S_i)}{\partial \boldsymbol{\theta}^T} \right),$$

where $T_{ik} = k \times f$, for $k = 1, \dots, K_i$, for equi-spaced time points, and $R(T_{ik}; S_i)$ denotes the reliability function. The asymptotic covariance matrix of the MLEs of the model parameters can be obtained by inverting the observed Fisher information matrix.

$$\boldsymbol{V} \equiv \boldsymbol{V}(\boldsymbol{\theta}; \boldsymbol{\zeta}) = (\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{\zeta}))^{-1}.$$

Using these expressions, the asymptotic variance of the MLE of reliability under normal operating conditions at a specific mission time t_0 can be computed by the delta method

$$V_R(\boldsymbol{\zeta}) \equiv AV(\widehat{R}(t_0; \boldsymbol{x}_0)) = \boldsymbol{P}_R^T \boldsymbol{V} \boldsymbol{P}_R,$$

where $\boldsymbol{P}_R = \frac{\partial R(t_0;\boldsymbol{x}_0)}{\partial \boldsymbol{\theta}}\Big|_{\widehat{\boldsymbol{\theta}}}$, and $\boldsymbol{x}_0$ is the vector of stress factors associated to the normal operating condition.

Suppose the budget for conducting a CSALT for one-shot device testing, the operation cost at testing condition S_i, the cost of devices (including the purchase of and testing cost), and the termination time are specified as C_{budget}, $C_{oper,i}$, C_{item} and T_{ter}, respectively. Then, for a given test plan, $\boldsymbol{\zeta}$, that includes the inspection frequency, f, the number of inspections at testing condition S_i, $K_i \geq 2$, and the allocation of devices, N_{ik}, for $i = 1, 2, ..., I$, the total cost of conducting the experiment is seen to be

$$TC(\boldsymbol{\zeta}) = C_{item} \sum_{i=1}^{I} \sum_{k=1}^{K_i} N_{ik} + f \left(\sum_{i=1}^{I} K_i C_{oper,i} \right).$$

In Balakrishnan et al. [2020a], an algorithm for the determination of an optimal CSALT subject to a specified budget ($TC(\boldsymbol{\zeta}) \leq C_{budget}$) and termination time is presented, by minimizing the asymptotic variance of the MLE of reliability; and applied to the case that the lifetimes of the devices follow a gamma or a Weibull distribution. In an extensive simulation study, this algorithm is evaluated, as well as its sensitivity over parameter misspecification. It is seen that, within moderate errors of the parameters, the designs of optimal CSALTs are quite robust. In this paper, the effect of model misspecification between gamma, Weibull, lognormal and BS distributions in the design of optimal CSALTs is also examined. Results do reveal that the assumption of lifetime distribution to be Weibull seems to be the more robust to model misspecification, while the assumption of lifetime distribution to be gamma seems to be the more non-robust or more sensitive.

Appendix B

Robust Inference for some other Statistical Models based on Divergences

This appendix briefly presents a series of results in the area of robust statistical information theory, which have also been obtained by the candidate during her Ph.D. studies. Section B.1 summarizes the results given in Castilla et al. [2020d]. Section B.2 deals with the diveregnce-based estimators in the logistic regression model. See Castilla et al. [2018a,b,c, 2020c] and Castilla and Chocano [2020]. Finally, Section B.3 contains three results (Castilla et al. [2018d, 2019, 2020b]) related with composite likelihood methods.

B.1 Multiple Linear Regression model

The multiple regression model (MRM) is one of the most known statistical models. Mathematically, let $(X_{i1},...,X_{ip},Y_i),\ \ i=1,...,n,$ be $(p+1)$-dimensional independent and identically distributed random variables verifying the condition

$$Y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + \varepsilon_i, \tag{B.1}$$

with $\boldsymbol{X}_i^T=(X_{i1},...,X_{ip})$ and $\boldsymbol{\beta}=(\beta_1,...,\beta_p)^T$ and $\varepsilon_i's$ are i.i.d. normal random variables with mean zero and variance σ^2 and independent of the $\boldsymbol{X}_i$. The $n\times p$ matrix with elements X_{ij} will be denoted by $\mathbb{X}$, i.e., $\mathbb{X}=(\boldsymbol{X}_1,...,\boldsymbol{X}_n)^T$. We can use the matrix and vector notation

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{B.2}$$

with $\boldsymbol{Y}=(Y_1,\dots,Y_n)^T$ and $\boldsymbol{\varepsilon}=(\varepsilon_1,\dots,\varepsilon_n)^T$.

As we have seen along the development of this Thesis, minimum distance estimators have been presented in different statistical models as an alternative to the classical MLE, which is known to have good efficiency properties, but not so good robustness properties. With this motivation, Durio and Isaia [2011] studied the minimum DPD estimators for the MRM. In the cited paper, the robustness of DPD estimators was analyzed from a simulation study, with no theoretical support. Minimum DPD estimators have also been used in order to define Wald-type tests as, for example, in Basu et al. [2016] and Ghosh et al. [2016]. Broniatowski et al. [2012] considered the RP in order to give robust estimators, minimum RP estimators, for the MRM.

Let $X_1,\dots,X_n$ be a random sample from a population having true density g which is being modeled by a parametric family of densities $f_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbb{R}^p$. The RP between the densities g and $f_{\boldsymbol{\theta}}$ is given by

$$R_\alpha\left(g,f_{\boldsymbol{\theta}}\right)=\frac{1}{\alpha+1}\log\left(\int f_{\boldsymbol{\theta}}(x)^{\alpha+1}dx\right)+\frac{1}{\alpha\left(\alpha+1\right)}\log\left(\int g(x)^{\alpha+1}dx\right)-\frac{1}{\alpha}\log\left(\int f_{\boldsymbol{\theta}}(x)^{\alpha}g(x)dx\right)$$

for $\alpha > 0$, whereas for $\alpha = 0$ it is given by

$$R_0(g, f_{\boldsymbol{\theta}}) = \lim_{\alpha \downarrow 0} R_\alpha(g, f_{\boldsymbol{\theta}}) = \int g(x) \log \frac{g(x)}{f_\theta(x)} dx,$$

i.e., the Kullback-Leibler divergence, $D_{Kullback}(g, f_{\boldsymbol{\theta}})$, between g and $f_{\boldsymbol{\theta}}$ (see Pardo [2005]). In Broniatowski et al. [2012] it was established that $R_\alpha(g, f_{\boldsymbol{\theta}}) \geq 0$, with $R_\alpha(g, f_{\boldsymbol{\theta}}) = 0$ if and only if $f_{\boldsymbol{\theta}} = g$.

The minimum RP estimator is obtained by minimizing the RP, $R_\alpha(\widehat{g}, f_{\boldsymbol{\theta}})$, with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\widehat{g}$ is an empirical estimator of g based on the available data.

In Castilla et al. [2020d], a new family of Wald-type tests was introduced, based on minimum Rényi pseudodistance estimators, for testing general linear hypotheses and the variance of the residuals in the multiple regression model. The classical Wald test, based on the maximum likelihood estimator, can be seen as a particular case inside this family. Theoretical results, supported by an extensive simulation study, point out how some tests included in this family have a better behaviour, in the sense of robustness, than the Wald test.

B.2 Multinomial Logistic Regression model

The multinomial logistic regression model, also known as polytomous logistic regression model (PLRM) is widely used in health and life sciences for analyzing nominal qualitative response variables (e.g., Daniels and Gatsonis [1997], Bertens et al. [2016], Dey and Raheem [2016] and the references therein). Such examples occur frequently in medical studies where disease symptoms may be classified as absent, mild or severe, the invasiveness of a tumor may be classified as in situ, locally invasive, or metastatic, etc. The qualitative response models specify the multinomial distribution for such a response variable with individual category probabilities being modeled as a function of suitable explanatory variables. One such popular model is the PLRM, where the logit function is used to link the category probabilities with the explanatory variables.

Mathematically, let us assume that the nominal outcome variable $\tilde{Y}$ has $d+1$ categories $C_1, ..., C_{d+1}$ and we observe $\tilde{Y}$ together with k explanatory variables with given values x_h, $h = 1, ..., k$. In addition, assume that $\boldsymbol{\beta}_j^T = (\beta_{0j}, \beta_{1j}, ..., \beta_{kj})$, $j = 1, ..., d$, is a vector of unknown parameters and $\boldsymbol{\beta}_{d+1}$ is a $(k+1)$-dimensional vector of zeros; i.e., the last category C_{d+1} has been chosen as the baseline category. Since the full parameter vector $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_d^T)$ is ν-dimensional with $\nu = d(k+1)$, the parameter space is $\Theta = \mathbb{R}^{d(k+1)}$. Let

$$\pi_j(\boldsymbol{x}, \boldsymbol{\beta}) = P(\tilde{Y} \in C_j \mid \boldsymbol{x}, \boldsymbol{\beta})$$

denote the probability that $\tilde{Y}$ belongs to the category C_j for $j = 1, ..., d+1$, when the vector of explanatory variable takes the value $\boldsymbol{x}^T = (x_0, x_1, \ldots, x_k)$, with $x_0 = 1$ being associated with the intercept β_{0j}. Then, the PLRM is given by

$$\pi_j(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}^T \boldsymbol{\beta}_j)}{1 + \sum_{h=1}^{d} \exp(\boldsymbol{x}^T \boldsymbol{\beta}_h)}, \quad j = 1, ..., d+1. \tag{B.3}$$

Now assume that we have observed the data on N individuals having responses $\tilde{y}_i$ with associated covariate values (including intercept) $\boldsymbol{x}_i \in \mathbb{R}^{k+1}$, $i = 1, ..., N$, respectively. For each individual, let us introduce the corresponding tabulated response $\boldsymbol{y}_i = (y_{i1}, ..., y_{i,d+1})^T$ with $y_{ir} = 1$ and $y_{is} = 0$ for $s \in \{1, ..., d+1\} - \{r\}$ if $\tilde{y}_i \in C_r$.

The most common estimator of $\boldsymbol{\beta}$ under the PLRM is the MLE, which is obtained by maximizing the loglikelihood function,

$$\log \mathcal{L}(\boldsymbol{\beta}) \equiv \sum_{i=1}^{N} \sum_{j=1}^{d+1} y_{ij} \log \pi_j(\boldsymbol{x}_i, \boldsymbol{\beta}).$$

One can then develop all the subsequent inference procedures based on the MLE $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

In Castilla et al. [2018a], a new family of estimators is defined as a generalization of the MLE for the PLRM. Based on these estimators, a family of Wald-type test statistics for linear hypotheses is introduced. Robustness properties of both the proposed estimators and the test statistics are theoretically studied through the classical influence function analysis and illustrated by real life examples and an extensive simulation study.

Note that in Castilla et al. [2018c] and Castilla et al. [2020c] new estimators and Wald-type tests are developed, in the context of Logistic Regression analysis with complex sample survey data, based on phi-divergence measures.

B.2.1 Robust inference for the multinomial logistic regression model with complex sample design based on divergence measures

In many practical applications, we come across data which have been collected through a complex survey scheme, like stratified sampling or cluster sampling, etc., rather than the simple random sampling. Such situations are quite common in large scale data collection, for example, within several states of a country or even among different countries. Suitable statistical methods are required to analyze these data by taking care of the stratified structure of the data; this is because there often exist several inter and intra-class correlations within such stratification and ignoring then may often lead to erroneous inference. Further, in many such complex surveys, stratified observations are collected on some categorical responses having two or more mutually exclusive unordered categories along with some related covariates and inference about their relationship is of up-most interest for insight generation and policy making. Polytomous logistic regression (PLR) model is a useful and popular tool in such situations to model categorical responses with associated covariates. However, most of classical literature deal with the cases of simple random sampling scheme. (e.g. McCullagh [1980], Agresti [2002]). The application of PLR model under complex survey setting can be found, for example, in Binder [1983], Roberts et al. [1987], Morel [1989] and Castilla et al. [2018b]; most of them, except the last one, are based on the quasi maximum likelihood approach.

Let us assume that the whole population is partitioned into H distinct strata and the data consist of n_h clusters in stratum h for each $h = 1, \ldots, H$. Further, for each cluster $i = 1, \ldots, n_h$ in the stratum h, we have observed the values of a categorical response variable (Y) for m_{hi} units. Assuming Y has $(d+1)$ categories, we denote these observed responses by a $(d+1)$-dimensional classification vector

$$\boldsymbol{y}_{hij} = \left(y_{hij1}, ..., y_{hij,d+1}\right)^T, \ h = 1, ..., H, i = 1, ..., n_h, \ j = 1, ..., m_{hi},$$

with $y_{hijr} = 1$ if the j-th unit selected from the i-th cluster of the h-th stratum falls in the r-th category and $y_{hijl} = 0$ for $l \neq r$. It is very common when working with dummy or qualitative explanatory variables to consider that the $k+1$ explanatory variables are common for all the individuals in the i-th cluster of the h-th stratum, being denoted as $\boldsymbol{x}_{hi} = (x_{hi0}, x_{hi1}, ..., x_{hik})^T$, with the first one, $x_{hi0} = 1$, associated with the intercept.

Let us denote the sampling weight from the i-th cluster of the h-th stratum by w_{hi}. For each i, h and j, the expectation of the r-th element of the random variable $\boldsymbol{Y}_{hij} = (Y_{hij1}, ..., Y_{hij,d+1})^T$, corresponding to the realization $\boldsymbol{y}_{hij}$, is determined by

$$\pi_{hir}\left(\boldsymbol{\beta}\right) = \mathrm{E}\left[Y_{hijr}|\boldsymbol{x}_{hi}\right] = \Pr\left(Y_{hijr} = 1|\boldsymbol{x}_{hi}\right) = \begin{cases} \dfrac{\exp\{\boldsymbol{x}_{hi}^T\boldsymbol{\beta}_r\}}{1+\sum_{l=1}^{d}\exp\{\boldsymbol{x}_{hi}^T\boldsymbol{\beta}_l\}}, & r = 1, ..., d \\ \dfrac{1}{1+\sum_{l=1}^{d}\exp\{\boldsymbol{x}_{hi}^T\boldsymbol{\beta}_l\}}, & r = d+1 \end{cases}, \quad \text{(B.4)}$$

with $\boldsymbol{\beta}_r = (\beta_{r0}, \beta_{r1}, ..., \beta_{rk})^T \in \mathbb{R}^{k+1}$, $r = 1, ..., d$ and the associated parameter space given by $\Theta = \mathbb{R}^{d(k+1)}$.

Note that, under homogeneity, the expectation of $\boldsymbol{Y}_{hij}$ does not depend on the unit number j, so from now we will denote by

$$\widehat{\boldsymbol{Y}}_{hi}=\sum_{j=1}^{m_{hi}}\boldsymbol{Y}_{hij}=\left(\sum_{j=1}^{m_{hi}}Y_{hij1},...,\sum_{j=1}^{m_{hi}}Y_{hij,d+1}\right)^T=(\widehat{Y}_{hi1},...,\widehat{Y}_{hi,d+1})^T$$

the random vector of counts in the i-th cluster of the h-th stratum and by $\boldsymbol{\pi}_{hi}\left(\boldsymbol{\beta}\right)$ the $(d+1)$-dimensional probability vector with the elements given in (B.4), $\boldsymbol{\pi}_{hi}\left(\boldsymbol{\beta}\right)=\left(\pi_{hi1}\left(\boldsymbol{\beta}\right),...,\pi_{hi,d+1}\left(\boldsymbol{\beta}\right)\right)^T$.

Even though the quasi weighted maximum likelihood estimator, is the main base of most of the existing literature on logistic models under complex survey designs, it is known to be non-robust with respect to the possible outliers in the data. In practice, with such a complex survey design, it is quite natural to have some outlying observations that make the likelihood based inference highly unstable. So, we often may need to make additional efforts to find and discard the outliers from the data before their analysis. A robust method providing stable solution even in presence of the outliers will be really helpful and more efficient in practice.

The cited work by Castilla et al. [2018b] has developed an alternative minimum divergence estimator based on ϕ-divergences, as well as new estimators for the intra-cluster correlation coefficient. A simulation study shows that the Binder's method for the intra-cluster correlation coefficient exhibits an excellent performance when the pseudo-minimum Cressie–Read divergence estimator (by considering the Cressie-Read family of ϕ-divergences), with $\lambda = 2/3$, is plugged. However, this paper does not lead with the problem of robustness. In Castilla et al. [2020a], the minimum quasi weighted DPD estimators for the multinomial logistic regression model with complex survey. This family of semiparametric estimators is a robust generalization of the maximum quasi likelihood estimator, by using the DPD measure. Their asymptotic distribution and accurate robustness properties are theoretically studied and empirically validated through a numerical example and an extensive Monte Carlo study. Recently, Castilla and Chocano [2020] studied the robustness of negative ϕ-divergences, through the boundedness of the influence function and extensive simulation experiments.

B.3 Composite Likelihood

The classical likelihood function requires exact specification of the probability density function but in most applications the true distribution is unknown. In some cases, where the data distribution is available in an analytic form, the likelihood function is still mathematically intractable due to the complexity of the probability density function. There are many alternatives to the classical likelihood function; one of them is the composite likelihood. Composite likelihood is an inference function derived by multiplying a collection of component likelihoods; the particular collection used is a conditional determined by the context. Therefore, the composite likelihood reduces the computational complexity so that it is possible to deal with large datasets and very complex models even when the use of standard likelihood methods is not feasible. Asymptotic normality of the composite maximum likelihood estimator (CMLE) still holds with Godambe information matrix to replace the expected information in the expression of the asymptotic variance-covariance matrix. This allows the construction of composite likelihood ratio test statistics, Wald-type test statistics as well as Score-type statistics.

We adopt here the notation by Joe et al. [2012], regarding composite likelihood function and the respective CMLE. In this regard, let $\{f(\cdot;\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta\subseteq\mathbb{R}^p,p\geq 1\}$ be a parametric identifiable family of distributions for an observation $\boldsymbol{y}$, a realization of a random m-vector $\boldsymbol{Y}$. In this setting, the composite density based on K different marginal or conditional distributions has the form

$$\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}) = \prod_{k=1}^{K} f_{A_k}^{w_k}(y_j, j \in A_k; \boldsymbol{\theta})$$

and the corresponding composite log-density has the form

$$c\ell(\boldsymbol{\theta},\boldsymbol{y}) = \sum_{k=1}^{K} w_k \ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}),$$

with $\ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}) = \log f_{A_k}(y_j, j \in A_k; \boldsymbol{\theta})$, where $\{A_k\}_{k=1}^{K}$ is a family of random variables associated either with marginal or conditional distributions involving some y_j, $j \in \{1,...,m\}$ and w_k, $k = 1,...,K$ are non-negative and known weights. If the weights are all equal, then they can be ignored. In this case all the statistical procedures produce equivalent results.

Let also $\boldsymbol{y}_1,...,\boldsymbol{y}_n$ be independent and identically distributed replications of $\boldsymbol{y}$. We denote by

$$c\ell(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n) = \sum_{i=1}^{n} c\ell(\boldsymbol{\theta},\boldsymbol{y}_i)$$

the composite log-likelihood function for the whole sample. In complete accordance with the classical MLE, the CMLE, $\widehat{\boldsymbol{\theta}}_c$, is defined by

$$\widehat{\boldsymbol{\theta}}_c = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} c\ell(\boldsymbol{\theta},\boldsymbol{y}_i) = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n}\sum_{k=1}^{K} w_k \ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}_i). \tag{B.5}$$

It can be also obtained by solving the equations.

$$\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n) = \mathbf{0}_p, \tag{B.6}$$

where

$$\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n) = \frac{\partial c\ell(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n)}{\partial\boldsymbol{\theta}} = \sum_{i=1}^{n}\sum_{k=1}^{K} w_k \frac{\partial \ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}_i)}{\partial\boldsymbol{\theta}}.$$

Composite likelihood methods have been successfully used in many applications concerning, for example, genetics (Fearnhead and Donnelly [2002]), generalized linear mixed models (Renard et al. [2004]), spatial statistics (Varin et al. [2005]), frailty models (Henderson and Shimakura [2003]), multivariate survival analysis (Li and Lin [2006]), etc.

B.3.1 Composite likelihood methods based on divergence measures

Let us consider the DPD measure, between the density function $g(\boldsymbol{y})$ and the composite density function $\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})$, i.e.,

$$d_\beta(g(.),\mathcal{CL}(\boldsymbol{\theta},.)) = \int_{\mathbb{R}^m} \left\{ \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} - \left(1+\frac{1}{\beta}\right)\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta} g(\boldsymbol{y}) + \frac{1}{\beta} g(\boldsymbol{y})^{1+\beta} \right\} d\boldsymbol{y} \tag{B.7}$$

for $\beta > 0$, while for $\beta = 0$ we have,

$$\lim_{\beta\to 0} d_\beta(g(.),\mathcal{CL}(\boldsymbol{\theta},.)) = d_{KL}(g(.),\mathcal{CL}(\boldsymbol{\theta},.)).$$

The composite minimum DPD estimator, $\widehat{\boldsymbol{\theta}}_c^{\beta}$, is defined by

$$\widehat{\boldsymbol{\theta}}_c^{\beta} = \arg\min_{\boldsymbol{\theta}\in\Theta} d_\beta(g(.),\mathcal{CL}(\boldsymbol{\theta},.)).$$

In the case of $\beta = 0$ it can be shown that it coincides with the CMLE.

In the case of testing composite null hypothesis is however, necessary to get and study the composite minimum DPD estimator which is restricted by some constraints of the type $\boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{0}_r$, where $\boldsymbol{m}$ is a function $\boldsymbol{m} : \Theta \subseteq \mathbb{R}^p \to \mathbb{R}^r$, r is an integer, with $r < p$, and $\mathbf{0}_r$ denotes the null vector of dimension r. The function $\boldsymbol{m}$ is a vector valued function such that the $p \times r$ matrix

$$\boldsymbol{M}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{m}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

exists and it is continuous in $\boldsymbol{\theta}$ with $rank(\boldsymbol{M}(\boldsymbol{\theta})) = r$. In this context the restricted composite minimum DPD estimator is defined by

$$\widetilde{\boldsymbol{\theta}}_c^{\beta} = \underset{\{\boldsymbol{\theta}\in\Theta:\boldsymbol{m}(\boldsymbol{\theta})=\mathbf{0}_r\}}{\arg\min} \; d_\beta(g\,(.)\,, \mathcal{CL}(\boldsymbol{\theta}, .)), \tag{B.8}$$

where $d_\beta(g\,(.)\,, \mathcal{CL}(\boldsymbol{\theta}, .))$ is defined by (B.7).

Composite minimum DPD estimators were defined in Castilla et al. [2018d], where the associated estimating system of equations and the asymptotic distribution were also provided. It was shown that the composite minimum DPD estimator is an M-estimator and it is asymptotically distributed as a normal with a variance-covariance matrix depending on the tuning parameter β. In this same paper, a robust family of Wald-type tests was introduced, based on the composite minimum DPD estimators, for testing both simple and a composite null hypothesis. The robustness of this new family of tests was studied on the basis of a simulation study.

Following this idea, Rao-type tests were also developed in Castilla et al. [2019]. In this case, when considering a composite null hypothesis, the restricted composite minimum DPD estimator will be needed. A simulation study was developed based on two numerical examples, and a comparison is done between these proposed Rao-type tests and the Wald-type tests developed in Castilla et al. [2018d]. Based on this simulation study, it seems that Wald-type tests are slightly better than the Rao-type tests, but due on the good behavior of both test statistics in relation to the robustness, we may select in each moment the easier test statistic.

B.3.2 Model selection in a composite likelihood framework based on divergence measures

Model selection criteria, for summarizing data evidence in favor of a model, is a very well studied subject in statistical literature, overall in the context of full likelihood. The construction of such criteria requires a measure of similarity between two models, which are typically described in terms of their distributions. This can be achieved if an unbiased estimator of the expected overall discrepancy is found, which measures the statistical distance between the true, but unknown model, and the entertained model. Therefore, the model with smallest value of the criterion is the most preferable model. The use of divergence measures, in particular Kullback-Leibler divergence (Kullback [1997]), to measure this discrepancy, is the main idea of some of the most known criteria: Akaike Information Criterion (AIC, Akaike [1973, 1974]), the criterion proposed by Takeuchi (TIC, Takeuchi [1976]) and other modifications of AIC Murari et al. [2019]. DIC criterion, based on the density power divergence (DPD), was presented in Mattheou et al. [2009] and, recently, Avlogiaris et al. [2019] presented a local BHHJ power divergence information criterion following Avlogiaris et al. [2016]. In the context of the composite likelihood there are some criteria based on Kullback-Leibler divergence, see for instance Varin and Vidoni [2005] and references therein.

In Castilla et al. [2020b] a new information criterion was presented, for model selection in the framework of composite likelihood based on DPD measure, which depends on a tuning parameter β. This criterion, called composite likelihood DIC criterion (CLDIC) coincides as an special case with the criterion given in Varin and Vidoni [2005] as a generalization of the classical criterion of Akaike. After introducing such a criterion, some asymptotic properties were established. A simulation study and two numerical examples were presented in order to point out the robustness properties of the introduced model selection criterion.

Bibliography

S. Aerts and G. Haesbroeck. Robust asymptotic tests for the equality of multivariate coefficients of variation. *Test*, 26(1):163–187, 2017.

A. Agresti. Categorical data analysis, 2nd edn.(john wiley & sons: Hoboken, nj.). 2002.

H. Akaike. Theory and an extension of the maximum likelihood principal. In *International symposium on information theory. Budapest, Hungary: Akademiai Kaiado*, 1973.

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.

T. W. Anderson and D. A. Darling. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, pages 193–212, 1952.

G. Avlogiaris, A. Micheas, and K. Zografos. On local divergences between two probability measures. *Metrika*, 79(3):303–333, 2016.

G. Avlogiaris, A. Micheas, and K. Zografos. A criterion for local model selection. *Sankhya A*, 81 (2):406–444, 2019.

N. Balakrishnan and M. H. Ling. EM algorithm for one-shot device testing under the exponential distribution. *Computational Statistics & Data Analysis*, 56(3):502–509, 2012a.

N. Balakrishnan and M. H. Ling. Multiple-stress model for one-shot device testing data under exponential distribution. *IEEE Transactions on Reliability*, 61(3):809–821, 2012b.

N. Balakrishnan and M. H. Ling. Expectation maximization algorithm for one shot device accelerated life testing with Weibull lifetimes, and variable parameters over stress. *IEEE Transactions on Reliability*, 62(2):537–551, 2013.

N. Balakrishnan and M. H. Ling. Gamma lifetimes and one-shot device testing analysis. *Reliability Engineering & System Safety*, 126:54–64, 2014a.

N. Balakrishnan and M. H. Ling. Best constant-stress accelerated life-test plans with multiple stress factors for one-shot device testing under a Weibull distribution. *IEEE Transactions on Reliability*, 63(4):944–952, 2014b.

N. Balakrishnan and Y. Peng. Generalized gamma frailty model. *Statistics in medicine*, 25(16): 2797–2816, 2006.

N. Balakrishnan, H. Y. So, and M. H. Ling. A Bayesian approach for one-shot device testing with exponential lifetimes under competing risks. *IEEE Transactions on Reliability*, 65(1):469–485, 2015a.

N. Balakrishnan, H. Y. So, and M. H. Ling. EM algorithm for one-shot device testing with competing risks under Weibull distribution. *IEEE Transactions on Reliability*, 65(2):973–991, 2015b.

N. Balakrishnan, H. Y. So, and M. H. Ling. EM algorithm for one-shot device testing with competing risks under Weibull distribution. *IEEE Transactions on Reliability*, 65(2):973–991, 2015c.

N. Balakrishnan, E. Castilla, N. Martín, and L. Pardo. Robust estimators for one-shot device testing data under gamma lifetime model with an application to a tumor toxicological data. *Metrika*, 82(8):991–1019, 2019a.

N. Balakrishnan, E. Castilla, N. Martín, and L. Pardo. Robust estimators and test statistics for one-shot device testing under the exponential distribution. *IEEE Transactions on Information Theory*, 65(5):3080–3096, 2019b.

N. Balakrishnan, E. Castilla, and M. H. Ling. Optimal designs of constant-stress accelerated life-tests for one-shot devices with model misspecification analysis. *Under revision*, 2020a.

N. Balakrishnan, E. Castilla, N. Martín, and L. Pardo. Robust inference for one-shot device testing data under Weibull lifetime model. *IEEE Transactions on Reliability*, 69(3):937–953, 2020b.

N. Balakrishnan, E. Castilla, N. Martín, and L. Pardo. Robust inference for one-shot device testing data under exponential lifetime model with multiple stresses. *Quality and Reliability Engineering International*, 2020c.

N. Balakrishnan, E. Castilla, N. Martin, and L. Pardo. Power divergence approach for one-shot device testing under competing risks. *arXiv preprint arXiv:2004.13372*, 2020d.

N. Balakrishnan, E. Castilla, N. Martin, and L. Pardo. Divergence-based robust inference under proportional hazards model for one-shot device life-test. *IEEE Transactions on Reliability*, DOI: 10.1109/TR.2021.3062289, 2021.

R. Bartnikas and R. Morin. Multi-stress aging of stator bars with electrical, thermal, and mechanical stresses as simultaneous acceleration factors. *IEEE transactions on energy conversion*, 19 (4):702–714, 2004.

P. Basak, I. Basak, and N. Balakrishnan. Estimation for the three-parameter lognormal distribution based on progressively censored data. *Computational Statistics & Data Analysis*, 53(10):3580–3592, 2009.

S. Basak, A. Basu, and M. C. Jones. On the 'optimal' density power divergence tuning parameter. *Journal of Applied Statistics*, 0(0):1–21, 2020.

A. Basu, I. R. Harris, N. L. Hjort, and M. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.

A. Basu, H. Shioya, and C. Park. *Statistical inference: the minimum distance approach*. CRC press, 2011.

A. Basu, A. Mandal, N. Martin, and L. Pardo. Generalized Wald-type tests based on minimum density power divergence estimators. *Statistics*, 50(1):1–26, 2016.

A. Basu, A. Ghosh, A. Mandal, N. Martin, L. Pardo, et al. A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator. *Electronic Journal of Statistics*, 11(2):2741–2772, 2017.

A. Basu, A. Ghosh, N. Martin, and L. Pardo. Robust Wald-type tests for non-homogeneous observations based on the minimum density power divergence estimator. *Metrika*, 81(5):493–522, 2018.

R. Beran et al. Minimum hellinger distance estimates for parametric models. *The annals of Statistics*, 5(3):445–463, 1977.

L. C. Bertens, K. G. Moons, F. H. Rutten, Y. van Mourik, A. W. Hoes, and J. B. Reitsma. A nomogram was developed to enhance the use of multinomial logistic regression modeling in diagnostic research. *Journal of clinical epidemiology*, 71:51–57, 2016.

A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

A. Bhattacharyya. On a measure of divergence between two multinomial populations. *Sankhyā: the indian journal of statistics*, pages 401–406, 1946.

D. A. Binder. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, pages 279–292, 1983.

L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

M. Broniatowski, A. Toma, and I. Vajda. Decomposable pseudodistances and applications in statistical estimation. *Journal of Statistical Planning and Inference*, 142(9):2574–2585, 2012.

E. Castilla and P. J. Chocano. A new robust approach for multinomial logistic regression with complex design model. *Under revision*, 2020.

E. Castilla, A. Ghosh, N. Martin, and L. Pardo. New robust statistical procedures for the polytomous logistic regression models. *Biometrics*, 74(4):1282–1291, 2018a.

E. Castilla, N. Martín, and L. Pardo. Pseudo minimum phi-divergence estimator for the multinomial logistic regression model with complex sample design. *AStA Adv. Stat. Anal.*, 102(3):381–411, 2018b.

E. Castilla, N. Martín, and L. Pardo. A logistic regression analysis approach for sample survey data based on phi-divergence measures. In *The Mathematics of the Uncertain*, pages 465–474. Springer, 2018c.

E. Castilla, N. Martín, L. Pardo, and K. Zografos. Composite likelihood methods based on minimum density power divergence estimator. *Entropy*, 20(1):18, 2018d.

E. Castilla, N. Martín, L. Pardo, and K. Zografos. Composite likelihood methods: Rao-type tests based on composite minimum density power divergence estimator. *Statistical Papers*, pages 1–39, 2019.

E. Castilla, A. Ghosh, N. Martín, S. Muñoz, and L. Pardo. Robust semiparametric inference for polytomous logistic regression with complex survey design. *Under revision*, 2020a.

E. Castilla, N. Martín, L. Pardo, and K. Zografos. Model selection in a composite likelihood framework based on density power divergence. *Entropy*, 22(3):270, 2020b.

E. Castilla, N. Martín, , and L. Pardo. Testing linear hypotheses in logistic regression analysis with complex sample survey data based on phi-divergence measures. *Communications in Statistics-Theory and Methods*, 2020c.

E. Castilla, N. Martín, S. Muñoz, and L. Pardo. Robust Wald-type tests based on minimum rényi pseudodistance estimators for the multiple regression model. *Under revision*, 2020d.

E. Chimitova and N. Balakrishnan. Goodness-of-fit tests for one-shot device testing data. In *Advanced Mathematical and Computational Tools in Metrology and Testing X*, pages 124–131. World Scientific, 2015.

N. Cressie and T. R. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):440–464, 1984.

M. Crowder. Competing risks. *Encyclopedia of Actuarial Science*, 1, John Wiley & Sons, Huboken, New Jersey, 2006.

I. Csiszar. Eine information's theoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoschen ketten, magyar tud. *Akad. Mat*, 1963.

M. J. Daniels and C. Gatsonis. Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine*, 16(20):2311–2325, 1997.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

S. Dey and E. Raheem. Multilevel multinomial logistic regression model for identifying factors associated with anemia in children 6–59 months in northeastern states of india. *Cogent Mathematics*, 3(1):1159798, 2016.

A. Durio and E. D. Isaia. The minimum density power divergence approach in building robust regression models. *Informatica*, 22(1):43–56, 2011.

T.-H. Fan, N. Balakrishnan, and C.-C. Chang. The Bayesian approach for highly reliable electro-explosive devices using one-shot device testing. *Journal of Statistical Computation and Simulation*, 79(9):1143–1154, 2009.

P. Fearnhead and P. Donnelly. Approximate likelihood methods for estimating local recombination rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):657–680, 2002.

M. E. Ghitany, B. Atieh, and S. Nadarajah. Lindley distribution and its application. *Mathematics and computers in simulation*, 78(4):493–506, 2008.

A. Ghosh and A. Basu. Robust estimation for non-homogeneous data and the selection of the optimal tuning parameter: the density power divergence approach. *Journal of Applied Statistics*, 42(9):2056–2072, 2015.

A. Ghosh and A. Basu. Robust bounded influence tests for independent non-homogeneous observations. *Statistica Sinica*, 28(3):1133–1155, 2018.

A. Ghosh, A. Basu, et al. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal of statistics*, 7:2420–2456, 2013.

A. Ghosh, A. Mandal, N. Martín, and L. Pardo. Influence analysis of robust Wald-type tests. *Journal of Multivariate Analysis*, 147:102–126, 2016.

P. K. Gupta and B. Singh. Parameter estimation of Lindley distribution with hybrid censored data. *International Journal of System Assurance Engineering and Management*, 4(4):378–385, 2013.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 1986.

R. Henderson and S. Shimakura. A serially correlated gamma frailty model for longitudinal count data. *Biometrika*, 90(2):355–366, 2003.

C. Hong and Y. Kim. Automatic selection of the turning parameter in the minimum density power divergence estimation. *Journal of the Korean Statistical Society*, 30(3):453–465, 2001.

P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.

G. J. Husak, J. Michaelsen, and C. Funk. Use of the gamma distribution to represent monthly rainfall in africa for drought monitoring applications. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(7):935–944, 2007.

H. Joe, N. Reid, P. Somg, D. Firth, and C. Varin. Composite likelihood methods. In *Report on the Workshop on Composite Likelihood*, 2012.

M. Jones, N. L. Hjort, I. R. Harris, and A. Basu. A comparison of related density-based minimum divergence estimators. *Biometrika*, 88(3):865–873, 2001.

R. Kodell and C. Nelson. An illness-death model for the study of the carcinogenic process using survival/sacrifice data. *Biometrics*, pages 267–277, 1980.

S. Kullback. *Information theory and statistics*. Courier Corporation, 1997.

K. Kwon and D. M. Frangopol. Bridge fatigue reliability assessment using probability density functions of equivalent stress range based on field monitoring data. *International Journal of Fatigue*, 32(8):1221–1232, 2010.

J. H. Lau, G. Harkins, D. Rice, J. Kral, and B. Wells. Experimental and statistical analyses of surface-mount technology plcc solder-joint reliability. *IEEE transactions on reliability*, 37(5): 524–530, 1988.

Y. Li and X. Lin. Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, 101(474):591–603, 2006.

F. Liese and I. Vajda. *Convex statistical distances*, volume 95. Teubner, 1987.

D. V. Lindley. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 102–107, 1958.

B. G. Lindsay et al. Efficiency versus robustness: the case for minimum hellinger distance and related methods. *The annals of statistics*, 22(2):1081–1114, 1994.

J. C. Lindsey and L. M. Ryan. A three-state multiplicative model for rodent tumorigenicity experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 42(2):283–300, 1993.

M. Ling and X. Hu. Optimal design of simple step-stress accelerated life tests for one-shot devices under Weibull distributions. *Reliability Engineering & System Safety*, 193:106630, 2020.

M. H. Ling. *Inference for one-shot device testing data*. PhD thesis, McMaster University, Hamilton, Ontario, Canada, 2012.

M. H. Ling. Optimal design of simple step-stress accelerated life tests for one-shot devices under exponential distributions. *Probability in the Engineering and Informational Sciences*, 33(1): 121–135, 2019.

M. H. Ling, H. Y. So, and N. Balakrishnan. Likelihood inference under proportional hazards model for one-shot device testing. *IEEE Transactions on Reliability*, 65(1):446–458, 2015.

R. E. Little and E. H. Jebe. *Statistical design of fatigue experiments*. Applied Science Publishers, 1975.

K. Mattheou, S. Lee, and A. Karagrigoriou. A model selection criterion based on the bhhj measure of divergence. *Journal of Statistical Planning and Inference*, 139(2):228–235, 2009.

J. Mazucheli and J. A. Achcar. The Lindley distribution applied to competing risks lifetime data. *Computer methods and programs in biomedicine*, 104(2):188–192, 2011.

P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

W. Q. Meeker. A comparison of accelerated life test plans for Weibull and lognormal distributions and type i censoring. *Technometrics*, 26(2):157–171, 1984.

W. Q. Meeker, L. A. Escobar, and C. J. Lu. Accelerated degradation tests: modeling and analysis. *Technometrics*, 40(2):89–99, 1998.

C. A. Meeter and W. Q. Meeker. Optimum accelerated life tests wth a nonconstant scale parameter. *Technometrics*, 36(1):71–83, 1994.

R. Mises. Les lois de probabilité pour les fonctions statistiques. In *Annales de l'institut Henri Poincaré*, volume 6, pages 185–212, 1936.

R. Mises. Sur les fonctions statistiques. *Bulletin de la Société Mathématique de France*, 67:177–184, 1939.

R. Mises. On the asymptotic distribution of differentiable statistical functions. *The annals of mathematical statistics*, 18(3):309–348, 1947.

D. Morales, L. Pardo, and I. Vajda. Asymptotic divergence of estimates of discrete distributions. *Journal of Statistical Planning and Inference*, 48(3):347–369, 1995.

J. G. Morel. Logistic regression under complex survey designs. *Survey Methodology*, 15(2):203–223, 1989.

A. Murari, E. Peluso, F. Cianfrani, P. Gaudio, and M. Lungaroni. On the use of entropy to improve model selection criteria. *Entropy*, 21(4):394, 2019.

W. Nelson. Accelerated life testing-step-stress models and data analyses. *IEEE transactions on reliability*, 29(2):103–108, 1980.

M. Newby. Monitoring and maintenance of spares and one shot devices. *Reliability Engineering & System Safety*, 93(4):588–594, 2008.

H. Ng, P. Chan, and N. Balakrishnan. Estimation of parameters from progressively censored data using EM algorithm. *Computational Statistics & Data Analysis*, 39(4):371–386, 2002.

E. Nogueira, M. Vázquez, and N. Núñez. Evaluation of algainp leds reliability based on accelerated tests. *Microelectronics Reliability*, 49(9-11):1240–1243, 2009.

D. Olwell and A. Sorell. Warranty calculations for missiles with only current-status data, using Bayesian methods. In *Annual Reliability and Maintainability Symposium. 2001 Proceedings. International Symposium on Product Quality and Integrity (Cat. No. 01CH37179)*, pages 133–138. IEEE, 2001.

L. Pardo. *Statistical inference based on divergence measures.* CRC press, 2005.

F. Pascual. Accelerated life test planning with independent Weibull competing risks with known shape parameter. *IEEE Transactions on Reliability*, 56(1):85–93, 2007.

D. Renard, G. Molenberghs, and H. Geys. A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis*, 44(4):649–667, 2004.

S. E. Rigdon, B. R. Englert, I. A. Lawson, C. M. Borror, D. C. Montgomery, and R. Pan. Experiments for reliability achievement. *Quality Engineering*, 25(1):54–72, 2012.

G. Roberts, N. Rao, and S. Kumar. Logistic regression analysis of sample survey data. *Biometrika*, 74(1):1–12, 1987.

J. Rodrigues, H. Bolfarine, and F. Louzada-Neto. Comparing several accelerated life models. *Communications in Statistics-Theory and Methods*, 22(8):2297–2308, 1993.

E. Ronchetti and P. Rousseeuw. The influence curve for tests. In *Research Report No 21.* Fachgruppe für Statistik ETH Zürich, 1979.

J. B. Seaborn. The confluent hypergeometric function. In *Hypergeometric Functions and Their Applications*, pages 41–51. Springer, 1991.

R. Shanker, F. Hagos, and S. Sujatha. On modeling of lifetimes data using exponential and Lindley distributions. *Biometrics & Biostatistics International Journal*, 2(5):1–9, 2015.

D. G. Simpson. Minimum hellinger distance estimation for the analysis of count data. *Journal of the American statistical Association*, 82(399):802–807, 1987.

D. G. Simpson. Hellinger deviance tests: efficiency, breakdown points, and examples. *Journal of the American Statistical Association*, 84(405):107–113, 1989.

H. Y. So. *Some Inferential Results for One-Shot Device Testing Data Analysis.* PhD thesis, McMaster University, Hamilton, Ontario, Canada, 2016.

M. Srinivas and T. Ramu. Multifactor aging of hv generator stator insulation including mechanical vibrations. *IEEE Transactions on Electrical Insulation*, 27(5):1009–1021, 1992.

E. W. Stacy and G. A. Mihram. Parameter estimation for a generalized gamma distribution. *Technometrics*, 7(3):349–358, 1965.

K. Takeuchi. Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, 153:12–18, 1976.

R. N. Tamura and D. D. Boos. Minimum hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, 81(393):223–229, 1986.

S.-T. Tseng, N. Balakrishnan, and C.-C. Tsai. Optimal step-stress accelerated degradation test plan for gamma degradation processes. *IEEE Transactions on Reliability*, 58(4):611–618, 2009.

C. Varin and P. Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005.

C. Varin, G. Høst, and Ø. Skare. Pairwise likelihood inference in spatial generalized linear mixed models. *Computational statistics & data analysis*, 49(4):1173–1191, 2005.

M. Vázquez, N. Núñez, E. Nogueira, and A. Borreguero. Degradation of alingap red leds under drive current and temperature accelerated life tests. *Microelectronics Reliability*, 50(9-11):1559–1562, 2010.

W. Wang and D. B. Kececioglu. Fitting the Weibull log-linear model to accelerated life-test data. *IEEE Transactions on Reliability*, 49(2):217–223, 2000.

J. Warwick and M. Jones. Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation*, 75(7):581–588, 2005.

J. Wolfowitz. Consistent estimators of the parameters of a linear structural relation. *Scandinavian Actuarial Journal*, 1952(3-4):132–151, 1952.

J. Wolfowitz. Estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics*, 5(1):9–23, 1953.

J. Wolfowitz. Estimation by the minimum distance method in nonparametric stochastic difference equations. *The Annals of Mathematical Statistics*, 25(2):203–217, 1954.

J. Wolfowitz. The minimum distance method. *The Annals of Mathematical Statistics*, pages 75–88, 1957.

M. Zelen. Factorial experiments in life testing. *Technometrics*, 1(3):269–288, 1959.

When you make the finding yourself – even if you're the last person on Earth to see the light – you'll never forget it.

CARL SAGAN