Estimators and Tests based on Likelihood-Depth

with Application to Weibull Distribution, Gaussian and Gumbel Copula

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik und Naturwissenschaften
der Universität Kassel

vorgelegt von
Dipl.-Math. Liesa Denecke
geboren am 18.06.1981 in Gießen

März 2010

Tag der mündlichen Prüfung
23. Juni 2010

Erstgutachter
Prof. Dr. Christine Müller
Technische Universität Dortmund

Zweitgutachter
Prof. Dr. Andreas Christmann
Universität Bayreuth

Liesa Denecke
Lehrstuhl für Statistik mit Anwendungen im Bereich der Ingenieurwissenschaften
Fakultät Statistik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund

denecke@statistik.tu-dortmund.de

Danksagung

Ich möchte mich bei Prof. Dr. Christine Müller für die hervorragende Betreuung und Förderung bedanken.

Bei Herrn Prof. Dr. Andreas Christmann bedanke ich mich für die Übernahme des Koreferats.

Diese Arbeit entstand in der Zeit von August 2006 bis März 2010 an der Universität Kassel und der Technischen Universität Dortmund. Ich möchte mich auf diesem Wege auch bei allen Kollegen des Fachbereichs Mathematik in Kassel und der Fakultät Statistik in Dortmund für die gute Zusammenarbeit und die kollegiale Atmosphäre bedanken.

Außerdem danke ich Nikolina Miteff-Denecke, Lea Kleinspehn und Stephan Schreiber für die Korrekturen der Arbeit.

Insbesondere gilt mein Dank meinen Eltern und meinem Freund Stephan für die Unterstützung in den letzten Jahren.

Dortmund, im März 2010 Liesa Denecke

Deutschsprachige Zusammenfassung

In dieser Arbeit werden mithilfe der Likelihood-Tiefen (ausreißer-)robuste Schätzfunktionen und Tests für den unbekannten Parameter $\theta \in \Theta \subseteq \mathbb{R}^q$ der stetigen Dichtefunktion f_θ einer Verteilung entwickelt. Die entwickelten Verfahren werden dann auf drei verschiedene Verteilungen angewandt.

Werden Daten aufgenommen, zum Beispiel durch eine Messung, so muss immer damit gerechnet werden, dass diese Daten zum Beispiel Messfehler enthalten oder dass extreme Werte auftreten. Diese Ausreißer können einige Schätzfunktionen und Tests erheblich beeinflussen. Beispielsweise kann das arithmetische Mittel schon durch *eine* extreme Beobachtung verfälscht werden. Daher gibt es Bestrebungen robuste Verfahren zu finden, die durch Ausreißer nicht oder kaum beeinflusst werden. Mithilfe der von Müller und Mizera in [MiMu 2004] und [Mue 2005] entwickelten Likelihood-Tiefen lassen sich solche Verfahren herleiten. Die Likelihood-Tiefen bilden Verallgemeinerungen der Datentiefe und erweitern Tukeys Halbraumtiefe, siehe [Tuk 1975], und die Simplex-Tiefe von Liu, siehe [Liu 1988, Liu 1990], welche ausreißer-robuste Verallgemeinerungen des Medians für multivariate Daten darstellen.

Die Größe der Likelihood-Tiefe soll ein Maß dafür sein, wie gut ein Parameter zum Datensatz passt. Je größer die Tiefe ist, desto besser passt der Parameter zum Datensatz. Wählt man als Schätzfunktion diejenige, die jedem Datensatz den Parameter mit der größten Tiefe zuordnet, stellt man jedoch fest, dass diese für einige Verteilungen verfälschte Ergebnisse liefert. Im Mittel wird in diesen Fällen dadurch nicht der wahre zu Grunde liegende Parameter geschätzt.

Für eindimensionale Parameter wird die Likelihood-Tiefe eines Parameters θ im Datensatz als das Minimum aus dem Anteil der Daten, für die die Ableitung der Loglikelihood-Funktion nach dem Parameter θ, $\frac{\partial}{\partial\theta} \ln f_\theta(z)$, nicht negativ ist, und dem Anteil der Daten, für die diese Ableitung nicht positiv ist, berechnet. Damit hat der Parameter die größte Tiefe, für den beide Anzahlen gleich groß sind. Asymptotisch hat der Parameter die größte Tiefe, für den die Wahrscheinlichkeit, dass für eine Beobachtung $\frac{\partial}{\partial\theta} \ln f_\theta(\cdot)$ nicht negativ ist, gleich $\frac{1}{2}$ ist. Wenn dies für den zu Grunde liegenden Parameter nicht der Fall ist, ist der Schätzer basierend auf der Likelihood-Tiefe verfälscht. In dieser Arbeit wird gezeigt, wie diese Verfälschung bestimmt werden und unter welchen Voraussetzungen sie korrigiert werden kann. Zudem wird die Korrektur berechnet. Des Weiteren wird für die neu konstruierten Schätzfunktionen gezeigt, wann sie konsistente Schätzungen bilden.

Die Arbeit besteht aus zwei Teilen. Im ersten Teil der Arbeit wird die allgemeine Theorie über die Schätzfunktionen und Tests dargestellt und zudem deren jeweiligen Konsistenz gezeigt. Es werden zwei verschiedene Theoreme für den Nachweis der Konsistenz der

Schätzfunktionen, unter unterschiedlichen Voraussetzungen an die zu Grunde liegende Verteilung, gegeben. Hierfür werden unter anderem eine Verallgemeinerung des Glivenko-Cantelli Lemmas und Vapnik-$\tilde{\text{C}}$ervonenkis Klassen benutzt.

Zur Entwicklung von Tests für den Parameter θ der Form $H_0 : \theta \in \Theta_0 \subset \Theta$ gegen $H_1 : \theta \notin \Theta_0$, wird die von Müller in [Mue 2005] entwickelte Simplex Likelihood-Tiefe, die eine U-Statistik ist, benutzt. Es zeigt sich, dass für dieselben Verteilungen, für die die Likelihood-Tiefe verfälschte Schätzer liefert, die Simplex Likelihood-Tiefe eine unverfälschte U-Statistik ist. Damit ist insbesondere die asymptotische Verteilung bekannt und es lassen sich Tests für verschiedene Hypothesen formulieren. Die Verschiebung in der Tiefe führt aber für einige Hypothesen zu einer schlechten Güte des zugehörigen Tests. Es werden daher korrigierte Tests eingeführt und Voraussetzungen angegeben, unter denen diese dann konsistent sind.

Im zweiten Teil wird die Theorie auf drei verschiedene Verteilungen angewandt: Die Weibull-Verteilung, die Gauß- und die Gumbel-Copula. Damit wird gezeigt, wie die Verfahren des ersten Teils genutzt werden können, um (robuste) konsistente Schätzfunktionen und Tests für den unbekannten Parameter der Verteilung herzuleiten.

Zunächst betrachten wir die zweiparametrige eindimensionale Weibull-Verteilung. Sie findet ihre größte Anwendung im Bereich der Analyse von Lebenszeiten. Dies geschieht insbesondere in den ingenieurwissenschaftlichen Anwendungen aber auch in klinischen Studien oder ähnlichem. Ihre Verteilungsfunktion hat eine vergleichsweise einfache Form und es lassen sich verschiedene Ausfallrisiko-Verläufe simulieren.
Es werden Schätzfunktionen und Test, sowohl für komplette Datensätze als auch für Typ-I rechtszensierte Daten mit fester Zensurzeit, für beide Parameter der Weibull-Verteilung entwickelt. Zensierte Daten entstehen dann, wenn Zeiten bis zu einem bestimmten Ereignis aufgenommen werden und für manche Daten nur festgehalten werden kann, dass das interessierende Ereignis bis zum Zensurzeitpunkt noch nicht eingetreten ist (z.B. da die Studie endet). Es wird für einen unbekannten Parameter als auch für die Situation, dass beide Parameter unbekannt sind, gezeigt, dass die auf der Likelihood-Tiefe basierten Schätzfunktionen konsistent sind, jeweils in der Situation von vollständig beobachteten Daten und Typ-I rechtszensierten Daten. Im nächsten Schritt werden Tests für Hypothesen über die Parameter hergeleitet. Dabei muss allerdings davon ausgegangen werden, dass der jeweils andere, nicht getestete Parameter bekannt ist. In diesem Fall können wir die Konsistenz der Tests für unzensierte Daten für die Tests bezüglich beider Parameter nachweisen. Für zensierte Daten kann die Konsistenz im Falle des sogenannten Formparameters nicht in allen Situationen nachgewiesen werden, für den Skalenparameter hingegen schon. In Simulationsstudien wird das Verhalten der Tests untersucht, wenn auch der jeweils andere, nicht getestete Parameter unbekannt ist. Hier zeigt sich, dass sich die Gütefunktionen genauso verhalten, wie in dem Fall, wenn der Parameter bekannt ist. Die neu entwickelte Schätzer und die Tests werden mit bekannten - auch mit robusten - Methoden verglichen. Es zeigt sich, dass die neue Methode, insbesondere in Bezug auf Robustheit im Vergleich mit dem Maximum-Likelihood-Schätzer, in einigen Fällen deutlich bessere Ergebnisse liefert als bestehende. Für zensierte Daten zeigt sich auch die Überlegenheit der neuen Methode gegenüber der ebenfalls betrachteten robusten

Methode basierend auf dem Median von He und Fung, siehe [HeFu 1999].

Im letzten Kapitel des zweiten Teils werden zwei spezielle Copula-Familien betrachtet: Die Gauß- und die Gumbel-Copula. Mithilfe von Copulas lassen sich Abhängigkeitsstrukturen modellieren. Sie finden ihre Hauptanwendung im Bereich der Finanz- und Versicherungsmathematik. Für zwei Zufallsvariablen mit Randverteilung gegeben durch die Standardnormalverteilung, und gemeinsamer Verteilung gegeben durch die zweidimensionale Normalverteilung, ist die zugehörige Copula die Gauß-Copula. Solche Variablen werden im ersten Teil des vierten Kapitels betrachtet. Der unbekannte Parameter ist hier der Korrelationskoeffizient. Die Techniken aus dem ersten Teil werden genutzt, um Schätzer und Tests für den unbekannten Korrelationskoeffizienten von zweidimensional normalverteilten Zufallsvariablen zu entwickeln. Für den Nachweis der Konsistenz der Schätzfunktion werden die Vapnik-$\tilde{\text{C}}$ervonenkis Klassen genutzt. Die Konsistenz der Tests für den Korrelationskoeffizienten kann mit den Methoden des ersten Teils nicht gezeigt werden. Simulationsstudien zeigen die Robustheit gegen kontaminierte Daten.
Abschließend wird die Theorie genutzt, um Schätzer und Tests für den Parameter der zweidimensionalen Gumbel-Copula herzuleiten. Hier gehen wir zunächst davon aus, dass die Randverteilungen bekannt sind. Es wird plausibel gemacht, dass die Voraussetzungen aus dem Theorie-Kapitel erfüllt sind, womit sich die Konsistenz der Schätzfunktion und Tests ergibt. In Simulationsstudien wird untersucht, ob das Schätzen der Randverteilung einen Einfluss auf den Schätzer bzw. die Güte der Tests hat. Dieses ist nicht der Fall.

Insgesamt zeigt sich, dass für die drei Verteilungen mithilfe der Likelihood-Tiefen robuste Schätzfunktionen und Tests gefunden werden können. In unverfälschten Daten sind vorhandene Standardmethoden zum Teil überlegen, jedoch zeigt sich der Vorteil der neuen Methoden in kontaminierten Daten und Daten mit Ausreißern.

Contents

1. Introduction **1**

I. General Theory 5

2. Likelihood-Depth **7**

2.1. Estimators . . . 10
2.2. Tests . . . 16
2.3. Open problems . . . 32

II. Application to special distributions 33

3. Weibull distribution **35**

3.1. Preliminaries . . . 35
3.2. Estimators for the parameters of the Weibull distribution . . . 37
3.2.1. Uncensored data with known shape parameter . . . 37
3.2.2. Uncensored data with known scale parameter . . . 40
3.2.3. Uncensored data, shape and scale parameter unknown . . . 44
3.2.4. Type-I right-censored data, shape or scale parameter known . . . 54
3.2.5. Type-I right-censored data, shape and scale parameter unknown . 58
3.3. Tests and confidence intervals for the shape parameter . . . 72
3.3.1. Uncensored data with known scale parameter . . . 73
3.3.2. Uncensored data with unknown scale parameter . . . 84
3.3.3. Type-I right-censored data with known scale parameter . . . 89
3.3.4. Type-I right-censored data with unknown scale parameter . . . 104
3.4. Tests and confidence intervals for the scale parameter . . . 111
3.4.1. Uncensored data with known shape parameter . . . 111
3.4.2. Uncensored data with unknown shape parameter . . . 119
3.4.3. Type-I right-censored data with known shape parameter . . . 121
3.4.4. Type-I right-censored data with unknown shape parameter . . . 127
3.5. Open problems . . . 128

4. Copulas **131**

4.1. Preliminaries . . . 131
4.2. Estimator for the correlation coefficient . . . 135
4.3. Tests and confidence intervals for the correlation coefficient . . . 149

4.4. Estimator for the parameter of the Gumbel copula . . . 160
4.4.1. Data with unknown margins . . . 168
4.5. Tests and confidence intervals for the parameter of the Gumbel copula . . 169
4.6. Open problems . . . 176

A. Weak convergence and empirical processes 179

B. R source code 183
B.1. Weibull distribution . . . 183
B.1.1. Estimators . . . 185
B.1.2. Tests . . . 189
B.2. Gaussian copula . . . 199
B.2.1. Estimator . . . 201
B.2.2. Tests . . . 201
B.3. The Gumbel copula . . . 203
B.3.1. Estimator . . . 205
B.3.2. Tests . . . 205

List of Symbols 209

Bibliography 211

Erklärung 215

1. Introduction

The subject of this thesis is to find robust estimators and tests for the unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^q$ of the continuous density function f_θ of identically independently distributed (i.i.d.) variables $Z_1, \ldots, Z_N$. These estimators and tests shall be based on the so-called likelihood-depth.

If the data are contaminated with outliers, for example very high or very low values (e.g. arising from measurement errors) or values coming from a different distribution, some estimators and tests can be rather unreliable. This is the case for example for the maximum likelihood estimator. The aim is to find (outlier) robust estimators and tests, which are not or only slightly affected by outliers.

Likelihood-depth and simplicial likelihood-depth are general notions of data depth, first used by Mizera and Müller [MiMu 2004] and Müller [Mue 2005]. They extend the half space depth of Tukey [Tuk 1975] and the simplicial depth of Liu [Liu 1988, Liu 1990], which lead to outlier robust generalizations of the median for multivariate data. The likelihood-depths belong to a broad class of depth notions introduced and studied in the last 20 years; see e.g. Rousseeuw and Hubert [RH 1999], Zuo and Serfling [ZoSe 2000a, ZoSe 2000b], Mizera [Miz 2002], and the book of Mosler [Mos 2002]. Although the likelihood-depth bases on a parametric approach, it can lead to distribution-free estimators and tests as Mizera and Müller in [MiMu 2004] demonstrated for location-scale estimation and Müller [Mue 2005] for regression. Müller [Mue 2005] also showed that the simplicial likelihood-depth is in particular appropriate for testing since it is an U-statistic. However, simplicial likelihood-depth is often a degenerated U-statistic so that the spectral decomposition of conditional expectations are needed for deriving the asymptotic distribution, which was done in Müller [Mue 2005], Wellman et. al. [WeHaMu 2009], and Wellmann and Müller [WeMu 2010] for regression. In these cases, the likelihood-depth estimator is asymptotically an unbiased estimator. In many other cases, the likelihood-depth estimator is asymptotically a biased estimator. Then the simplicial likelihood-depth is not a degenerated U-statistic and its asymptotic distribution is simply the normal distribution. Hence, asymptotic α-level tests can be easily derived. Thereby, rather general hypotheses can be tested and the resulting tests are outlier robust. But these tests have a bad power for some alternatives due to the bias. In particular, the power at such alternatives is not converging to one with growing sample size as this should be for consistent tests.

The Weibull distribution is often used in survival analysis, especially in life-testing and reliability studies. It was introduced by Weibull in 1951; see [Wei 1951]. Moreover, the distribution function is one-dimensional and depends on two parameters. The Weibull

model is frequently used in engineering applications, but also in biology or other fields. For an introduction see e.g. the textbooks of Rinne [Rin 2009], Murthy, Xie and Jiang [MXJ 2004], Lee and Wang [LeWa 2003] or Lawless [Law 2003]. With the help of the Weibull distribution constant as well as de- and increasing Hazard-functions can be modeled. Because of this and the fact that the survival-function has a simple form it is used in many applications as for instance the study of durability of materials in engineerings, medical lifetime studies, etc.
Most times the maximum likelihood estimator is used for parametric estimation; it can be found e.g. in Cohen [Coh 1965] or the textbook of Lee and Wang [LeWa 2003]. The maximum likelihood estimator can also be defined for censored data. Other methods can be found in the textbook of Rinne [Rin 2009], where a survey of estimation procedures is given. In this textbook also exists a hint to a robust estimator for the parameters of the Weibull distribution developed by He and Fung in [HeFu 1999], an estimator based on the so-called method of medians. He and Fung propose that their estimator is not infected by right censoring, but only if less than 16 % of the largest observations are censored.
There exist also other methods to define robust estimators for the Weibull distribution, for example Dixit [Dix 1994] proposes a Bayesian approach, Shier and Lawrence [ShLa 1984] estimate via regression, Marks [Ma 2005] gives an estimator based on quantiles and Boudt, Caliskan and Croux [BCC 2009] present estimators like the quantile least squares, repeated median and median/Q_n estimator. Seki and Yokoyama [SeYo 1996] state a bootstrap method and Cacciari et al. [CMMJ 2002] use a modified Thiel method. However, all these robust procedures are proposed for complete data and not for censored ones. Only Homan and Lachenbruch [HoLa 1986] introduce a robust estimator for the parameter of the exponential distribution, which can also be used for censored data. Note that the exponential distribution is a special Weibull distribution.
The textbook of Rinne [Rin 2009] gives a good overview of the developed test procedures for the parameters of the Weibull distribution, for complete and for censored data. Tests based on maximum likelihood procedures are presented there. Besides, many articles deal with tests and confidence intervals for the parameters of the Weibull distribution in censored and uncensored data, see for example Balakrishnan and Stehlik [BaSt 2008], Chen [Che 1997], Wong and Wong [WoWo 1982] or Kahle [Ka 1996]. But still these methods are not robust against contamination. Only He and Fung [HeFu 1999] give an outlier robust confidence interval for the shape parameter.

The copula model has a variety of applications because it models dependence structures. For example in finance, in the analysis of credit risks, the insolvency of several debtors at the same time or for insurances the risk of appearance of different claims at the same time have to be modeled to insure solvency of the bank and insurance, respectively, all the time. Copulas can also be used in the simulation of technical production processes to model the occurrence of coupled failures. Some applications of copulas can be found in Aas [Aas 2004], Andresen [And 2005], Cízek, Härdle and Weron [CHW 2005], Dobríc and Schmid [DSch 2005] or Malvergne and Sornette [MaSo 2006]. For an introduction to copulas see the textbook of Nelsen [Nel 2006] or Joe [Joe 1997].
Different estimation procedures for copulas were introduced. Parametric, semi-parametric

and nonparametric methods are proposed. Most of the parametric and semi-parametric methods are two-stage estimations, as presented in Andresen [And 2005], Durrleman, Nikeghbali, and Roncalli [DNR 2000], Genest, Ghoudi and Rivest [GGR 1995], Hoff [Hoff 2007] or Kim, Silvapulle and Silvapulle [KSS 2007] for example. Usually, a first step is the estimation of the margins by parametric or non-parametric methods. Afterwards an estimation procedure for the parameter of the copula is presented. Genest and Segers [GeSe 2009] present rank-based estimators in the situation of unknown margins. An example for a nonparametric estimation model for the copula is the empirical copula, see Durrleman, Nikeghbali and Roncalli [DNR 2000] or Capéraà, Fougères and Genest [CFG 1997]. Some goodness-of-fit-tests can be found in Dobríc and Schmid [DSch 2005], Fermian [Fer 2005] or Panchenko [Pan 2005]. But to the authors knowledge, nothing is known about the robustness behavior, when contamination with data from other distributions occurs.

This thesis is divided into two parts. The first part deals with the general theory about estimators and tests based on likelihood-depth, while the second shows the application of this general theory to three distributions, namely the one-dimensional Weibull distribution, the two-dimensional Gumbel copula and the two-dimensional Gaussian copula or more precisely the Gaussian copula with normal distributed margins, what is the two-dimensional normal distribution.

We consider continuous density functions. In the first part we find outlier-robust estimators and tests based on the likelihood-depth. In some cases the results of Mizera and Müller [Mue 2005] and Müller [MiMu 2004] can not be used directly. More precisely, it can happen that the maximum likelihood-depth estimator is biased. We determine the bias, correct it and show that the resulting estimators are strongly consistent. Therefore, we use among other things an extension of the Glivenko-Cantelli-Lemma; see van der Vaart and Wellner [VaWe 1996]. To prove the assumptions of this lemma so-called Vapnik-C̃ervonenkis classes are used; see also [VaWe 1996]. This is done in Chapter 2 after an introduction to likelihood-depth. The tests based on simplicial likelihood-depth are established in Section 2.2. In situations, where the maximum likelihood-depth estimator is biased the simplicial depth is a non degenerated U-statistic, so its distribution is especially known. But for some hypotheses the tests must be corrected, too. We determine this correction, examine the asymptotic power of the corrected tests and show that the correction improves the power of the test. Furthermore, we prove consistency of the tests.

In the second part of this work the results of Chapter 2 are used in three different situations. We start in Chapter 3 with the Weibull distribution and develop robust estimators for uncensored and type-I right-censored data with fixed censor time. Especially, we consider the cases, where one of the two parameters has to be estimated and where both parameters are unknown. In all cases we prove that the resulting estimators are consistent. The estimation of the two parameters of the Weibull distribution can be done one after the other, so we can apply the results of Chapter 2. Further, we show in simulation studies that the new estimators are superior to the maximum likelihood estimators for contaminated data and that they receive better results than the estimators based on the

method of medians by He and Fung for censored data, even for data without outliers. Consistent tests for the shape and scale parameter are given for complete data for both parameters, considering the other parameter to be known, and in case of the scale parameter also for censored data, again considering the shape parameter to be known. These tests are also compared with different existing methods. Again, simulation studies show that the new methods are robust and that they also work very well for censored data unlike the tests based on the method of medians by He and Fung [HeFu 1999].
Copulas and in particular the Gaussian and the Gumbel copula are the matter of Chapter 4. A short introduction to copulas is given in Section 4.1 and the specific definition of the Gaussian and the Gumbel copula are presented. In both cases we restrict ourselves to the case of two-dimensional variables. The next two Subsections 4.2 and 4.3 deal with estimation of and tests for the correlation coefficient, which is the unknown parameter in case of the two-dimensional Gaussian copula with normal distributed margins. In Section 4.4 and 4.5 we give estimators and tests for the Gumbel copula. For both distributions we examine in simulation studies the outlier robustness of the new methods and their behavior for finite sample sizes and compare the new methods to existing procedures. Moreover, for the Gumbel copula we consider also the case, where the margin distributions are unknown and have to be estimated first. We see that the estimation has no influence on the estimator, respectively the power of the tests.

For all three distributions the new estimators and tests are little worse than standard methods, concerning uncontaminated data. But considering contaminated data and data with outliers, they are more robust than these standard methods like the maximum likelihood estimator.

In the Appendix A one of the Glivenko-Cantelli theorems from van der Vaart and Wellner [VaWe 1996] and the definitions used there are given. Also we define the Vapnik-C̃ervonenkis classes and cite the results that show that these classes fulfill the assumptions of the Glivenko-Cantelli theorem. Some of the methods are implemented in R, see [R 2009], and the R source code can be found in the Appendix B.
The symbols used are explained in the List of Symbols on page 209.

Part I.

General Theory

2. Likelihood-Depth

We present the likelihood-depth that will be used to develop estimators and tests for an unknown parameter θ of the density function f_θ, known except for θ.

In his work, [Miz 2002], Mizera introduces general definitions of depth. He gives definitions of global, local and tangent depth. The depth shall be a measure of data-analytic admissibility and is based on the principle of Rousseeuw and Hubert [RH 1999], that says "the depth of θ is the smallest number of observations that would need to be removed to make θ a nonfit". Mizera introduces for this so-called criterial functions for every observation, such that the lower the value of the criterial function is at θ, the better θ fits the observation. Müller and Mizera show in their work "Location-Scale Depth", [MiMu 2004], that these criterial functions can be based on likelihood-functions. They introduce the likelihood-nonfit, the global and the tangential likelihood-depth. For testing Müller introduces the simplicial likelihood-depth, see [Mue 2005].

To motivate the use of depth, we will introduce data depth and Tukey's half space depth, see [Tuk 1975]. Data depth is a concept to generalize the median to multivariate data. The median of data $x_* := (x_1, \dots, x_N)$, $x_i \in \mathbb{R}$, $i = 1, \dots, N$, is lying in the middle of the data, i.e. for the ordered data $x_{(1)}, \dots, x_{(N)}$ the median is defined as

$$x_{med} := \begin{cases} x_{\left(\frac{N+1}{2}\right)}, & N \text{ odd} \\ \frac{1}{2}\left(x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}\right), & N \text{ even} \end{cases} .$$

Now for univariate data the data depth of a location parameter μ in the data is defined as the minimum of the part of the data that is smaller than or equal to μ and the part of the data that is greater than of equal to μ, i.e.

$$d(\mu, x_*) := \frac{1}{N} \min \left(\sharp\{n; x_n \leq \mu\}, \sharp\{n; x_n \geq \mu\} \right).$$

The median is that parameter that maximizes the data depth. But what is the middle of the data for multivariate data? To generalize the concept of data depth to multivariate data, Tukey introduces in [Tuk 1975] the half space depth. For a location parameter μ in multivariate data $z_* = (z_1, \dots, z_N)$ the half space depth is the minimum part of observations that lie in a half space containing μ, i.e.

$$d_H(\mu, z_*) := \frac{1}{N} \min_H \sharp \{n; z_n \text{ lies in a half space } H \text{ containing } \mu\}.$$

The Tukey median are all parameter with maximum half space depth. To generalize the half space depth to non-location parameters, the concept of nonfit by Rousseeuw and

Hubert [RH 1999] is used. An equivalent definition of the half space depth is given as the minimum part of observations that must be omitted so that μ becomes a nonfit in the data left, i.e. we can find a parameter $\tilde{\mu}$ that has a smaller distance to all observations left. By generalizing the nonfit, respectively the "distance", the half space depth is generalized to non-location parameters. One generalization are the likelihood-depths.

Here in this work we will always consider identically and independently distributed (i.i.d.) m-dimensional variables $Z_1, \dots, Z_N$. The density f_θ of Z_i shall be known except for the parameter $\theta \in \mathbb{R}^q$. We assume that $f_\theta(\cdot)$ is continuous, $f_{(\cdot)}(z)$ uniformly continuous, differentiable and that the partial derivatives are continuous.

Notations. *We will use the following notations throughout the whole work: Variables are always denoted with capital letters, their realizations with lower case, for* $(Z_1, \dots, Z_N)$ *we write* $Z_{*,N}$ *or* Z_* *(and* $z_{*,N}$ *or* z_* *for the data).*

The density function of a variable Z_n *is called* f_θ*, the likelihood-function* $L(\theta, z_n)$*, and* $h(\theta, z_n) = \ln L(\theta, z_n)$ *is the log-likelihood-function.*

If $\theta \in \mathbb{R}$*, we will write* $h'(\theta, z)$ *instead of* $\nabla_\theta h(\theta, z) = (\frac{\partial}{\partial \theta_1} h(\theta, z), \dots, \frac{\partial}{\partial \theta_q} h(\theta, z))$.

We define a nonfit based on the likelihood-function, see e.g. Mizera [Miz 2002] or Mizera and Müller [MiMu 2004]:

Definition 2.1. *We call* $\theta \in \mathbb{R}^q$ *a* likelihood-nonfit *in* $z_* = (z_1, \dots, z_N)$*, if there exists* $\theta' \neq \theta$ *such that*

$$L(\theta', z_n) > L(\theta, z_n) \text{ for all } n = 1, \dots, N. \tag{2.1}$$

We use an equivalent condition to (2.1):

$$h(\theta', z_n) > h(\theta, z_n) \text{ for all } n = 1, \dots, N. \tag{2.2}$$

Now we define the global likelihood-depth of a parameter in a dataset, similar to Mizera [Miz 2002] and Mizera and Müller [MiMu 2004]:

Definition 2.2. *The* global likelihood-depth *of a parameter* θ *within observations* $z_* = (z_1, \dots, z_N)$ *is the minimal number* m *of* $z_{i_1}, \dots, z_{i_m}$ *that has to be removed so that* θ *becomes a likelihood-nonfit in the remaining data, i.e.* θ *is a likelihood-nonfit within* $\{z_1, \dots, z_N\} \setminus \{z_{i_1}, \dots, z_{i_m}\}$.

In large datasets the calculation of the global likelihood-depth can be complicated. A sufficient condition for θ being a likelihood-nonfit is, that there exists an $u \in \mathbb{R}^q$ such that $\nabla_\theta h(\theta, z_n)^T u > 0$ for all $n = 1, \dots, N$, i.e. the gradients of the log-likelihood-function all lie in one subspace with dimension $\leq q$. This leads to the tangent likelihood-depth as in Mizera and Müller [MiMu 2004] which is easier to handle. Müller, see [Mue 2005], also introduces the simplicial likelihood-depth:

Definition 2.3. *(i) The* (tangent) likelihood-depth *of* $\theta \in \mathbb{R}^q$ *within* $z_* = (z_1, \ldots, z_N)$ *is*

$$d_T(\theta, z_*) := \frac{1}{N} \inf_{u \neq 0, u \in \mathbb{R}^q} \sharp\{n; u^T \nabla_\theta h(\theta, z_n)^T \leq 0\}.$$

Recall that $h(\theta, z) = \ln L(\theta, z) = \ln f_\theta(z)$.

(ii) The simplicial likelihood-depth *of* $\theta \in \mathbb{R}^q$ *within observations* $z_* := (z_1, \ldots, z_N)$ *is defined as*

$$d_S(\theta, z_*) := \binom{N}{q+1}^{-1} \sharp\{\{n_1, \ldots, n_{q+1}\} \subset \{1, \ldots, N\}; d_T(\theta, (z_{n_1}, \ldots, z_{n_{q+1}})) > 0\},$$

i.e. the number of $q+1$*-subsets in which* θ *has a positive tangent likelihood-depth.*

In the parameter-space $\Theta \subset \mathbb{R}^q$, an estimator for the parameter θ can be chosen as the one that has maximum (tangent) likelihood-depth.

We are going to treat especially one- and two-dimensional parameters θ, where the case of two-dimensional parameters θ will be traced back to the one-dimensional case.

If $\theta \in \mathbb{R}$, the depths are calculated by just counting the observations z_n, $n \in \{1, \ldots, N\}$, for which $\frac{\partial}{\partial\theta} h(\theta, z_n) = h'(\theta, z_n)$ is positive, negative and zero respectively. These numbers will be denoted by

$$N^\theta_{pos} := \sharp\{n; h'(\theta, z_n) > 0\}, N^\theta_{neg} := \sharp\{n; h'(\theta, z_n) < 0\} \text{ and } N^\theta_0 := \sharp\{n; h'(\theta, z_n) = 0\}.$$

In this notation we have the following lemma.

Lemma 2.4. *The (tangent) likelihood-depth of* $\theta \in \mathbb{R}$ *in data* z_* *is*

$$d_T(\theta, z_*) = \frac{1}{N} \left(\min(N^\theta_{pos}, N^\theta_{neg}) + N^\theta_0 \right).$$

Proof: See the discussion above this lemma. □

Calculating the simplicial likelihood-depth means determining the tangent depth of each pair of observations (z_{i_1}, z_{i_2}), $i_1, i_2 \in \{1, \ldots, N\}, i_1 \neq i_2$, where the tangent likelihood-depth of θ within two observations x, y is non-zero only if $h'(\theta, x)h'(\theta, y) \leq 0$.

Lemma 2.5. *The simplicial likelihood-depth of* $\theta \in \mathbb{R}$ *in data* $z_* = (z_1, \ldots, z_N)^T$, $z_i \in \mathbb{R}^m, i = 1, \ldots, N$, *is*

$$\begin{aligned} d_S(\theta, z_*) &= \frac{1}{\binom{N}{2}} (N^\theta_{pos} N^\theta_{neg} + N^\theta_{pos} N^\theta_0 + N^\theta_{neg} N^\theta_0 + \binom{N^\theta_0}{2}) \\ &= \frac{2}{N(N-1)} (N^\theta_{pos} N^\theta_{neg} + N^\theta_{pos} N^\theta_0 + N^\theta_{neg} N^\theta_0 + \binom{N^\theta_0}{2}). \end{aligned}$$

Proof: See the lines above this lemma. □

For simplification we assume N to be even in this work. With the last to lemmas it is obvious that the likelihood-depth and the simplicial likelihood-depth are maximized by the parameter θ for which the number of observations $x \in \{z_1, \dots, z_N\}$ with $h'(\theta, x) \geq 0$ is equal to the number of observations $y \in \{z_1, \dots, z_N\}$ with $h'(\theta, y) \leq 0$.

2.1. Estimators

Consider $Z_* = (Z_1, \dots, Z_n)$ i.i.d. with continuous f_θ known except for $\theta \in \mathbb{R}$, resp. later also for $\theta \in \mathbb{R}^q$, $q \geq 1$, but we start with considering one-dimensional θ. We use the likelihood-depth to define a "good" robust estimator for θ. As the depth shall be a measure for how well a parameter fits the data, the first idea is to choose that parameter $\theta \in \Theta$ as an estimator, which has maximum depth.

Definition 2.6. *The* maximum likelihood-depth estimator *is the parameter $\tilde{\theta}$ which has maximum depth , i.e.*

$$\tilde{\theta} \in \arg\max d_T(\theta, z_*),$$

where $x' \in \arg\max_\theta f(x)$ iff $f(x') = \max_{x \in D} f(x)$, D being the domain of f.

We already discussed in the end of the last subsection that the likelihood-depth is maximized by the parameter θ with

$$\sharp\{z_n; h'(\theta, z_n) \geq 0\} = \sharp\{z_n; h'(\theta, z_n) \leq 0\} = \frac{N}{2},$$

when N is considered to be even. As the density function is continuous, the likelihood-depth is (asymptotically, considering $N_0^\theta = 0$) maximized by that parameter $\tilde{\theta}$ for which $N_{pos}^{\tilde{\theta}} = N_{neg}^{\tilde{\theta}}$. The law of large numbers provides $\frac{1}{N} N_{pos}^{\tilde{\theta}} \xrightarrow{\theta} P_\theta(h'(\tilde{\theta}, Z) > 0)$ and $\frac{1}{N} N_{neg}^{\tilde{\theta}} \xrightarrow{\theta} P_\theta(h'(\tilde{\theta}, Z) < 0)$ as $N \to \infty$, if θ denotes the underlying parameter. Hence, the estimator $\tilde{\theta}$ is asymptotically unbiased, if $P_{\tilde{\theta}}(h'(\tilde{\theta}, Z) \geq 0) = \frac{1}{2}$ holds. Unfortunately this is not always the case.

To simplify the presentation we introduce some abbreviations.

Notations. *The set, where $h'(\theta, z) = \frac{\partial}{\partial \theta} \ln L(\theta, z)$ is positive or zero (negative or zero), will be denoted by T_{pos}^θ (T_{neg}^θ), i.e.*

$$T_{pos}^\theta := \{z \in \mathbb{R}^m; h'(\theta, z) \geq 0\} \; (T_{neg}^\theta = \{z \in \mathbb{R}^m; h'(z, \theta) \leq 0\}).$$

With this we can define

$$p_{\theta,\theta'} := P_\theta(T_{pos}^{\theta'}) := P_\theta(Z \in T_{pos}^{\theta'}) = 1 - P_\theta(T_{neg}^\theta) \;\; \textit{and} \;\; p_\theta := p_{\theta,\theta}.$$

The asymptotic mean value for the depth of a parameter θ' in data with underlying distribution θ is the minimum of $p_{\theta,\theta'}$ and $1 - p_{\theta,\theta'}$. If $p_\theta \neq \frac{1}{2}$ holds, then there exists $s(\theta) \neq \theta$ with

$$p_{\theta,s(\theta)} = P_\theta(T_{pos}^{s(\theta)}) = \frac{1}{2}. \tag{2.3}$$

This $s(\theta)$ is the asymptotic value for the parameter with maximum depth in a dataset with density f_θ.

For one-dimensional θ the next proposition shows that the maximum likelihood-depth estimator is a consistent estimator for $s(\theta)$. But before we note it, we arrange some more abbreviations.

Notations. *Let be θ_0 the parameter of the underlying distribution. With $\lambda_N^\pm$ we shorten the part of observations for that $h'(\theta, z)$ is non-negative or non-positive respectively, i.e.*

$$\lambda_N^+(\theta, z_{*,N}) := \tfrac{1}{N}\sharp\{n; h'(\theta, z_n) \geq 0\}, \quad \lambda_N^-(\theta, z_{*,N}) := \tfrac{1}{N}\sharp\{n; h'(\theta, z_n) \leq 0\}.$$

Furthermore $\lambda_{\theta_0}^\pm$ denotes the probability, that for one data $h'(\theta, \cdot)$ is non-negative or non-positive respectively, i.e.

$$\lambda_{\theta_0}^+(\theta) := P_{\theta_0}(h'(\theta, Z) \geq 0) = p_{\theta_0,\theta}, \quad \lambda_{\theta_0}^-(\theta) = P_{\theta_0}(h'(\theta, Z) \leq 0).$$

Because of the continuity of f_θ, it is obvious that $\lambda_{\theta_0}^+(\theta) = 1 - \lambda_{\theta_0}^-(\theta)$.

Proposition 2.7. *Let be*

$$d_T(\theta, Z_{*,N}) = \min\{\lambda_N^+(\theta, Z_{*,N}), \lambda_N^-(\theta, Z_{*,N})\}$$

and $\tilde{\theta}_N(Z_{*,N}) = \arg\max_\theta d_T(\theta, Z_{*,N})$. If $\lambda_N^+(\cdot, Z_{*,N})$ is decreasing, $\lambda_{\theta_0}^+(\cdot)$ is strictly decreasing, $\lambda_N^-(\cdot, Z_{*,N})$ is increasing and $\lambda_{\theta_0}^-(\cdot)$ is strictly increasing with

$$\lambda_{\theta_0}^+(s(\theta_0)) = \frac{1}{2} = \lambda_{\theta_0}^-(s(\theta_0)), \tag{2.4}$$

then $\tilde{\theta}_N(Z_{*,N})$ tends to $s(\theta_0)$ almost surely, as N tends to infinity.

Proof: The strong law of large numbers provides

$$\lim_{N\to\infty} \lambda_N^\pm(\theta, Z_{*,N}) = \lambda_{\theta_0}^\pm(\theta) \text{ almost surely for all } \theta \in \Theta. \tag{2.5}$$

Let be $\varepsilon > 0$. Since $\lambda_{\theta_0}^+$ and $\lambda_{\theta_0}^-$ are monotone functions satisfying (2.4), there exists $\delta > 0$ with

$$\begin{aligned}
\lambda_{\theta_0}^+(s(\theta_0) + \varepsilon) &< \frac{1}{2} - \delta,\\
\lambda_{\theta_0}^+(s(\theta_0) - \varepsilon) &> \frac{1}{2} + \delta,\\
\lambda_{\theta_0}^-(s(\theta_0) + \varepsilon) &> \frac{1}{2} + \delta,\\
\lambda_{\theta_0}^-(s(\theta_0) - \varepsilon) &< \frac{1}{2} - \delta.
\end{aligned}$$

(2.5) implies that

$$A_\varepsilon := \left\{ \omega \in \Omega; \begin{array}{c} |\lambda_N^\pm(s(\theta_0) \pm \varepsilon, Z_{*,N}(\omega)) - \lambda_{\theta_0}^\pm(s(\theta_0) \pm \varepsilon)| < \frac{\delta}{3} \\ \text{and } |\lambda_N^\pm(s(\theta_0), Z_{*,N}(\omega)) - \lambda_{\theta_0}^\pm(s(\theta_0))| < \frac{\delta}{3} \text{ for almost all N} \end{array} \right\}$$

has probability one.

Let $\omega \in A_\varepsilon$. Then there exists N_0, such that for $N \geq N_0$:

$$\begin{aligned} &\lambda_N^+(s(\theta_0) + \varepsilon, Z_{*,N}(\omega)) < \frac{1}{2} - \frac{2}{3}\delta, \\ &\lambda_N^+(s(\theta_0), Z_{*,N}(\omega)) \in \left(\frac{1}{2} - \frac{1}{3}\delta, \frac{1}{2} + \frac{1}{3}\delta\right), \\ &\lambda_N^+(s(\theta_0) - \varepsilon, Z_{*,N}(\omega)) > \frac{1}{2} + \frac{2}{3}\delta, \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} &\lambda_N^-(s(\theta_0) + \varepsilon, Z_{*,N}(\omega)) > \frac{1}{2} + \frac{2}{3}\delta, \\ &\lambda_N^-(s(\theta_0), Z_{*,N}(\omega)) \in \left(\frac{1}{2} - \frac{1}{3}\delta, \frac{1}{2} + \frac{1}{3}\delta\right), \\ &\lambda_N^-(s(\theta_0) - \varepsilon, Z_{*,N}(\omega)) < \frac{1}{2} - \frac{2}{3}\delta. \end{aligned} \tag{2.7}$$

Since $\lambda_N^+(\cdot, Z_{*,N}(\omega))$ and $\lambda_N^-(\cdot, Z_{*,N}(\omega))$ are monotone decreasing resp. increasing we have

$$\begin{aligned} &\lambda_N^+(\theta, Z_{*,N}(\omega)) < \frac{1}{2} - \frac{2}{3}\delta \text{ for all } \theta > s(\theta_0) + \varepsilon, \\ &\lambda_N^-(\theta, Z_{*,N}(\omega)) < \frac{1}{2} - \frac{2}{3}\delta \text{ for all } \theta < s(\theta_0) - \varepsilon. \end{aligned}$$

Thus,

$$\min\{\lambda_N^+(\theta, Z_{*,N}(\omega)), \lambda_N^-(\theta, Z_{*,N}(\omega))\} < \frac{1}{2} - \frac{2}{3}\delta$$

for all $\theta \notin [s(\theta_0) - \varepsilon, s(\theta_0) + \varepsilon]$. On the other hand (2.6) and (2.7) imply

$$\min\{\lambda_N^+(\theta, Z_{*,N}(\omega)), \lambda_N^-(\theta, Z_{*,N}(\omega))\} > \frac{1}{2} - \frac{1}{3}\delta$$

for $\theta = s(\theta_0)$. Therefore $\tilde{\theta}_N(Z_{*,N}(\omega)) \in [s(\theta_0) - \varepsilon, s(\theta_0) + \varepsilon]$ for $N \geq N_0$.

Let $B_\varepsilon := \left\{\omega \in \Omega; \tilde{\theta}_N(Z_{*,N}(\omega)) \in [s(\theta_0) - \varepsilon, s(\theta_0) + \varepsilon] \text{ for almost all } N\right\}$. Hence it follows that

$$1 = P_{\theta_0}(A_\varepsilon) \leq P_{\theta_0}(B_\varepsilon)$$

for all $\varepsilon > 0$. In particular, we have

$$
\begin{aligned}
1 &= \lim_{k\to\infty} P_{\theta_0}\left(B_{\frac{1}{k}}\right) = P_{\theta_0}\left(\bigcap_{k=1}^{\infty} B_{\frac{1}{k}}\right) \\
&= P_{\theta_0}\left(\left\{\omega\in\Omega; \forall k\in\mathbb{N}\ \exists N_0\ \forall N\geq N_0 : |\tilde{\theta}_N(Z_{*,N}(\omega)) - s(\theta_0)| < \frac{1}{k}\right\}\right) \\
&= P_{\theta_0}\left(\left\{\omega\in\Omega; \lim_{N\to\infty}\tilde{\theta}_N(Z_{*,N}(\omega)) = s(\theta_0)\right\}\right),
\end{aligned}
$$

which completes the proof. □

Even if $\lambda_N^{\pm}$ and $\lambda_{\theta_0}^{\pm}$ are not monotone, the next theorem yields the consistency of the estimators. Here we use the definition of outer almost surely convergence, that can be found in Definition A.4 on page 180, and the definition of outer probability, which is given in Definition A.1 on page 179.

Proposition 2.8. *Let be* $\tilde{\theta}_N := \arg\max_\theta(d_T(\theta, z_{*,N}))$ *and*

$\lambda_N^{\pm}(\cdot, Z_{*,N})$ *converge uniformly (outer) almost surely to* $\lambda_{\theta_0}^{\pm}(\cdot)$

as $N\to\infty$*. Furthermore, let* $\lambda_{\theta_0}^{\pm}$ *be such that for all* $\varepsilon > 0$ *there exists* $\delta > 0$*, such that*

$$\lambda_{\theta_0}^{-}(\theta) < \frac{1}{2} - \delta \text{ for } \theta < s(\theta_0) - \varepsilon$$

and

$$\lambda_{\theta_0}^{+}(\theta) < \frac{1}{2} - \delta \text{ for } \theta > s(\theta_0) + \varepsilon$$

(or $\lambda_{\theta_0}^{-}(\theta) < \frac{1}{2} - \delta$ *for* $\theta > s(\theta_0) + \varepsilon$ *and* $\lambda_{\theta_0}^{+}(\theta) < \frac{1}{2} - \delta$ *for* $\theta < s(\theta) - \varepsilon$*), where as before* $s(\theta_0)$ *is such that* $\lambda_{\theta_0}^{+}(s(\theta_0)) = \frac{1}{2} = \lambda_{\theta_0}^{-}(s(\theta_0))$*.*

Then $\tilde{\theta}_N$ *converges to* $s(\theta)$ *(outer) almost surely, as* N *tends to infinity.*

Proof: We assume $\lambda_{\theta_0}^{-}(\theta) < \frac{1}{2}$ for $\theta < s(\theta_0)$ and $\lambda_{\theta_0}^{+}(\theta) < \frac{1}{2}$ for $\theta > s(\theta_0)$. The proof of the other case works analogously. Let be $\varepsilon > 0$ and $\delta > 0$ such that $\lambda_{\theta_0}^{+}(\theta) < \frac{1}{2} - \delta$ for $\theta > s(\theta_0) + \varepsilon$ and $\lambda_{\theta_0}^{-}(\theta) < \frac{1}{2} - \delta$ for $\theta < s(\theta_0) - \varepsilon$. As $\lambda_N^{\pm}$ converges (outer) uniformly to $\lambda_{\theta_0}^{\pm}$,

$$A := \left\{\omega\in\Omega; \sup_\theta |\lambda_N^{\pm}(\theta, Z_{*,N}(\omega)) - \lambda_{\theta_0}^{\pm}(\theta)| < \frac{\delta}{3} \text{ for almost all } N\right\}$$

has outer probability one, i.e. $P_{\theta_0}^*(A) = 1$. Let $\omega\in A$. There exists N_0 such that we have for $N \geq N_0$:

$$\lambda_N^{+}(s(\theta_0), Z_{*,N}(\omega)) \in \left(\frac{1}{2} - \frac{1}{3}\delta, \frac{1}{2} + \frac{1}{3}\delta\right), \tag{2.8}$$

$$\lambda_N^{-}(s(\theta_0), Z_{*,N}(\omega)) \in \left(\frac{1}{2} - \frac{1}{3}\delta, \frac{1}{2} + \frac{1}{3}\delta\right). \tag{2.9}$$

Further, there is N_1 for every θ, such that for all $N \geq N_1$,

$$|\lambda_N^+(\theta, Z_{*,N}(\omega)) - \lambda_{\theta_0}^+(\theta)| < \frac{\delta}{3},$$
$$|\lambda_N^-(\theta, Z_{*,N}(\omega)) - \lambda_{\theta_0}^-(\theta)| < \frac{\delta}{3}.$$

Consequently, for $N \geq \max(N_0, N_1)$, it holds

$$\lambda_N^+(\theta, Z_{*,N}(\omega)) < \frac{1}{2} - \delta + \frac{1}{3}\delta \text{ for all } \theta > s(\theta_0) + \varepsilon$$

and

$$\lambda_N^-(\theta, Z_{*,N}(\omega)) < \frac{1}{2} - \delta + \frac{1}{3}\delta \text{ for all } \theta < s(\theta_0) - \varepsilon.$$

This leads to

$$\min\{\lambda_N^+(\theta, Z_{*,N}(\omega)), \lambda_N^-(\theta, Z_{*,N}(\omega))\} < \frac{1}{2} - \frac{2}{3}\delta$$

for $\theta \notin [s(\theta_0) - \varepsilon, s(\theta_0) + \varepsilon]$.

For $\theta = s(\theta_0)$, (2.8) and (2.9) show

$$\min\{\lambda_N^+(\theta, Z_{*,N}(\omega)), \lambda_N^-(\theta, Z_{*,N}(\omega))\} > \frac{1}{2} - \frac{1}{3}\delta,$$

hence $\tilde{\theta}_N \in [s(\theta_0) - \varepsilon, s(\theta_0) + \varepsilon]$ for $N \geq \max(N_0, N_1)$. Let

$$B_\varepsilon := \left\{\omega \in \Omega; \tilde{\theta}_N(Z_{*,N}(\omega)) \in [s(\theta_0) - \varepsilon, s(\theta_0) + \varepsilon] \text{ for almost all } N\right\}.$$

Thus, it follows that

$$1 = P_{\theta_0}^*(A) \leq P_{\theta_0}^*(B_\varepsilon)$$

for all $\varepsilon > 0$. In particular, we have

$$\begin{aligned}
1 &= \lim_{k\to\infty} P_{\theta_0}^*\left(B_{\frac{1}{k}}\right) = P_{\theta_0}^*\left(\bigcap_{k=1}^{\infty} B_{\frac{1}{k}}\right) \\
&= P_{\theta_0}^*\left(\left\{\omega \in \Omega; \forall k \in \mathbb{N}\ \exists N_0\ \forall N \geq N_0 :\ |\tilde{\theta}_N(Z_{*,N}(\omega)) - s(\theta_0)| < \frac{1}{k}\right\}\right) \\
&= P_{\theta_0}^*\left(\left\{\omega \in \Omega; \lim_{N\to\infty} \tilde{\theta}_N(Z_{*,N}(\omega)) = s(\theta_0)\right\}\right),
\end{aligned}$$

which completes the proof. □

To show uniform convergence of $\lambda_N^{\pm}$ we can use a generalization of the Glivenko-Cantelli-Lemma. This can be found in Appendix A. A special class that fulfills the condition of this Theorem A.8 on page 181 are the Vapnik-Červonenkis classes, or simply VC-classes, see Definition A.11 on page 182 and Theorem A.14 in the Appendix A.

An example for a VC-class is given by $\mathscr{C} = \{T_{pos}^{\theta}; \theta \in \Theta\}$, if $T_{pos}^{\theta} \subset T_{pos}^{\theta'}$ for $\theta' < \theta$:

Proposition 2.9. *Let be $T_{pos}^{\theta} \subset T_{pos}^{\theta'}$ for $\theta' < \theta$. $\mathscr{C} := \{T_{pos}^{\theta}; \theta \in \Theta\}$ is a VC-class. Thus,*

$$\lambda_N^{\pm}(\cdot, Z_{*,N}) \text{ converge uniformly outer almost surely to } \lambda_{\theta_0}^{\pm}(\cdot).$$

Proof: $\mathscr{C}$ shatters no set of two points $\{x_1, x_2\}$, i.e. $V(\mathscr{C}) = 2$: Let be without loss of generality $x_1 \in T_{pos}^{\theta_1}$, $x_2 \in T_{pos}^{\theta_2}$ with $\theta_1 < \theta_2$. Then $x_2 \in T_{pos}^{\theta_2} \subset T_{pos}^{\theta_1}$. Therefore, there exists no $C \in \mathscr{C}$ such that $C \cap \{x_1, x_2\} = \{x_1\}$. □

By this theory we have a method to prove uniform convergence of $\lambda_N^{\pm}$. It is used later on in the case of the Gaussian copula.

When the density-function is not only depending on one but on more parameters, the following lines show how we can prove consistency also in these cases, using the same methods as before. If $\theta = (\theta_1, \ldots, \theta_q)$ is multivariate and the depth can be calculated component wise and the parameter with maximum depth can also be found component wise, i.e. $\tilde{\theta}_N(z_{*,N}) = (\tilde{\theta}_1, \ldots, \tilde{\theta}_q)$ with $\tilde{\theta}_i$ the parameter with maximum likelihood-depth, see Definition 2.6 on page 10, $i = 1, \ldots, q$, we get the following

Proposition 2.10. *Let be $\lambda_N^{i,+}(\theta, z_*)$ the part of observations, such that $\frac{\partial}{\partial \theta_i} \ln f_\theta(\cdot)$ is non-negative and $\lambda_N^{i,-}(\theta, z_*)$ the part of observations, such that $\frac{\partial}{\partial \theta_i} \ln f_\theta(\cdot)$ is non-positive. Further let be*

$$d_T^i(z_{*,N}) = \min\{\lambda_N^{i,+}(\theta_i, z_{*,N}), \lambda_N^{i,-}(\theta_i, z_{*,N})\}$$

the likelihood-depth of θ_i and

$$\lambda_{\theta_0}^{i,+}(\theta_i) = P_{\theta_0}(T_{pos}^{\theta_i}) := P_{\theta_0}(\frac{\partial}{\partial \theta_i} \ln f_\theta(Z) \geq 0),$$

$$\lambda_{\theta_0}^{i,-}(\theta_i) := P_{\theta_0}(T_{neg}^{\theta_i}) = P_{\theta_0}(\frac{\partial}{\partial \theta_i} \ln f_\theta(Z) \leq 0),$$

$i = 1 \ldots, q$. If for all $i = 1, \ldots, q$, $\lambda_N^{i,+}(\cdot, Z_{*,N})$ is decreasing, $\lambda_{\theta_0}^{i,+}(\cdot)$ is strictly decreasing and $\lambda_N^{i,-}(\cdot, Z_{*,N})$ is increasing and $\lambda_{\theta_0}^{i,-}(\cdot)$ is strictly increasing or if $\lambda_N^{i,\pm}$ and $\lambda_{\theta_0}^{i,\pm}$ fulfill the assumptions of Proposition 2.8 and if $\lambda_{\theta_0}^{i,+}(s_i(\theta_0)) = \frac{1}{2} = \lambda_{\theta_0}^{i,-}(s_i(\theta_0))$ with $s(\theta_0) = (s_1(\theta_0), \ldots, s_q(\theta_0))^T$, then $\tilde{\theta}_N^i(Z_{*,N})$ converges to $s_i(\theta_0)$ (outer) almost surely as $N \to \infty$ for all $i = 1, \ldots, q$, i.e.

$$\tilde{\theta}_N(Z_{*,N}) = (\tilde{\theta}_N^1(Z_{*,N}), \ldots, \tilde{\theta}_N^q(Z_{*,N})) \to s(\theta_0), \text{ for } N \to \infty,$$

(outer) almost surely.

Proof: With the strong law of large number we have for $i = 1, \ldots, q$

$$\lim_{N \to \infty} \lambda_N^{i,\pm}(\theta, Z_{*,N}) = \lambda_{\theta_0}^{i,\pm}(\theta)$$

almost surely for all $\theta \in \Theta$. Now use Proposition 2.7 resp. Proposition 2.8 component wise. □

If $s(\theta_0) \neq \theta_0$, we need a correction of the estimator. The next proposition shows, in which cases such a correction exists so that the corrected estimator is still consistent. We use the notations: $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,q})^T, \theta = (\theta_1, \ldots, \theta_q)^T$ and

$$\Lambda(\theta_0, \theta) := \left(\lambda_{\theta_0}^{1,+}(\theta_1) - \frac{1}{2}, \ldots, \lambda_{\theta_0}^{q,+}(\theta_q) - \frac{1}{2}\right)^T.$$

Proposition 2.11. *Assume additionally to the assumptions of the Proposition 2.10 that Λ is continuously differentiable in a neighborhood $\mathcal{N}$ of $(\theta_0, s(\theta_0))$ and that*

$$\frac{\partial}{\partial \theta}\Lambda(\theta, s(\theta_0))\bigg|_{\theta=\theta_0} \quad \text{and} \quad \frac{\partial}{\partial \theta}\Lambda(\theta_0, \theta)\bigg|_{\theta=s(\theta_0)}$$

are regular. Then there exists a neighborhood $\mathcal{U}$ around $s(\theta_0)$ and a neighborhood $\mathcal{V}$ around θ_0 and a continuous $s^{-1} : \mathcal{U} \to \mathcal{V}$ such that

$$s^{-1}(\tilde{\theta}_N(Z_{*,N})) \underset{N\to\infty}{\longrightarrow} \theta_0 \text{ almost surely.}$$

Proof: Proposition 2.10 provides $\Lambda(\theta_0, s(\theta_0)) = 0$ and $\tilde{\theta}_N(Z_{*,N}) \to s(\theta_0)$ almost surely. The implicit function theorem provides a neighborhood $\mathcal{V}$ of θ_0 and a neighborhood $\mathcal{U}$ of $s(\theta_0)$ and unique continuous functions $f : \mathcal{V} \to \mathcal{U}$ and $f^{-1} : \mathcal{U} \to \mathcal{V}$ such that $f(\theta_0) = s(\theta_0)$, $f^{-1}(s(\theta_0)) = \theta_0$, $\Lambda(\theta, f(\theta)) = 0$ for all $\theta \in \mathcal{V}$ and $\Lambda(f^{-1}(\theta), \theta) = 0$ for all $\theta \in \mathcal{U}$. In particular $\Lambda(\theta, f(\theta)) = \Lambda(f^{-1}(f(\theta)), f(\theta))$ for all $\theta \in \mathcal{V}$, such that $f^{-1}(f(\theta)) = \theta$ for all $\theta \in \mathcal{V}$. Since $f(\theta_0) = s(\theta_0)$ we can set $s = f$ on $\mathcal{V}$ and $s^{-1} = f^{-1}$ on $\mathcal{U}$. Hence, s^{-1} is the inverse function of s. Since s^{-1} is continuous we have with $\tilde{\theta}_N(Z_{*,N}) \to s(\theta_0)$, $s^{-1}(\tilde{\theta}_N(Z_{*,N})) \to s^{-1}(s(\theta_0)) = \theta_0$. □

We summarize the results in the following theorem.

Theorem 2.12. *Let be $s^{-1} : \Theta \to \Theta, s^{-1}(s(\theta)) = \theta$ and the conditions of Proposition 2.10 and Proposition 2.11 be satisfied. Then the estimator based on the likelihood-depth given by $\hat{\theta} = s^{-1}(\arg\max d_T(\theta, z_*))$ is a consistent estimator for θ.*

We call this new estimator likelihood-depth estimator (LDE).

2.2. Tests

Let still $Z_1, \ldots, Z_N$ be i.i.d. with (continuous) density function f_θ, $\theta \in \Theta \subset \mathbb{R}$. The aim of this section is to find tests for the one-dimensional parameter θ for hypotheses of type $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \notin \Theta_0$.

The likelihood-depth of θ in $\{z_1, z_2\}$ is asymptotically, assuming N_0^θ=0, recall that N_0^θ denotes the number of observations for that $h'(\theta, \cdot) = 0$, according to Lemma 2.4

$$d_T(\theta, z_* = (z_1, z_2)) = 1_{T_{pos}^\theta}(z_1)1_{T_{neg}^\theta}(z_2) + 1_{T_{neg}^\theta}(z_1)1_{T_{pos}^\theta}(z_2),$$

where 1_A denotes the indicator function of A, namely $1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$. The simplicial likelihood-depth can be written as

$$d_S(\theta, z_*) = \frac{1}{\binom{N}{2}} \sum_{1 \leq n_1 < n_2 \leq N} d_T(\theta, (z_{n_1}, z_{n_2})),$$

thus, it is the U-statistic belonging to the tangent likelihood-depth, which is the symmetric kernel. The distribution of the U-statistic is especially known, if it is non-degenerated. Then, the theorem of Hoeffding, see e.g. Witting, Mülller-Funk [WMF 1995], states:

Theorem 2.13 (Hoeffding). *Let be $X_1, \ldots, X_n$ i.i.d. with distribution P and U_n the U-statistic with a symmetric kernel $\psi \in L_2(P^{(m)} = \otimes P)$ of length m. Then with $\gamma := \mathrm{E}(\psi(X_1, \ldots, X_m))$, $\psi_1(x_1) := \mathrm{E}(\psi(X_1, \ldots, X_m) | X_1 = x_1)$ and $\sigma_1^2 := \mathrm{Var}(\psi_1(X_1))$ (all dependent on F but not on n):*

$$\sqrt{n}(U_n - \gamma) \xrightarrow{\mathscr{D}} X \sim \mathcal{N}(0, m^2 \sigma_1^2).$$

We apply this theorem to define new asymptotic tests with level α for all $\theta \in \Theta$ with $p_\theta = P_\theta(T^\theta_{pos}) \neq \frac{1}{2}$. First, we define a test statistic and show with the help of Hoeffding's theorem, that it is asymptotically normal distributed. Then, we define different tests.

Lemma 2.14. *Let be $\theta \in \Theta$ with $p_\theta \neq \frac{1}{2}$ and*

$$T(\theta, z_*) := \sqrt{N} \frac{\frac{2}{N(N-1)} \sum_{1 \leq n_1 < n_2 \leq N} d_T(\theta, (z_{n_1}, z_{n_2})) - 2p_\theta(1 - p_\theta)}{2\sqrt{(1 - p_\theta)p_\theta(1 - 2p_\theta)^2}}.$$

Then, it holds $T(\theta, Z_{*,N}) \xrightarrow{\mathscr{D}} X \sim \mathcal{N}(0, 1)$.

Proof: Let $\theta \in \Theta$ with $p_\theta \neq \frac{1}{2}$. It holds

$$P_\theta(d_T(\theta, Z_* = (Z_1, Z_2)) = 1) = 2p_\theta(1 - p_\theta),$$
$$P_\theta(d_T(\theta, Z_* = (Z_1, Z_2) = 1 | Z_1 \in T^\theta_{pos}) = P_\theta(Z_2 \in T^\theta_{neg}) = (1 - p_\theta) \neq \tfrac{1}{2},$$
$$P_\theta(d_T(\theta, Z_* = (Z_1, Z_2) = 1 | Z_1 \in T^\theta_{neg}) = P_\theta(Z_2 \in T^\theta_{pos}) = p_\theta \neq \tfrac{1}{2},$$

and therefore

$$P_\theta(d_T(\theta, Z_* = (Z_1, Z_2) = 1 | Z_1 = z_1)) = (1 - p_\theta) 1_{T^\theta_{pos}}(z_1) + p_\theta 1_{T^\theta_{neg}}(z_1) \neq \tfrac{1}{2}$$

with probability one. To show that $T(\theta, Z_{*,N})$ is asymptotically normal distributed, the theorem of Hoeffding is used. As already mentioned, the simplicial depth is a U-statistic with likelihood-depth as kernel. In this situation the emergent quantities are:

$$\begin{aligned}
\psi_\theta(z_1, z_2) &:= 1_{\{d_T(\theta, z_* = (z_1, z_2)) = 1\}}(z_1, z_2) = d_T(\theta, z_* = (z_1, z_2)), \\
\gamma_\theta &:= \mathrm{E}(\psi_\theta(Z_1, Z_2)) = \mathrm{E}(1_{T^\theta_{pos}}(Z_1) 1_{T^\theta_{neg}}(Z_2) + 1_{T^\theta_{neg}}(Z_1) 1_{T^\theta_{pos}}(Z_2)) \\
&= 2p_\theta(1 - p_\theta), \\
\psi_1(z_1) &:= \mathrm{E}(\psi_\theta(Z_1, Z_2) | Z_1 = z_1) = (1 - p_\theta) 1_{T^\theta_{pos}}(z_1) + p_\theta 1_{T^\theta_{neg}}(z_1)
\end{aligned}$$

and

$$
\begin{aligned}
\sigma_\theta^2 &:= \operatorname{Var}(\psi_1(Z_1)) = \operatorname{Var}((1-p_\theta)1_{T^\theta_{pos}}(Z_1) + p_\theta 1_{T^\theta_{neg}}(Z_1)) \\
&= \operatorname{Var}((1-p_\theta)1_{T^\theta_{pos}}(Z_1)) + \operatorname{Var}(p_\theta 1_{T^\theta_{neg}}(Z_1)) \\
&\quad +2\operatorname{Cov}((1-p_\theta)1_{T^\theta_{pos}}(Z_1), p_\theta 1_{T^\theta_{neg}}(Z_1)) \\
&= (1-p_\theta)^3 p_\theta + p_\theta^3(1-p_\theta) + 2p_\theta(1-p_\theta)[\overbrace{\mathrm{E}(1_{T^\theta_{pos}}(Z_1)1_{T^\theta_{neg}}(Z_1))}^{=0} \\
&\quad -\mathrm{E}(1_{T^\theta_{pos}}(Z_1))\mathrm{E}(1_{T^\theta_{neg}}(Z_1))] \\
&= (1-p_\theta)^3 p_\theta + p_\theta^3(1-p_\theta) - 2p_\theta(1-p_\theta)p_\theta(1-p_\theta) \\
&= (1-p_\theta)p_\theta((1-p_\theta)^2 + p_\theta^2 - 2p_\theta(1-p_\theta)) \\
&= (1-p_\theta)p_\theta(1-2p_\theta)^2.
\end{aligned}
$$

The requirements of the theorem of Hoeffding are fulfilled as the U-statistic is not degenerated, because $\psi_1(z_1)$ is not independent of z_1. We get

$$
\sqrt{N}\frac{1}{\binom{N}{2}} \sum_{1\leq n1<n2\leq N} (d_T(\theta,(Z_{n1},Z_{n2})) - \gamma_\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\sigma_\theta^2),
$$

i.e. the test statistic defined as

$$
T(\theta, z_*) = \sqrt{N}\frac{\frac{2}{N(N-1)}\sum_{1\leq n1<n2\leq N} d_T(\theta,(z_{n1},z_{n2})) - \gamma_\theta}{2\sigma_\theta}
$$

is approximately normal distributed with mean $\mu = 0$ and variance $\sigma^2 = 1$. □

Theorem 2.15. *Let be the test statistic $T(\theta, z_*)$ as in Lemma 2.14 and $p_\theta \neq \frac{1}{2}$ for $\theta \in \Theta$. The test*

$$
\varphi(z_*) := 1_{\{\sup_{\theta\in\Theta_0} T(\theta,z_*)<\Phi^{-1}(\alpha)\}}(z_*)
$$

is an asymptotic α-level test for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \notin \Theta_0$.

Proof: Let be $\theta \in \Theta_0$, then

$$
\begin{aligned}
P_\theta(\varphi(Z_{*,N}) = 1) &= P_\theta(\sup_{\tilde\theta\in\Theta_0} T(\tilde\theta, Z_{*,N}) < \Phi^{-1}(\alpha)) \\
&\leq P_\theta(T(\theta, Z_{*,N}) < \Phi^{-1}(\alpha)) \underset{N\to\infty}{\longrightarrow} \Phi(\Phi^{-1}(\alpha)) = \alpha,
\end{aligned}
$$

what shows that φ has asymptotically level α. □

This theorem leads to various tests based on likelihood-depth, which are given in the next two corollaries.

Corollary 2.16. *Let be $p_\theta \neq \frac{1}{2}$, then*

$$\varphi_{\theta_0}^{0,=}(z_*) := 1_{\{T(\theta_0,z_*)<\Phi^{-1}(\alpha)\}}(z_*)$$

satisfies $\lim_{N\to\infty} P_{\theta_0}\left(\varphi_{\theta_0}^{0,=}(Z_{*,N}) = 1\right) \leq \alpha$, *i.e. is an asymptotic α-level test for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.*

Corollary 2.17. *If $p_\theta \neq \frac{1}{2}$ for $\theta \in \Theta$, it holds that*

$$\varphi_{\theta_0}^{0,\geq}(z_*) := 1_{\{\sup_{\theta\geq\theta_0} T(\theta,z_*)<\Phi^{-1}(\alpha)\}}(z_*)$$

is a test with asymptotic level α for $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$. Also

$$\varphi_{\theta_0}^{0,\leq}(z_*) := 1_{\{\sup_{\theta\leq\theta_0} T(\theta,z_*)<\Phi^{-1}(\alpha)\}}(z_*)$$

is an asymptotic α-level test for $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

We obtain asymptotic α-level tests for hypotheses like $H_0 : \theta = \theta_0$ and $H_0 : \theta \leq \theta_0$, if $p_\theta \neq \frac{1}{2}$ for $\theta \in \Theta$. But $p_\theta \neq \frac{1}{2}$ also means that the simplicial depth and thereby the test statistic reach their maximum for $s(\theta) \neq \theta$ with $P_\theta(T_{pos}^{s(\theta)}) = \frac{1}{2} = P_\theta(T_{neg}^{s(\theta)})$; see Section 2.1. If $s(\theta) > \theta$, then we expect that the power of the test for $H_0 : \theta \leq \theta_0$ is quite good, but the test for $H_0 : \theta \geq \theta_0$ may have bad power. Since in the asymptotic case the test statistic is maximal for $s(\theta) > \theta$, it can happen that the hypothesis $H_0 : \theta \geq \theta_0$ is falsely not rejected for $\theta < \theta_0$. For $s(\theta) < \theta$ it is just the other way around, i.e. the power is supposed to be good for $H_0 : \theta \geq \theta_0$ and bad for $H_0 : \theta \leq \theta_0$. A correction of the power, if $s(\theta) > \theta$ of the test for $H_0 : \theta \geq \theta_0$ and if $s(\theta) < \theta$ of the test for $H_0 : \theta \leq \theta_0$ is needed.

The idea to improve the power is to find $c_\alpha^1(\theta_0)$, resp. $c_\alpha^2(\theta_0)$, as the **maximum**, resp. the **minimum**, value θ, such that under θ_0 the probability, that the test statistic takes a value smaller than the α-quantile of the standard normal distribution, is asymptotically smaller than α.

Definition 2.18. *We define*

$$c_\alpha^1(\theta_0) := \max\{\theta; \lim_{N\to\infty} P_{\theta_0}\left(T(\theta, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \alpha\}$$

and

$$\check{c}_\alpha^1(\theta_0) := \min\{\theta; \lim_{N\to\infty} P_{\theta}\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \alpha\},$$

respectively

$$c_\alpha^2(\theta_0) := \min\{\theta; \lim_{N\to\infty} P_{\theta_0}\left(T(\theta, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \alpha\}$$

and

$$\check{c}_\alpha^2(\theta_0) := \max\{\theta; \lim_{N\to\infty} P_{\theta}\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \alpha\}.$$

The next definition presents the new tests.

Definition 2.19. *Let be $p_\theta \neq \frac{1}{2}$ for all $\theta \in \Theta$ and $T(\theta, z_*)$ as in Lemma 2.14. Recall the definition of $\varphi^{0,=}_{\theta_0}(z_*)$, $\varphi^{0,\geq}_{\theta_0}(z_*)$ and $\varphi^{0,\leq}_{\theta_0}(z_*)$ of Corollary 2.16 and Corollary 2.17. If $s(\theta) > \theta$, we use instead of $\varphi^{0,=}_{\theta_0}$ and $\varphi^{0,\geq}_{\theta_0}$ the following two corrected tests:*

$$\varphi^{=}_{\theta_0} := \max\{1_{\{T(\theta_0,z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*), 1_{\{T(c^1_{\frac{\alpha}{2}}(\theta_0),z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*)\}$$

and

$$\varphi^{\geq}_{\theta_0}(z_*) := 1_{\{\sup_{\theta \geq c^1_\alpha(\theta_0)} T(\theta,z_*)<\Phi^{-1}(\alpha)\}}(z_*).$$

If $s(\theta) < \theta$, we use c^2_α to define alternative corrected tests for $\varphi^{0,=}_{\theta_0}$ and $\varphi^{0,\leq}_{\theta_0}$:

$$\varphi^{=}_{\theta_0} := \max\{1_{\{T(\theta_0,z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*), 1_{\{T(c^2_{\frac{\alpha}{2}}(\theta_0),z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*)\}$$

and

$$\varphi^{\leq}_{\theta_0}(z_*) := 1_{\{\sup_{\theta \leq c^2_\alpha(\theta_0)} T(\theta,z_*)<\Phi^{-1}(\alpha)\}}(z_*).$$

The remaining part of this section deals with the asymptotic power of the tests defined above. We make some more assumptions and we start with showing that the corrected tests are asymptotic α-level tests.

Theorem 2.20. *If $p_\theta \neq \frac{1}{2}$ for all $\theta \in \Theta$ and $c^1_\alpha(\cdot)$ is increasing, then the test $\varphi^{\geq}_{\theta_0}$ is an asymptotic α-level test for $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$.*

Proof: Let be $\theta \geq \theta_0$. We get

$$\begin{aligned}
\lim_{N\to\infty} P_\theta\left(\varphi^{\geq}_{\theta_0}(Z_{*,N}) = 1\right) &= \lim_{N\to\infty} P_\theta\left(\sup_{\tilde\theta \geq c^1_\alpha(\theta_0)} T(\tilde\theta, Z_{*,N}) < \Phi^{-1}(\alpha)\right)\\
&\leq \lim_{N\to\infty} P_\theta\left(\sup_{\tilde\theta \geq c^1_\alpha(\theta)} T(\tilde\theta, Z_{*,N}) < \Phi^{-1}(\alpha)\right)\\
&\leq P_\theta\left(T(c^1_\alpha(\theta), Z_{*,N}) < \Phi^{-1}(\alpha)\right)\\
&\leq \alpha.
\end{aligned}$$

□

Theorem 2.21. *If $p_\theta \neq \frac{1}{2}$ for all $\theta \in \Theta$ and $c^2_\alpha(\cdot)$ is increasing, then the test $\varphi^{\leq}_{\theta_0}$ is an asymptotic α-level test for $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.*

Proof: Let be $\theta \leq \theta_0$. Then $c^2_\alpha(\theta) \leq c^2_\alpha(\theta_0)$ and

$$\begin{aligned}
\lim_{N\to\infty} P_\theta\left(\varphi^{\leq}_{\overline{\theta}_0}(Z_{*,N}) = 1\right) &= \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\leq c^2_\alpha(\theta_0)} T(\tilde{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \\
&\leq \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\leq c^2_\alpha(\theta)} T(\tilde{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \\
&\leq \lim_{N\to\infty} P_\theta\left(T(c^2_\alpha(\theta), Z_{*,N}) < \Phi^{-1}(\alpha)\right) \\
&\leq \alpha.
\end{aligned}$$

□

Further we have

Theorem 2.22. *If $p_\theta \neq \frac{1}{2}$ for all $\theta \in \Theta$, then $\varphi^{=}_{\overline{\theta}_0}$ is a test with asymptotic level α.*

Proof: Let be $\varphi^{=}_{\overline{\theta}_0}(z_*) = \max\{1_{\{T(\theta_0,z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*), 1_{\{T(c^1_{\frac{\alpha}{2}}(\theta_0),z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*)\}$. It holds

$$\begin{aligned}
&\lim_{N\to\infty} P_{\theta_0}\left(\varphi^{=}_{\overline{\theta}_0}(Z_{*,N}) = 1\right) \\
&= \lim_{N\to\infty} P_{\theta_0}\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right) \vee T(c^1_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right) \\
&\leq \lim_{N\to\infty} P_{\theta_0}\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right) + \lim_{N\to\infty} P_{\theta_0}\left(T(c^1_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right) \\
&\leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

The proof works analogously for $\varphi^{=}_{\overline{\theta}_0} = \max\{1_{\{T(\theta_0,z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*), 1_{\{T(c^2_{\frac{\alpha}{2}}(\theta_0),z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(z_*)\}$.

□

Next we show that the tests $\varphi^{0,\geq}_{\theta_0}$ for $\theta < s(\theta)$, resp. $\varphi^{0,\leq}_{\theta_0}$ for $\theta > s(\theta)$, and $\varphi^{0,=}_{\theta_0}$ have a bad asymptotic power and that the the asymptotic power of the alternative/corrected tests $\varphi^{\geq}_{\overline{\theta}_0}, \varphi^{\leq}_{\overline{\theta}_0}, \varphi^{=}_{\overline{\theta}_0}$ is really improved in comparison to the old ones and that they are consistent. We make use of the following three lemmas.

Lemma 2.23. *For $Z_1, \ldots, Z_N$ i.i.d., $Z_i \sim F_{\theta'}$, $i = 1, \ldots, N$, $p_{\theta',\theta''} = P_{\theta'}(T^{\theta''}_{pos}) \neq \frac{1}{2}$, it holds*

$$\sqrt{N}\frac{d_S(\theta'', Z_{*,N}) - 2p_{\theta',\theta''}(1 - p_{\theta',\theta''})}{2\sqrt{p_{\theta',\theta''}(1 - p_{\theta',\theta''})(1 - 2p_{\theta',\theta''})^2}} \xrightarrow{\mathscr{D}} X \sim \mathcal{N}(0, 1).$$

Proof: The proof is a direct consequence of the theorem of Hoeffding and the proof of Lemma 2.14. We just have to replace $p_\theta = p_{\theta,\theta}$ by $p_{\theta',\theta''}$. □

Lemma 2.24. *Let $\frac{1}{2} < p \leq \frac{1}{2} + \frac{1}{\sqrt{8}} \approx 0.85$.*

(a) *If* $1-p \leq q \leq p$, *then*

$$q(1-q) \geq p(1-p)$$

and

$$q(1-q)(1-2q)^2 \leq p(1-p)(1-2p)^2.$$

(b) *If* $q \notin [1-p, p]$, *then* $q(1-q) < p(1-p)$.

Proof: (a) Let be $f(x) := x(1-x)$ for $x \in [0,1]$. It is f a parabola that is opened downwards with peak in $(0.5, 0.25)$, see also Figure 2.1. Thus, $f(q) = q(1-q) \geq p(1-p) = f(p)$ for all $q \in \{q; 1-p < q < p\}$, as $p > \frac{1}{2}$.

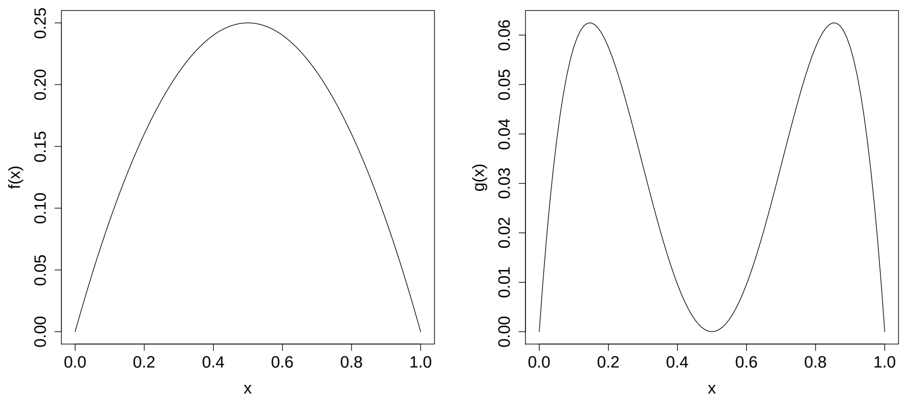

Figure 2.1.: Plot of f and g.

Now let be $g(x) := x(1-x)(1-2x)^2 = x - 5x^2 + 8x^3 - 4x^4$ for $x \in [0,1]$. Then the polynomial g of degree four has at most three extremal points. It reaches local maxima in the points $x_{1/2} = \frac{1}{2} \pm \sqrt{\frac{1}{8}}$ and a local minimum in $x_3 = 0$, see also Figure 2.1 on the right. Henceforth, between $\frac{1}{2} - \sqrt{\frac{1}{8}}$ and 0, g is monotone decreasing and monotone increasing between 0 and $\frac{1}{2} + \sqrt{\frac{1}{8}}$. Thus, $g(q) = q(1-q)(1-2q)^2 < p(1-p)(1-2p)^2 = g(p)$ for q, such that $1-p < q < p$, as $\frac{1}{2} < p \leq \frac{1}{2} + \sqrt{\frac{1}{8}}$

(b) i) Let be $q < 1-p < \frac{1}{2}$, f is monotone increasing for $x < 0.5$, hence, $q(1-q) < (1-p)(1-1+p) = p(1-p)$.

ii) Let be $q > p > 0.5$, f is monotone decreasing for $x > 0.5$, thus, $p(1-p) > q(1-q)$.

□

Under the conditions of the last lemma we can determine c^1_α and $\check{c}^1_\alpha$, respectively c^2_α and $\check{c}^2_\alpha$.

Lemma 2.25. *If $p_{(\cdot),\theta_0} = P_{(\cdot)}(T^{\theta_0}_{pos})$ is strictly increasing from 0 to 1, $p_{\theta_0,(\cdot)}$ strictly decreasing and $\frac{1}{2} < p_{\theta_0} \leq \frac{1}{2} + \frac{1}{\sqrt{8}}$, then for $\alpha < 0.5$, $c^1_\alpha(\theta_0)$ is the value θ, such that*

$$1 - p_\theta = p_{\theta_0,\theta},$$

and $\check{c}^1_\alpha(\theta_0)$ is the value θ, such that

$$1 - p_{\theta_0} = p_{\theta,\theta_0}.$$

In particular we have $c^1_\alpha(\check{c}^1_\alpha(\theta_0)) = \theta_0 = \check{c}^1_\alpha(c^1_\alpha(\theta_0))$.

Proof: We start with determining c^1_α. Therefore we transform $P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha))$ such that Lemma 2.23 can be used, i.e.

$$\begin{aligned}
&P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha)) = P_{\theta_0}\left(\sqrt{N}\frac{d_S(\theta, Z_*) - 2p_\theta(1-p_\theta)}{2\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}} < \Phi^{-1}(\alpha)\right)\\
&= P_{\theta_0}\left(\sqrt{N}\frac{d_S(\theta, Z_*) - 2p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{2\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}\right.\\
&\qquad\left. < \frac{\Phi^{-1}(\alpha)\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}} + \sqrt{N}\frac{p_\theta(1-p_\theta) - p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}\right).
\end{aligned}$$

Since we are looking for θ, such that $\lim_{N\to\infty} P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha)) < \alpha$, we have to ensure that

$$\frac{\Phi^{-1}(\alpha)\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}} + \sqrt{N}\frac{p_\theta(1-p_\theta) - p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}} \not\to \infty$$

as $N \to \infty$. Hence, $\frac{p_\theta(1-p_\theta) - p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}$ should be smaller than or equal to zero, i.e.

$$p_\theta(1-p_\theta) - p_{\theta_0,\theta}(1-p_{\theta_0,\theta}) \leq 0.$$

Lemma 2.24 states that this is true, if and only if $1 - p_\theta \leq p_{\theta_0,\theta} \leq p_\theta$. Let be θ such that these inequalities hold, then using Lemma 2.23 on page 21, we get

$$\begin{aligned}
&P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha))\\
&\leq P_{\theta_0}\left(\sqrt{N}\frac{d_S(\theta, Z_*) - 2p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{2\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}} < \frac{\Phi^{-1}(\alpha)\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}\right)\\
&\overset{N\to\infty}{\longrightarrow} \Phi\left(\frac{\Phi^{-1}(\alpha)\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}\right).
\end{aligned}$$

As $\Phi^{-1}(\alpha) < 0$ for $\alpha < 0.5$ and since we Lemma 2.24 yields:

$$\frac{\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}} \geq 1,$$

it is $\Phi\left(\frac{\Phi^{-1}(\alpha)\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}\right) \leq \alpha$. As $c_\alpha^1(\theta_0)$ is the biggest value θ for that these considerations are true, we have to determine the maximum value θ, such that

$$1 - p_\theta \leq p_{\theta_0,\theta} \leq p_\theta.$$

Since $p_{\theta_0,(\cdot)}$ is decreasing, we obtain $c_\alpha^1(\theta_0)$ as the solution of $p_{\theta_0,\theta} = 1 - p_\theta$ for θ.

We use similar arguments to determine $\check{c}_\alpha^1(\theta_0)$. It holds

$$\begin{aligned}
&P_\theta(T(\theta_0, Z_*) < \Phi^{-1}(\alpha)) \\
&= P_\theta\Bigg(\sqrt{N}\frac{d_S(\theta_0, Z_*) - 2p_{\theta,\theta_0}(1-p_{\theta,\theta_0})}{2\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}} \\
&\qquad < \frac{\Phi^{-1}(\alpha)\sqrt{p_{\theta_0}(1-p_{\theta_0})(1-2p_{\theta_0})^2}}{\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}} + \sqrt{N}\frac{p_{\theta_0}(1-p_{\theta_0}) - p_{\theta,\theta_0}(1-p_{\theta,\theta_0})}{\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}}\Bigg).
\end{aligned}$$

We have to determine θ such that $p_{\theta_0}(1-p_{\theta_0}) \leq p_{\theta,\theta_0}(1-p_{\theta,\theta_0})$. Lemma 2.24 states that this is true, if

$$1 - p_{\theta_0} \leq p_{\theta,\theta_0} \leq p_{\theta_0}. \tag{2.10}$$

In this case it is $\frac{\sqrt{p_{\theta_0}(1-p_{\theta_0})(1-2p_{\theta_0})^2}}{\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}} \geq 1$ and since $\Phi^{-1}(\alpha) < 0$ for $\alpha < 0.5$ we have

$$\begin{aligned}
&P_\theta(T(\theta_0, Z_*) < \Phi^{-1}(\alpha)) \\
&\leq P_\theta\left(\sqrt{N}\frac{d_S(\theta_0, Z_*) - 2p_{\theta,\theta_0}(1-p_{\theta,\theta_0})}{2\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}} < \frac{\Phi^{-1}(\alpha)\sqrt{p_{\theta_0}(1-p_{\theta_0})(1-2p_{\theta_0})^2}}{\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}}\right) \\
&\overset{N\to\infty}{\to} \Phi(\Phi^{-1}(\alpha)\frac{\sqrt{p_{\theta_0}(1-p_{\theta_0})(1-2p_{\theta_0})^2}}{\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}}) \\
&\leq \Phi(\Phi^{-1}(\alpha)) = \alpha.
\end{aligned}$$

As $\check{c}_\alpha^1(\theta_0) = \min\{\theta; \lim_{N\to\infty} P_\theta(T(\theta_0, Z_*) < \Phi^{-1}(\alpha)) \leq \alpha\}$ and $p_{(\cdot),\theta_0}$ is strictly increasing, $\check{c}_\alpha^1$ is the value, such that $p_{\theta,\theta_0} = 1 - p_{\theta_0}$.

If $\tilde{\theta} = \check{c}_\alpha^1(\theta_0)$, then we just showed $1 - p_{\theta_0} = p_{\tilde{\theta},\theta_0}$, i.e. $\theta_0 = c_\alpha^1(\tilde{\theta})$. If it holds $\bar{\theta} = c_\alpha^1(\theta_0)$, then we get $1 - p_{\bar{\theta}} = p_{\theta_0,\bar{\theta}}$ and this is $\theta_0 = \check{c}_\alpha^1(\bar{\theta})$. □

Corollary 2.26. *Under the assumptions of Lemma 2.25, $c_\alpha^1(\cdot)$ is strictly increasing if and only if $\check{c}_\alpha^1(\cdot)$ is strictly increasing.*

Lemma 2.27. *If $p_{(\cdot),\theta_0}$ is strictly increasing from 0 to 1, $p_{\theta_0,(\cdot)}$ is strictly decreasing and $\frac{1}{2} < 1 - p_{\theta_0} \leq \frac{1}{2} + \frac{1}{\sqrt{8}}$, then for $\alpha < 0.5$ $c_\alpha^2(\theta_0)$ is the value θ such that*

$$p_{\theta_0,\theta} = 1 - p_\theta,$$

and $\check{c}_\alpha^2(\theta_0)$ is the value θ, for that holds

$$p_{\theta,\theta_0} = 1 - p_{\theta_0}.$$

In particular it is

$$c_\alpha^2(\check{c}_\alpha^2(\theta_0)) = \theta_0 = \check{c}_\alpha^2(c_\alpha^2(\theta_0)).$$

Proof: We use a similar proof to the one of Lemma 2.25 on page 23. It holds

$$\begin{aligned} &P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha)) \\ &= P_{\theta_0}\left(\sqrt{N}\frac{d_s(\theta, Z_*) - 2p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{2\sqrt{p_{\theta_0,\theta(1-p_{\theta,\theta_0})(1-2p_{\theta_0,\theta})^2}}}\right. \\ &\qquad \left.< \Phi^{-1}(\alpha)\frac{\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}} + \sqrt{N}\frac{p_\theta(1-p_\theta) - p_{\theta_0,\theta}(1-p_{\theta_0,\theta})}{\sqrt{p_{\theta_0,\theta}(1-p_{\theta_0,\theta})(1-2p_{\theta_0,\theta})^2}}\right). \end{aligned}$$

Lemma 2.24 on page 21 yields, by exchanging the roles of p_0 and $1 - p_0$,

$$p(1-p) \geq p_0(1-p_0) \text{ and } p(1-p)(1-2p)^2 \leq (1-p_0)p_0(1-2p_0)^2,$$

for

$$\frac{1}{2} < 1 - p_0 < \frac{1}{2} + \frac{1}{\sqrt{8}}$$

only if $p_0 \leq p \leq 1 - p_0$, as it is $(1 - 2(1 - p_0))^2 = (2p_0 - 1)^2 = (1 - 2p_0)^2$. Hence,

$$\lim_{N\to\infty} P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha)) \leq \alpha,$$

only if θ such that $p_\theta \leq p_{\theta_0,\theta} \leq 1 - p_\theta$. As $p_{\theta_0,\cdot}$ decreasing and

$$c_\alpha^2(\theta_0) = \min\{\theta; \lim_{N\to\infty} P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha)) < \alpha\},$$

it is $c_\alpha^2(\theta_0)$ equal to that value θ, such that $p_{\theta_0,\theta} = 1 - p_\theta$. Analog we show $\check{c}_\alpha^2$ being that value θ, such that $p_{\theta,\theta_0} = 1 - p_\theta$. □

Corollary 2.28. *Under the assumptions of Lemma 2.27 we have $c_\alpha^2(\cdot)$ strictly increasing if and only if $\check{c}_\alpha^2(\cdot)$ is strictly increasing.*

Lemma 2.29. *If c_α^i and $\check{c}_\alpha^i$, $i = 1, 2$, exist for every $\theta \in I$, $I \subset \Theta$ an interval, and are strictly monotone functions, then $c_\alpha^i(\cdot)$ and $\check{c}_\alpha^i(\cdot)$, $i = 1, 2$, are continuous on I.*

Proof: We write c_α and $\check{c}_\alpha$ instead of c_α^i and $\check{c}_\alpha^i$, $i = 1, 2$, because the proof is the same in both cases. Without loss of generality, assume c_α and $\check{c}_\alpha$ to be strictly increasing. Assume further c_α being not continuous. Then there exists $\theta_0 \in I$ and $\varepsilon > 0$, such that for all $\delta > 0$ there exists θ with $|\theta - \theta_0| < \delta$ so that $|c_\alpha(\theta_0) - c_\alpha(\theta)| > \varepsilon$, i.e. there exists a jump in θ_0. Without loss of generality we can assume that the jump is left of θ_0, i.e. $\lim_{\delta \searrow 0} c_\alpha(\theta_0 - \delta) < c_\alpha(\theta_0) - \varepsilon$, as c_α is increasing. Thus, there exists $\tilde{\theta} \in (\lim_{\delta \searrow 0} c_\alpha(\theta_0 - \delta), c_\alpha(\theta_0) - \varepsilon) \neq \emptyset$ for that no θ' can be found, such that $c_\alpha(\theta') = \tilde{\theta}$. But as $c_\alpha(\check{c}_\alpha(\theta)) = \theta$ for all $\theta \in I$, it is $c_\alpha(\theta') = \tilde{\theta}$, with $\theta' = \check{c}_\alpha(\tilde{\theta})$, what contradicts the assumption that c_α is not continuous. The proof that $\check{c}_\alpha$ is continuous works the same way. □

We assume, when analyzing tests for $H_0 : \theta \in \Theta_0$, that c_α^1 and $\check{c}_\alpha^1$, respectively c_α^2 and $\check{c}_\alpha^2$ exist for all $\theta \in \Theta_0$.

Now the power-functions of the various tests are studied more precisely. We start with the situation where $\theta < s(\theta)$.

Theorem 2.30. *If $s(\theta_0) > \theta_0$ holds, $p_{(\cdot),\theta_0}$ is strictly increasing from 0 to 1, $p_{\theta_0,(\cdot)}$ is strictly decreasing and $\frac{1}{2} < p_{\theta_0} \leq \frac{1}{2} + \frac{1}{\sqrt{8}}$, if further $c_\alpha^1(\theta_0) > \theta_0$ and c_α^1 is increasing, it holds*

$$\lim_{N\to\infty} P_\theta\left(\varphi_{\theta_0}^{0,=}(Z_{*,N}) = 1\right) = \lim_{N\to\infty} P_\theta\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}(\alpha)\right)$$
$$= \begin{cases} = 1, & \theta < \check{c}_\alpha^1(\theta_0) \\ = \alpha, & \theta = \check{c}_\alpha^1(\theta_0) \\ = 0, & \check{c}_\alpha^1(\theta_0) < \theta < \theta_0 \\ = \alpha, & \theta = \theta_0 \\ = 1, & \theta > \theta_0 \end{cases}.$$

Proof: We use the definition of $\check{c}_\alpha^1$, the proof of Lemma 2.25 and Corollary 2.16 on page 19: It is $\check{c}_\alpha^1(\theta_0)$ the smallest value θ with

$$\lim_{N\to\infty} P_\theta\left(\varphi_{\theta_0}^{0,=}(Z_{*,N}) = 1\right) \leq \alpha.$$

If $\theta < \check{c}_\alpha^1(\theta_0)$ or $\theta > \theta_0$, the proof of Lemma 2.25 shows $p_{\theta,\theta_0} \notin [1 - p_{\theta_0}, p_{\theta_0}]$ as $p_{(\cdot),\theta_0}$ is strictly increasing, hence, it is with a glance at Lemma 2.24 (b)

$$p_{\theta,\theta_0}(1 - p_{\theta,\theta_0}) < p_{\theta_0}(1 - p_{\theta_0}).$$

Using the proof of Lemma 2.25 again we get for these θ

$$\lim_{N\to\infty} P_\theta(\varphi_{\theta_0}^{0,=}(Z_{*,N}) = 1) = \lim_{N\to\infty} P_\theta(T(\theta_0, Z_{*,N}) < \Phi^{-1}(\alpha))$$
$$= \lim_{N\to\infty} P_\theta\left(\sqrt{N}\frac{d_S(\theta_0, Z_{*,N}) - 2p_{\theta,\theta_0}(1 - p_{\theta,\theta_0})}{2\sqrt{p_{\theta,\theta_0}(1 - p_{\theta,\theta_0})(1 - 2p_{\theta,\theta_0})^2}}\right.$$
$$\left. < \frac{\Phi^{-1}(\alpha)\sqrt{p_{\theta_0}(1 - p_{\theta_0})(1 - 2p_{\theta_0})^2}}{\sqrt{p_{\theta,\theta_0}(1 - p_{\theta,\theta_0})(1 - 2p_{\theta,\theta_0})^2}} + \sqrt{N}\underbrace{\frac{p_{\theta_0}(1 - p_{\theta_0}) - p_{\theta,\theta_0}(1 - p_{\theta,\theta_0})}{\sqrt{p_{\theta,\theta_0}(1 - p_{\theta,\theta_0})(1 - 2p_{\theta,\theta_0})^2}}}_{>0}\right)$$

$$\underset{N\to\infty}{\longrightarrow} \Phi(\infty) = 1.$$

If $\check{c}^1_\alpha(\theta_0) < \theta < \theta_0$, it is $p_{\theta_0}(1-p_{\theta_0}) < p_{\theta,\theta_0}(1-p_{\theta,\theta_0})$ what leads to

$$\sqrt{N}\frac{p_{\theta_0}(1-p_{\theta_0}) - p_{\theta,\theta_0}(1-p_{\theta,\theta_0})}{\sqrt{p_{\theta,\theta_0}(1-p_{\theta,\theta_0})(1-2p_{\theta,\theta_0})^2}} \underset{N\to\infty}{\longrightarrow} -\infty,$$

i.e. $\lim_{N\to\infty} P_\theta\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}(\alpha)\right) = 0$. If $\theta = \check{c}^1_\alpha(\theta_0)$, then $1 - p_{\theta_0} = p_{\theta,\theta_0}$ thus, $\frac{(1-2p_{\theta_0})^2}{(1-p_{\theta,\theta_0})^2} = \frac{(1-2p_{\theta_0})^2}{(2p_{\theta_0}-1)^2} = 1$. Consequently, it holds for $\theta = \check{c}^1_\alpha(\theta_0)$ and for $\theta = \theta_0$:

$$\lim_{N\to\infty} P_\theta(\varphi^{0,=}_{\theta_0}(Z_{*,N}) = 1) = \Phi(\Phi^{-1}(\alpha)) = \alpha.$$

□

Hence the power of $\varphi^{0,=}_{\theta_0}$ is bad for $\theta \in [\check{c}^1_\alpha(\theta_0), \theta_0)$. But it can be used as a test for $H_0 : \theta \in [\check{c}^1_\alpha(\theta_0), \theta_0]$.

Corollary 2.31. *Under the assumptions of Theorem 2.30, the test* $\varphi^{0,=}_{\theta_0}(z_*) = 1_{\{T(\theta_0,z_*)<\Phi^{-1}(\alpha)\}}(z_*)$ *for* $H_0 : \theta \in [\check{c}^1_\alpha(\theta_0), \theta_0]$ *against* $H_1 : \theta \notin [\check{c}^1_\alpha(\theta_0), \theta_0]$ *is a consistent test with asymptotic level* α*.*

The next theorem shows that the power really improves if we use $\varphi^{=}_{\theta_0}$ instead of $\varphi^{0,=}_{\theta_0}$.

Theorem 2.32. *Under the assumptions of Theorem 2.30, we have*

$$\lim_{N\to\infty} P_\theta\left(\max\{1_{\{T(\theta_0,Z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(Z_*), 1_{\{T(c^1_{\frac{\alpha}{2}}(\theta_0),Z_*)<\Phi^{-1}\left(\frac{\alpha}{2}\right)\}}(Z_*)\}\right)$$
$$= \lim_{N\to\infty} P_\theta\left(\varphi^{=}_{\theta_0}(Z_{*,N}) = 1\right)\begin{cases} = 1, & \theta \neq \theta_0 \\ \leq \alpha, & \theta = \theta_0 \end{cases},$$

i.e. $\varphi^{=}_{\theta_0}$ *is a consistent test for* $H_0 : \theta = \theta_0$*.*

Proof: We already proved in Theorem 2.22 that $\lim_{N\to\infty} P_{\theta_0}\left(\varphi^{=}_{\theta_0}(Z_{*,N}) = 1\right) \leq \alpha$. Now let be $\theta \neq \theta_0, \theta \in \Theta$. Using the results of Theorem 2.30 and the fact that $\check{c}^1_{\frac{\alpha}{2}}(c^1_{\frac{\alpha}{2}}(\theta_0)) = \theta_0$, see Lemma 2.25, leads to the following lines:

$$\lim_{N\to\infty} P_\theta\left(\varphi^{=}_{\theta_0}(Z_{*,N}) = 0\right)$$
$$= \lim_{N\to\infty} P_\theta\left(T(\theta_0, Z_{*,N}) \geq \Phi^{-1}\left(\frac{\alpha}{2}\right), T\left(c^1_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}\right) \geq \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)$$
$$\leq \min\{\lim_{N\to\infty} P_\theta\left(T(\theta_0, Z_{*,N}) \geq \Phi^{-1}\left(\frac{\alpha}{2}\right)\right), \lim_{N\to\infty} P_\theta\left(T\left(c^1_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}\right) \geq \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)\}$$
$$= \min\{\lim_{N\to\infty}(1 - P_\theta\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)),$$
$$\lim_{N\to\infty}(1 - P_\theta\left(T\left(c^1_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}\right) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right))\}$$

$$= \begin{cases} 0, & \theta < \check{c}^1_{\frac{\alpha}{2}}(\theta_0) \\ \min\{\lim_N (1 - P_\theta \left(T(\theta_0, Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)), 0\} = 0, & \theta = \check{c}^1_{\frac{\alpha}{2}}(\theta_0) \\ \min\{1, 0\} = 0, & \check{c}^1_{\frac{\alpha}{2}}(\theta_0) < \theta < \theta_0 \\ \min\{0, 1\} = 0, & \theta_0 < \theta < c^1_{\frac{\alpha}{2}}(\theta_0) \\ \min\{0, \lim_N (1 - P_\theta \left(T\left(c^1_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}\right) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right))\} = 0, & \theta = c^1_{\frac{\alpha}{2}}(\theta_0) \\ 0, & \theta > c^1_{\frac{\alpha}{2}}(\theta_0) \end{cases} .$$

Thus, we have for $\theta \neq \theta_0$

$$\lim_{N \to \infty} P_\theta \left(\varphi^{=}_{\theta_0}(Z_{*,N}) = 1\right) = 1 - \lim_{N \to \infty} P_\theta \left(\varphi^{=}_{\theta_0}(Z_{*,N}) = 0\right) \geq 1 - 0 = 1,$$

what proves the claim. □

Theorem 2.33. *If $s(\theta_0) > \theta_0$ holds, $p_{(\cdot),\theta}$ is strictly increasing from 0 to 1, $p_{\theta,(\cdot)}$ strictly decreasing, $\frac{1}{2} < p_\theta < \frac{1}{2} + \frac{1}{\sqrt{8}}$, and $c^1_\alpha(\theta) > \theta$ for all $\theta \leq \theta_0$, if further $c^1_\alpha(\cdot)$ is strictly increasing, $p_\theta \neq \frac{1}{2}$ and additionally p_θ continuous for all $\theta \in \Theta$, then*

$$\lim_{N \to \infty} P_\theta \left(\varphi^{0,\leq}_{\theta_0}(Z_{*,N}) = 1\right) \begin{cases} = 0, & \theta < \theta_0 \\ \leq \alpha, & \theta = \theta_0 \\ = 1, & \theta > \theta_0 \end{cases} .$$

Proof: Let be $\theta < \theta_0$. As $c^1_\alpha(\cdot)$ is increasing, also $\check{c}^1_\alpha(\cdot)$ is increasing. Hence, we find with the help of Lemma 2.29 $\overline{\theta} \leq \theta_0$ such that $\check{c}^1_\alpha(\overline{\theta}) < \theta < \overline{\theta}$ as $\check{c}^1_\alpha(\tilde{\theta}) < \tilde{\theta}$ for all $\tilde{\theta}$. Then we know with Theorem 2.30 that $\lim_{N\to\infty} P_\theta \left(T(\overline{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right) = 0$, i.e. it holds

$$\lim_{N \to \infty} P_\theta \left(\sup_{\tilde{\theta} \leq \theta_0} T(\tilde{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \lim_{N \to \infty} P_\theta \left(T(\overline{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right) = 0.$$

For $\theta = \theta_0$ we already showed in Theorem 2.21 $\lim_{N\to\infty} P_\theta \left(\varphi^{0,\leq}_{\theta_0}(Z_{*,N}) = 1\right) \leq \alpha$. Now let be $\theta > \theta_0$. We prove $\lim_{N\to\infty} P_\theta \left(\sup_{\tilde{\theta}\leq\theta_0} T(\tilde{\theta}, Z_{*,N}) \geq \Phi^{-1}(\alpha)\right) = 0$, where we use that for all z_*

$$\sup_{\tilde{\theta} \leq \theta_0} T(\tilde{\theta}, z_*) = \sup_{\tilde{\theta} \in \{((-\infty,\theta_0]\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} T(\tilde{\theta}, z_*). \tag{2.11}$$

Assume that

$$\sup_{\tilde{\theta} \leq \theta_0} T(\tilde{\theta}, z_*) > \sup_{\tilde{\theta} \in \{((-\infty,\theta_0]\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} T(\tilde{\theta}, z_*).$$

Then there exists $\overline{\theta} \in ((-\infty, \theta_0] \cap \Theta) \setminus \mathbb{Q}$ with

$$T(\overline{\theta}, z_*) = \sup_{\tilde{\theta} \leq \theta_0} T(\tilde{\theta}, z_*).$$

As $h'(\cdot, z)$ is continuous and $d_S(\cdot, z_*)$ is a step function, there exists $\tilde{\theta} \in \mathbb{Q}$ near $\overline{\theta}$, such that $d_S(\overline{\theta}, z_*) = d_S(\tilde{\theta}, z_*)$. Further we have p_θ being continuous and as

$$T(\theta, z_*) = \sqrt{N} \frac{d_S(\theta, z_*) - 2p_\theta(1-p_\theta)}{2\sqrt{(1-p_\theta)p_\theta(1-2p_\theta)^2}},$$

we get for every $\varepsilon > 0$ a $\tilde{\theta} \in \mathbb{Q}$ near $\overline{\theta}$, as $\mathbb{Q}$ is dense in $\mathbb{R}$, such that $|T(\tilde{\theta}, z_*) - T(\overline{\theta}, z_*)| < \varepsilon$. This contradicts our assumption and consequently (2.11) holds. Now let $\theta > \theta_0$. Then

$$\begin{aligned}
&\lim_{N\to\infty} P_\theta\left(\varphi_{\theta_0}^{\leq}(Z_{*,N}) = 0\right) = \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\leq\theta_0} T(\tilde{\theta}, Z_{*,N}) \geq \Phi^{-1}(\alpha)\right) \\
&= \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\in\{((-\infty,\theta_0]\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} T(\tilde{\theta}, z_*) \geq \Phi^{-1}(\alpha)\right) \\
&= \lim_{N\to\infty} P_\theta\left(\bigcup_{\tilde{\theta}\in\{((-\infty,\theta_0]\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} \{\tilde{\theta}; T(\tilde{\theta}, Z_{*,N}) \geq \Phi^{-1}(\alpha)\}\right) \\
&\leq \sum_{\tilde{\theta}\in\{((-\infty,\theta_0]\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} \lim_{N\to\infty} \underbrace{P_\theta(T(\tilde{\theta}, Z_{*,N}) \geq \Phi^{-1}(\alpha))}_{=0,\text{ see Theorem 2.30}} = 0.
\end{aligned}$$

Which leads directly to

$$\lim_{N\to\infty} P_\theta\left(\varphi_{\theta_0}^{0,\leq}(Z_{*,N}) = 1\right) = 1 - \lim_{N\to\infty} P_\theta\left(\varphi^{0,\leq}(Z_{*,N}) = 0\right) \geq 1.$$

□

Theorem 2.34. *Let be $s(\theta_0) > \theta_0$, $p_{(\cdot),\theta}$ strictly increasing from 0 to 1, $p_{\theta,(\cdot)}$ strictly decreasing, $\frac{1}{2} < p_\theta < \frac{1}{2} + \frac{1}{\sqrt{8}}$, $c_\alpha^1(\cdot)$ strictly increasing and $c_\alpha^1(\theta) > \theta$ for all $\theta \geq \theta_0$, let further be $p_\theta \neq \frac{1}{2}$ and additionally p_θ continuous for all $\theta \in \Theta$.*

(a) It holds

$$\lim_{N\to\infty} P_\theta\left(\varphi_{\theta_0}^{0,\geq}(Z_{*,N}) = 1\right) = \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\geq\theta_0} T(\tilde{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right)$$
$$\begin{cases} = 1, & \theta < \check{c}_\alpha^1(\theta_0) \\ \leq \alpha, & \theta = \check{c}_\alpha^1(\theta_0) \\ = 0, & \theta > \check{c}_\alpha^1(\theta_0) \end{cases}.$$

(b) $\varphi_{\theta_0}^{\geq}$ is a consistent test with asymptotic level α for $H_0 : \theta \geq \theta_0$, i.e.

$$\lim_{N\to\infty} P_\theta\left(\varphi_{\theta_0}^{\geq}(Z_{*,N}) = 1\right) = \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\geq c_\alpha^1(\theta_0)} T(\tilde{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right)$$
$$\begin{cases} = 1, & \theta < \theta_0 \\ \leq \alpha, & \theta = \theta_0 \\ = 0, & \theta > \theta_0 \end{cases}.$$

Proof: (a) For $\theta < \check{c}^1_\alpha(\theta_0)$ we prove $\lim_{N\to\infty} P_\theta\left(\sup_{\tilde\theta\geq\theta_0} T(\tilde\theta, Z_{*,N}) \geq \Phi^{-1}(\alpha)\right) = 0$, where we use that for all z_*

$$\sup_{\tilde\theta\geq\theta_0} T(\tilde\theta, z_*) = \sup_{\tilde\theta\in\{([\theta_0,\infty)\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} T(\tilde\theta, z_*).$$

The proof for this works analogously to the proof of (2.11). Let be $\theta < \check{c}^1_\alpha(\theta_0)$. Then

$$\begin{aligned}
&\lim_{N\to\infty} P_\theta\left(\varphi^{0,\geq}_{\theta_0}(Z_{*,N}) = 0\right) = \lim_{N\to\infty} P_\theta\left(\sup_{\tilde\theta\geq\theta_0} T(\tilde\theta, Z_{*,N}) \geq \Phi^{-1}(\alpha)\right) \\
&\leq \sum_{\tilde\theta\in\{([\theta_0,\infty)\cap\Theta)\cap\mathbb{Q}\}\cup\{\theta_0\}} \lim_{N\to\infty} \underbrace{P_\theta(T(\tilde\theta, Z_{*,N}) \geq \Phi^{-1}(\alpha))}_{=0,\text{ see Theorem 2.30}} = 0
\end{aligned}$$

which yields

$$\lim_{N\to\infty} P_\theta\left(\varphi^{0,\geq}_{\theta_0}(Z_{*,N}) = 1\right) = 1 - \lim_{N\to\infty} P_\theta\left(\varphi^{0,\geq}(Z_{*,N}) = 0\right) \geq 1.$$

For $\theta = \check{c}^1_\alpha(\theta_0)$ we apply the definition of $\check{c}^1_\alpha$ and get

$$\begin{aligned}
&\lim_{N\to\infty} P_\theta\left(\varphi^{0,\geq}_{\theta_0}(Z_{*,N}) = 1\right) = \lim_{N\to\infty} P_\theta\left(\sup_{\tilde\theta\geq\theta_0} T(\tilde\theta, Z_{N,*}) < \Phi^{-1}(\alpha)\right) \\
&\leq \lim_{N\to\infty} P_\theta\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \alpha.
\end{aligned}$$

Now consider $\theta > \check{c}^1_\alpha(\theta_0)$. As $\check{c}^1_\alpha$ is strictly increasing, we find $\overline{\theta} \geq \theta_0$ such that $\check{c}^1_\alpha(\overline{\theta}) < \theta < \overline{\theta}$, where the continuity of $\check{c}^1_\alpha$ for $\theta \in \Theta_0$, see Lemma 2.29, is used. Thus, with Theorem 2.30 yields

$$\lim_{N\to\infty} P_\theta\left(\sup_{\tilde\theta\geq\theta_0} T(\tilde\theta, Z_{*,N}) < \Phi^{-1}(\alpha)\right) \leq \lim_{N\to\infty} P_\theta\left(T(\overline{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right) = 0.$$

(b) We use the same arguments as above, as in the proof of Theorem 2.33, and the fact that $\check{c}^1_\alpha(c^1_\alpha(\theta_0)) = \theta_0$. □

Now we study the power-functions of the various tests in the situation where $\theta > s(\theta)$.

Theorem 2.35. *If $\theta_0 > s(\theta_0)$ holds, $p_{(\cdot),\theta_0}$ is strictly increasing from 0 to 1, $p_{\theta_0,(\cdot)}$ is strictly decreasing, $\frac{1}{2} < 1 - p_{\theta_0} < \frac{1}{2} + \frac{1}{\sqrt{8}}$ and $p_\theta \neq \frac{1}{2}$ for all $\theta \in \Theta$ and further, if c^2_α is strictly increasing and $c^2_\alpha(\theta_0) < \theta_0$, then it holds*

(a)

$$\lim_{N\to\infty} P_\theta\left(\varphi^{0,=}_{\theta_0}(Z_{*,N}) = 1\right) \begin{cases} = 1, & \theta < \theta_0 \\ = \alpha, & \theta = \theta_0 \\ = 0, & \theta_0 < \theta < \check{c}^2_\alpha(\theta_0) \\ = \alpha, & \theta = \check{c}^2_\alpha(\theta_0) \\ = 1, & \theta = \check{c}^2_\alpha(\theta_0) \end{cases}$$

and

(b) $\varphi^{=}_{\theta_0}$ *is a consistent test with asymptotic level* α*, i.e.*

$$\lim_{N\to\infty} P_\theta\left(\varphi^{=}_{\theta_0}(Z_{*,N}) = 1\right)$$
$$= \lim_{N\to\infty} P_\theta\left(T(\theta_0, Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right) \vee T(c^2_{\frac{\alpha}{2}}(\theta_0), Z_{*,N}) < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)$$
$$\begin{cases} = 1, & \theta \neq \theta_0 \\ \leq \alpha, & \theta = \theta_0 \end{cases}.$$

Proof: The proof works analogously to the proof of Theorem 2.30, see page 26, respectively the proof of Theorem 2.32 on page 27. □

Corollary 2.36. $\varphi^{0,=}_{\theta_0}$ *is a consistent test with asymptotic level* α *for* $H_0 : \theta \in [\theta_0, \check{c}^2_\alpha(\theta_0)]$.

Analog to the case $s(\theta) > \theta$ we get that $\varphi^{0,\geq}_{\theta_0}$ is a consistent test with asymptotic level α.

Theorem 2.37. *If* $s(\theta_0) < \theta_0$ *holds,* $p_{(\cdot),\theta}$ *is strictly increasing from 0 to 1,* $p_{\theta,(\cdot)}$ *strictly decreasing,* $\frac{1}{2} < p_\theta < \frac{1}{2} + \frac{1}{\sqrt{8}}$*, and* $c^2_\alpha(\theta) < \theta$ *for all* $\theta \geq \theta_0$*, if further* $c^2_\alpha(\cdot)$ *is strictly increasing,* $p_\theta \neq \frac{1}{2}$ *and additionally* p_θ *continuous for all* $\theta \in \Theta$*, it holds*

$$\lim_{N\to\infty} P_\theta\left(\varphi^{0,\geq}_{\theta_0}(Z_{*,N}) = 1\right) \begin{cases} = 1, & \theta < \theta_0 \\ \leq \alpha, & \theta = \theta_0 \\ = 0, & \theta > \theta_0 \end{cases}.$$

Proof: The proof is carried through analogously to the proof of Theorem 2.33 on page 28. □

The next theorem shows that in the case of $s(\theta) < \theta$ the power of the test for $H_0 : \theta \leq \theta_0$ is also improved by the introduction of c^2_α.

Theorem 2.38. *If* $s(\theta_0) < \theta_0$ *holds,* $p_{(\cdot),\theta}$ *is strictly increasing from 0 to 1,* $p_{\theta,(\cdot)}$ *strictly decreasing,* $\frac{1}{2} < p_\theta < \frac{1}{2} + \frac{1}{\sqrt{8}}$*,* $c^2_\alpha(\cdot)$ *is strictly increasing and* $c^2_\alpha(\theta) < \theta$ *for all* $\theta \leq \theta_0$*, if further* $p_\theta \neq \frac{1}{2}$ *and additionally* p_θ *continuous for all* $\theta \in \Theta$*, then it holds*

(a)

$$\lim_{N\to\infty} P_\theta\left(\varphi^{0,\leq}_{\theta_0}(Z_{*,N}) = 1\right) \begin{cases} = 0, & \theta < \check{c}^2_\alpha(\theta_0) \\ \leq \alpha, & \theta = \check{c}^2_\alpha(\theta_0) \\ = 1, & \theta > \check{c}^2_\alpha(\theta_0) \end{cases}$$

and

(b)

$$\lim_{N\to\infty} P_\theta\left(\varphi^{\leq}_{\theta_0}(Z_{*,N}) = 1\right) = \lim_{N\to\infty} P_\theta\left(\sup_{\tilde{\theta}\leq\check{c}^2_\alpha(\theta_0)} T(\tilde{\theta}, Z_{*,N}) < \Phi^{-1}(\alpha)\right)$$
$$\begin{cases} = 0, & \theta < \theta_0 \\ \leq \alpha, & \theta = \theta_0 \\ = 1, & \theta > \theta_0 \end{cases}.$$

Proof: We use analog arguments as in the proof of Theorem 2.34 on page 29. □

In both of the cases $s(\theta) > \theta$ and $s(\theta) < \theta$, we developed tests that have good asymptotic power for different hypotheses.

2.3. Open problems

We introduced the likelihood-depth and estimators based on it. Under some assumptions to $\lambda_N^{\pm}(\theta) = \frac{1}{N}\sharp\{n; \frac{\partial}{\partial\theta} \ln f_\theta(z_n) \gtreqless 0\}$ and $\lambda_{\theta_0}^{\pm}(\theta) = P_{\theta_0}(\frac{\partial}{\partial\theta} \ln f_\theta(Z) \gtreqless 0)$, it was proven that the resulting estimators are consistent. One question could be, if consistency could also be proven under different conditions, which are chosen to be appropriate for other applications.

In the next section, we also defined tests based on likelihood-depth for the parameter of the underlying distribution, when the maximum likelihood-depth estimator is biased. We gave a formula to calculate the correction of the tests by introducing c_α^i, $i = 1, 2$, but only for the case that $p_{\theta,(\cdot)} = \lambda_\theta(\cdot)$ and $p_{(\cdot),\theta} = \lambda_{(\cdot)}(\theta)$ are strictly monotone. Here an open question is, if a rule to determine c_α^i, $i = 1, 2$ can also be found if $p_{(\cdot),\theta}$ and $p_{\theta,(\cdot)}$ are not strictly monotone, as in the case of the two-dimensional normal distribution. As a consequence, proofs for consistency of the tests without assuming monotonicity could be of interest.

Part II.

Application to special distributions

3. Weibull distribution

3.1. Preliminaries

The Weibull distribution $\text{Wei}(a, b), a, b > 0$ is one of the most used distributions in the analysis of lifetime data, where the lifetime describes the time until an event of interest occurs. It is commonly used for the durability analysis of materials or manufactured products and often applied in biological or medical studies.

Let be $T_1, \dots, T_N$ i.i.d. lifetime-variables, $T_i \sim \text{Wei}(a, b)$, $i = 1, \dots, N$. The density of T_i, $i = 1, \dots, N$, for $t \geq 0$ is given by

$$f_{a,b}(t) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1} \exp\left(-\left(\frac{t}{b}\right)^{a}\right),$$

with $a, b > 0$. For $a = 1$ the Weibull distribution is equal to the exponential distribution. The parameter a is called shape parameter and b is called the scale parameter. This can easily be explained, if we take a look at the distribution function for different parameters. In Figure 3.1, on the left we see that the density function changes its shape, if we change the "shape" parameter a. If we consider changes in the "scale" parameter, we see in the right graphic that only the scale on the horizontal axis changes. The survival-function,

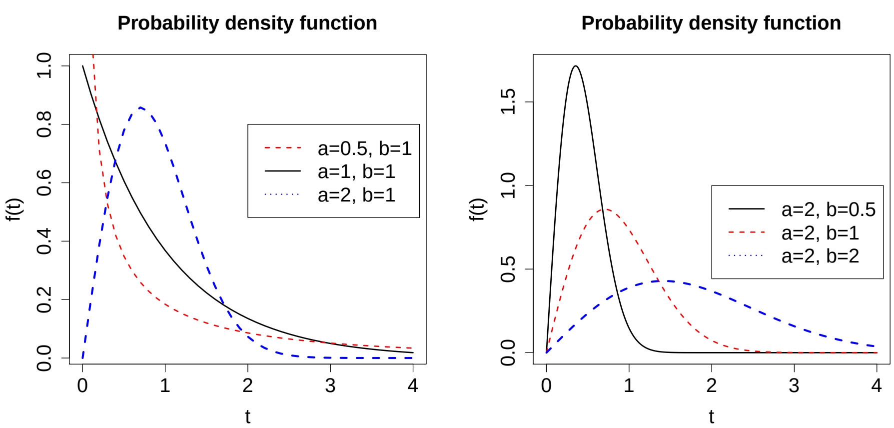

Figure 3.1.: Various density functions of the Weibull distribution.

that gives the probability for one individual to survive time t is for Weibull distributed variables T given by

$$S_{a,b}(t) := P_{a,b}(T \geq t) = 1 - P_{a,b}(T < t) = 1 - F_{a,b}(t) = \exp\left(-\left(\frac{t}{b}\right)^a\right), t \geq 0.$$

For the characterization of lifetime distributions the so-called hazard-function is an important tool. It displays the change in the risk of failure in dependence of the lifetime. For continuous lifetimes it is defined as

$$\lambda(t) := \lim_{\delta \searrow 0} \frac{P(t \leq T \leq t+\delta | T \geq t)}{\delta} = \frac{f(t)}{S(t)}, \quad t \geq 0.$$

For the last equality see for example Lawless [Law 2003], page 9. In case of the Weibull distribution, the hazard function is just determined as

$$\lambda_{a,b}(t) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1}, \quad t \geq 0.$$

Figure 3.2 shows some hazard functions for the Weibull distribution with different shape parameters.

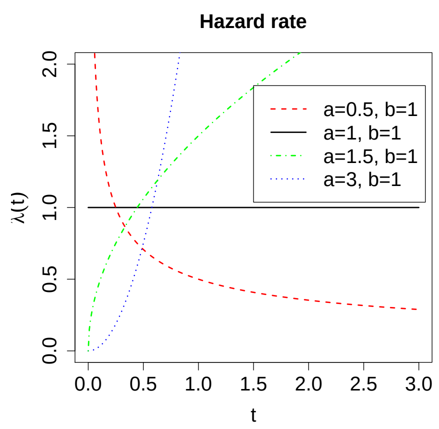

Figure 3.2.: Hazard rates for different shape parameter.

We can simulate increasing ($a > 1$), decreasing ($a < 1$) and constant ($a = 1$) risks of failure with the help of the Weibull distribution, what makes the Weibull model very flexible. This and the rather simple form of the density-, distribution- and hazard-function are the main reasons, why the Weibull distribution is so important in the analysis of survival data.

In studies of lifetimes, it can happen that the real lifetime can not be observed. For example, the lifetime can exceed the study period or an object can not be observed anymore for other reasons than the occurrence of the event of interest. This is called "right-censoring". E.g. people in a clinical study move away or machines can break down for other causes than the studied ones. In this case we only know that the lifetime of this individual has exceeded a certain time, the so-called “censoring-time”. Here, we will only consider type-I right-censoring with one fixed censoring-time. Type-I censoring means, that each individual has a fixed potential censoring time c_i, $i = 1, \dots, N$, and the real lifetime is observed if $T_i \leq c_i$, otherwise we only know $T_i > c_i$. As we will consider one fixed censoring-time c_0, it is $c_1 = \dots = c_N = c_0$ in our case. That means c_0 is the maximum time of study for each individual and we do not consider individuals that fall out of the study for other reasons than the interested ones.

For $i = 1, \dots, N$ we introduce new variables $Y_i := \min(T_i, c_0)$ for the observed lifetime and indicator variables

$$\Delta_i := \begin{cases} 1, & T_i \leq c_0 \\ 0, & T_i > c_0 \end{cases},$$

that indicate, if T_i is censored or not. This leads to variables $Z_i = (Y_i, \Delta_i)$.

The likelihood-function for an observation $z = (y, \delta)$, realization of a variable $Z = (Y, \Delta)$ with distribution f_θ (here $\theta = (a, b)$), from a type-I right-censored sample with censoring-time c_0 is given by

$$L(\theta, z) = f_\theta(y)^\delta S_\theta(c_0)^{1-\delta},$$

see for example Lawless [Law 2003], page 53. This is used in the next sections, when we determine estimators and tests for censored data. In all these section we assume, that less than half the data is censored.

3.2. Estimators for the parameters of the Weibull distribution

We use the likelihood-depth in order to find estimators for the parameters of the Weibull distribution. For the theoretical results see Section 2.1. We start with the estimation for uncensored data and then use the results to find also estimators for type-I right-censored data. Both estimations can be done step by step.

3.2.1. Uncensored data with known shape parameter

In the following let be $T_1, \dots, T_N$ i.i.d. with $T_i \sim \text{Wei}(a, b)$, $i = 1, \dots, N$. With $t_* = t_{*,N} = (t_1, \dots, t_N)$ the realizations of $T_1, \dots, T_N$ are denoted. We start with estimating b and suppose $a = a_0$ to be fixed. In order to determine the likelihood-depth of b in $t_* = (t_1, \dots, t_N)$, $h'_{a_0}(b, \cdot) = \frac{\partial}{\partial b} \ln f_{a_0,b}(\cdot)$ has to be specified.

Lemma 3.1. *For $t \geq 0$, $a_0, b > 0$, it holds*

$$h'_{a_0}(b, t) = -\frac{a_0}{b} + a_0 t^{a_0} \frac{1}{b^{a_0+1}}.$$

Proof: Using the definition of $h'_{a_0}(b,\cdot)$, we obtain

$$\begin{aligned}
h'_{a_0}(b,t) &= \frac{\partial}{\partial b}\ln f_{a_0,b}(t)\\
&= \frac{\partial}{\partial b}\ln\left(\frac{a_0}{b}\left(\frac{t}{b}\right)^{a_0-1}\exp\{-\left(\frac{t}{b}\right)^{a_0}\}\right)\\
&= \frac{\partial}{\partial b}\left(\ln a_0-\ln b+(a_0-1)\ln\frac{t}{b}-\left(\frac{t}{b}\right)^{a_0}\right)\\
&= -\frac{1}{b}-(a_0-1)\frac{b}{t}\frac{t}{b^2}-a_0\left(\frac{t}{b}\right)^{a_0-1}\left(-\frac{t}{b^2}\right)\\
&= -\frac{a_0}{b}+a_0t^{a_0}\frac{1}{b^{a_0+1}}.
\end{aligned}$$

□

Henceforth, we can show that the likelihood-depth of b in $t_*=(t_1,\dots,t_N)$ has a simple form and that it is independent of a_0. Thus, the parameter with maximum depth for b can be determined without knowing a_0.

Theorem 3.2. *The likelihood-depth of b in t_* is the minimum of the number of observations that are smaller than or equal to b and the number of observations that are greater than or equal to b, i.e.*

$$d_T(b,t_*)=\frac{1}{N}\min(\sharp\{n;t_n\le b\},\sharp\{n;t_n\ge b\}).$$

The likelihood-depth is maximized by $\tilde{b}_N=\mathrm{med}(t_1,\dots,t_N)$, where $\mathrm{med}(t_1,\dots,t_N)$ denotes the median of the data $t_1,\dots,t_N$.

Proof: The second claim is easy to see: If the likelihood-depth of b in t_* is given by $\frac{1}{N}\min(\sharp\{n;t_n\le b\},\sharp\{n;t_n\ge b\})$, it is maximized by that parameter for that $\sharp\{n;t_n\le b\}\ge\frac{N}{2}$ and $\sharp\{n;t_n\ge b\}\ge\frac{N}{2}$, what is the definition of the median.

To see that $d_T(b,t_*)=\min(\sharp\{n;t_n\le b\},\sharp\{n;t_n\ge b\})$, we apply the definition of the likelihood-depth, that is

$$d_T(b,t_*)=\min\left(\sharp\{n;h'_{a_0}(b,t_n)\le 0\},\sharp\{n;h'_{a_0}(b,t_n)\ge 0\}\right).$$

Using Lemma 3.1 and $t,b>0$, we have

$$h'_{a_0}(b,t_n)\ge 0\Leftrightarrow -\frac{a_0}{b}+a_0t^{a_0}\frac{1}{b^{a_0+1}}\ge 0\Leftrightarrow t^{a_0}\frac{1}{b^{a_0}}\ge 1\Leftrightarrow t\ge b.$$

Thus, the likelihood-depth of b in t_* is independent of a_0 and maximized by the median of the data. □

Remark 3.3. *Analog to the notations in Section 2.1, it is $T^b_{pos}=\{t;t\ge b\}$, $T^b_{neg}=\{t;t\le b\}$.*

We can find an estimator for b based on likelihood-depth that is independent of a_0. Unfortunately this estimator is biased.

Lemma 3.4. *The median is a biased estimator for the scale parameter of the Weibull distribution.*

Proof: We know from Proposition 2.7 on page 11 that the parameter with maximum likelihood-depth is a consistent estimator, if $s(b) = b$, i.e. $p_b := P_{a_0,b}(T^b_{pos}) = \frac{1}{2}$. But it is easy to show that in our case $p_b \neq \frac{1}{2}$:

$$p_b = P_{a_0,b}(T^b_{pos}) = P_{a_0,b}(T \geq b) = S_{a_0,b}(b) = \exp\left(-\left(\frac{b}{b}\right)^{a_0}\right) = \exp(-1) \approx 0.368. \quad \square$$

Notations. *As $p_b = P_{a_0,b}(T^b_{pos})$ is independent of b, we denote it by p_{scale}.*

We apply Theorem 2.12 on page 16 to derive a consistent estimator for b. To find a correction, we need to solve $P_{a_0,b}(T^{s(b)}_{pos}) = \frac{1}{2}$ for b. This leads to the following estimator.

Theorem 3.5. *Let $\tilde{b}_N$ denote the median of $t_* = (t_1, \ldots, t_N)$. If $a = a_0$ is known, then*

$$\hat{b}_N = \frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{a_0}}}$$

is a strong consistent estimator for the scale parameter of the Weibull distribution $\mathrm{Wei}(a_0, b_0)$.

Proof: Let $\lambda^+_N(b, t_{*,N}) = \frac{1}{N}\sharp\{n; t_n \geq b\}$, $\lambda^-_N(b, t_{*,N}) = \frac{1}{N}\sharp\{t_n \leq b\}$, and

$$\lambda^+_{b_0}(b) = P_{a_0,b_0}(T^b_{pos}) = \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right), \ \lambda^-_{b_0}(b) = P_{a_0,b_0}(T^b_{neg}) = 1 - \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right).$$

Then $\lambda^+_N(\cdot, t_{*,N})$ is decreasing, $\lambda^+_{b_0}(\cdot)$ is strictly decreasing, $\lambda^-_N(\cdot, t_{*,N})$ is increasing, and $\lambda^-_{b_0}(\cdot)$ is strictly increasing. The strong law of large numbers provides

$$\lim_{N\to\infty} \lambda^{\pm}_N(b, T_{*,N}) = \lambda^{\pm}_{b_0}(b) \text{ almost surely for all } b > 0.$$

Moreover,

$$\lambda^+_{b_0}(s(b_0)) = \frac{1}{2} \Leftrightarrow \exp\left(-\left(\frac{s(b_0)}{b_0}\right)^{a_0}\right) = \frac{1}{2} \Leftrightarrow \left(\frac{s(b_0)}{b_0}\right)^{a_0} = \ln 2 \Leftrightarrow s(b_0) = b_0(\ln 2)^{\frac{1}{a_0}}.$$

Theorem 2.7 on page 11 provides $\tilde{b}_N = \arg\max d_T(b, T_{*,N}) \underset{N\to\infty}{\longrightarrow} s(b_0)$ almost surely. It is obvious that $s^{-1}(b) = \frac{b}{(\ln 2)^{\frac{1}{a_0}}}$ is continuous, which yields

$$\frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{a_0}}} = s^{-1}(\tilde{b}_N) \underset{N\to\infty}{\longrightarrow} s^{-1}(s(b_0)) = b_0 \text{ almost surely.}$$

This holds for all $b_0 > 0$, so that $\frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{a_0}}}$ is a strongly consistent estimator for b_0. $\square$

Remark. *The above evolved estimator for the scale parameter coincides with the estimator of He and Fung [HeFu 1999]. They obtain a robust estimator by equating the sample median of* $\frac{\partial}{\partial\theta}\ln(f(t,\theta))$*, where* $\theta = (a,b)$*, with the population median. This procedure is called the "method of medians". It leads to estimators* $\hat{a}_N, \hat{b}_N$ *as the solutions of the two equations below*

$$\operatorname{med}\left(\left(1-\left(\frac{t_1}{\hat{b}_N}\right)^{\hat{a}_N}\right)\ln\left(\frac{t_1}{\hat{b}_N}\right)^{\hat{a}_N},\ldots,\left(1-\left(\frac{t_N}{\hat{b}_N}\right)^{\hat{a}_N}\right)\ln\left(\frac{t_N}{\hat{b}_N}\right)^{\hat{a}_N}\right) = c \tag{3.1}$$

$$\hat{b}_N = \frac{\operatorname{med}(t_1,\ldots,t_N)}{(\ln 2)^{\frac{1}{\hat{a}_N}}}, \tag{3.2}$$

where $c = \operatorname{med}((1-Y)\ln Y) \approx -0.51$ *and* Y *has an exponential distribution.*

The maximum likelihood-depth estimator for b is independent of a_0, but the correction function of the bias is not. If a and b are both unknown, we can base the estimation of a only on the biased estimator for b, because, as the next section shows, the depth of a is depending on b.

3.2.2. Uncensored data with known scale parameter

When we determine the estimator for a based on likelihood-depth, we start with considering $b = b_0$ as to be known. In order to calculate the depth of a in $t_* = (t_1,\ldots,t_N)$ we need to determine $h'_{b_0}(a,\cdot) = \frac{\partial}{\partial a}h_{b_0}(a,\cdot)$.

Lemma 3.6. *The partial derivative of the log-likelihood-function with respect to* a *is*

$$h'_{b_0}(a,t) = \frac{1}{a} + \ln\frac{t}{b_0} - \ln\left(\frac{t}{b_0}\right)\left(\frac{t}{b_0}\right)^a.$$

Proof: To prove the claim, we have to differentiate the log-likelihood-function with respect to a.

$$\begin{aligned} h'_{b_0}(a,t) &= \frac{\partial}{\partial a}\ln f_{a,b_0}(t) \\ &= \frac{\partial}{\partial a}\left(\ln a - \ln b_0 + (a-1)\ln\left(\frac{t}{b_0}\right) - \left(\frac{t}{b_0}\right)^a\right) \\ &= \frac{1}{a} + \ln\left(\frac{t}{b_0}\right) - \left(\frac{t}{b_0}\right)^a \ln\left(\frac{t}{b_0}\right). \end{aligned}$$

□

In order to determine the depth, we identify for which $t > 0$ the function $h'_{b_0}(a,t)$ is positive or zero. It is shown that the depth of a in t_* is not independent of b_0. The following lemma will is used to determine the zeros of h'_{b_0}.

Lemma 3.7. *Let be $f(c) = (c-1)\ln(c)$ for $c > 0$. Then it holds $f(c) = 1$ for exact two points c_1, c_2 with $0.259 \approx c_1 < 1 < c_2 \approx 2.240$.*

Proof: The derivative of f, $f'(c) = \ln c + \frac{c-1}{c}$, is positive for $c > 1$, zero for $c = 1$, and negative for $c < 1$. Therefore, f is strictly decreasing for $c < 1$ and strictly increasing for $c > 1$. As $f(\frac{1}{\exp(1)^2}) = 2(1 - \frac{1}{\exp(1)^2}) > 1$ and $\frac{1}{\exp(1)^2} < 1$, $f(1) = 0$, and f being continuous (and strictly decreasing), there exists only one $c_1 < 1$ with $f(c_1) = 1$. On the other hand, it is $f(1) = 0$ and $f(\exp(1)) = \exp(1) - 1 > 1$, $\exp(1) > 1$. Hence, because of the continuity and the monotonicity of f, there exists exactly one $c_2 > 1$, such that $f(c_2) = 1$ holds. □

Lemma 3.8. *For $t \geq 0$, $a > 0$, it holds that $h'_{b_0}(a,t)$ is positive or zero, if and only if*

$$t_{01}^{a,b_0} := c_1^{\frac{1}{a}} b_0 \leq t \leq c_2^{\frac{1}{a}} b_0 =: t_{02}^{a,b_0},$$

where $0.259 \approx c_1 < 1 < c_2 \approx 2.240$ and c_1, c_2 being the solutions of $\ln c = \frac{1}{c-1}$.

Proof: Let be $x := \frac{t}{b_0}$. We have to solve

$$\frac{1}{a} + \ln x - x^a \ln x = 0$$

for $x > 0$. This equation is equivalent to $(1 - x^a)\ln x = -\frac{1}{a}$. Let be $c > 0$ the such that $\ln c = \frac{1}{c-1}$. Then we get:

$$\ln c = \frac{1}{c-1} \Leftrightarrow (1-c)\ln(c)\frac{1}{a} = -\frac{1}{a} \Leftrightarrow (1-c)\ln(c^{\frac{1}{a}}) = -\frac{1}{a}.$$

$x = c^{\frac{1}{a}}$ solves $(1 - x^a)\ln x = -\frac{1}{a}$. Lemma 3.7 shows that there exist only two c_1, c_2 with $\ln c = \frac{1}{c-1}$. This leads to the zeros of $h'_{b_0}(a,\cdot) = \frac{1}{a} + \ln\left(\frac{\cdot}{b_0}\right)\left(1 - \left(\frac{\cdot}{b_0}\right)^a\right)$, which are the points $t_{01}^{a,b_0} = c_1^{\frac{1}{a}} b_0$ and $t_{02}^{a,b_0} = c_2^{\frac{1}{a}} b_0$. Since $c_1 < 1 < c_2$, we obtain $t_{01}^{a,b_0} < b_0 < t_{02}^{a,b_0}$. Because $h'_{b_0}(a, b_0) = \frac{1}{a}$ is positive for all $a > 0$, we end up with

$$h'_{b_0}(a,t) \geq 0 \Leftrightarrow t_{01}^{a,b_0} \leq t \leq t_{02}^{a,b_0}.$$

□

We determine the likelihood-depth of a in data t_*. The last lemma shows that it depends on b_0.

Corollary 3.9. *The likelihood-depth of a in $t_* = (t_1, \ldots, t_N)$ is calculated as*

$$d_T^{b_0}(a, t_*) = \frac{1}{N}\min\left(\sharp\left\{n; t_{01}^{a,b_0} \leq t_n \leq t_{02}^{a,b_0}\right\}, \sharp\left\{n; t_n \leq t_{01}^{a,b_0} \text{ or } t_n \geq t_{02}^{a,b_0}\right\}\right).$$

Again we have to check, if the maximum likelihood-depth estimator for a is biased or not. The next lemma shows, it is.

Lemma 3.10. *The maximum likelihood-depth estimator for the shape parameter of the Weibull distribution is a biased estimator.*

Proof: We have to show that $p_a^{b_0} := P_{a,b_0}(T_{pos}^{a,b_0}) \neq \frac{1}{2}$. Using the results from the last corollary yields

$$\begin{aligned}
P_{a,b_0}(T_{pos}^{a,b_0}) &= \int_{t_{01}^{a,b_0}}^{t_{02}^{a,b_0}} \frac{a}{b_0}\left(\frac{t}{b_0}\right)^{a-1} \exp\left(-\left(\frac{t}{b_0}\right)^a\right) dt \\
&= -\exp\left(-\left(\frac{t_{02}^{a,b_0}}{b_0}\right)^a\right) + \exp\left(-\left(\frac{t_{01}^{a,b_0}}{b_0}\right)^a\right) \\
&= -\exp\left(-\left(\frac{c_2^{\frac{1}{a}} b_0}{b_0}\right)^a\right) + \exp\left(-\left(\frac{c_1^{\frac{1}{a}} b_0}{b_0}\right)^a\right) \\
&= \exp(-c_1) - \exp(-c_2) \neq \frac{1}{2}.
\end{aligned}$$

The last inequality can be seen, if we plug in the numerical values for c_1, c_2, being $c_1 \approx 0.259$ and $c_2 \approx 2.240$. □

The next step is to determine the correction function s^{-1} for the bias. This will be shown to be independent of b. One has to consider two cases. The first one is $b = b_0$ known and the second one that b is unknown and can only be estimated by the biased median $\tilde{b}_N$. We also show that we obtain a consistent estimation procedure. In order to prepare the next theorem, we proof

Lemma 3.11. *If $c_1 < 1 < c_2$, $b_0 > 0$, $a_1 < a_2$, then*

$$\left\{t \in \mathbb{R}; c_1^{\frac{1}{a_2}} b_0 \leq t \leq c_2^{\frac{1}{a_2}} b_0\right\} \subsetneq \left\{t \in \mathbb{R}; c_1^{\frac{1}{a_1}} b_0 \leq t \leq c_2^{\frac{1}{a_1}} b_0\right\}.$$

Proof: Since $a_1 < a_2$ we have $-\frac{1}{a_1} < -\frac{1}{a_2}$. Thus,

$$c_1^{\frac{1}{a_1}} = \exp\left(\frac{1}{a_1} \underbrace{\ln(c_1)}_{<0}\right) < \exp\left(\frac{1}{a_2}\ln(c_1)\right) = c_1^{\frac{1}{a_2}}$$

and

$$c_2^{\frac{1}{a_1}} = \exp\left(\frac{1}{a_1} \underbrace{\ln(c_2)}_{>0}\right) > \exp\left(\frac{1}{a_2}\ln(c_2)\right) = c_2^{\frac{1}{a_2}}.$$

This yields $c_1^{\frac{1}{a_1}} b_0 < c_1^{\frac{1}{a_2}} b_0 < c_2^{\frac{1}{a_2}} b_0 < c_2^{\frac{1}{a_1}} b_0$. □

We develop a new consistent estimator for the shape parameter.

Theorem 3.12. *If $b = b_0$ is known, the correction of the bias for the maximum likelihood-depth estimator for a is*

$$s^{-1}(a) = \kappa \cdot a,$$

where κ is the unique solution of $\exp\left(-c_1^{\kappa}\right) - \exp\left(-c_2^{\kappa}\right) = \frac{1}{2}$. It is $\kappa \approx 0.691$. We obtain that

$$\hat{a}_N = \arg\max_{a>0} d_T(a, t_{*,N}) \cdot \kappa$$

is a strong consistent estimator for a.

Proof: We define

$$\lambda_N^+(a, t_{*,N}) := \frac{1}{N} \sharp\{n; c_1^{\frac{1}{a}} b_0 \leq t_n \leq c_2^{\frac{1}{a}} b_0\},$$

and

$$\lambda_N^-(a, t_{*,N}) := \frac{1}{N} \sharp\{n; t_n \leq c_1^{\frac{1}{a}} b_0 \vee t_n \geq c_2^{\frac{1}{a}} b_0\},$$

further let be

$$\lambda_{a_0}^+(a) := P_{a_0,b_0}(T_{pos}^a) = \exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-c_2^{\frac{a_0}{a}}\right).$$

Then $\lambda_{a_0}^-(a) := P_{a_0,b_0}(T_{neg}^a) = 1 - \lambda_{a_0}^+(a)$. According to Lemma 3.11 we have $\lambda_N^+(\cdot, t_{*,N})$ decreasing, $\lambda_{a_0}^+(\cdot)$ strictly decreasing, $\lambda_N^-(\cdot, t_{*,N})$ increasing, and $\lambda_{a_0}^-(\cdot)$ strictly increasing. The strong law of large numbers provides

$$\lim_{N\to\infty} \lambda_N^{\pm}(a, T_{*,N}) = \lambda_{a_0}^{\pm}(a) \text{ almost surely.}$$

Solving $\lambda_{a_0}^+(s(a_0)) = \exp\left(-c_1^{\frac{a_0}{s(a_0)}}\right) - \exp\left(-c_2^{\frac{a_0}{s(a_0)}}\right) = \frac{1}{2}$, we receive $\frac{a_0}{s(a_0)} = \kappa$, i.e. $s(a_0) = \frac{a_0}{\kappa}$ and $s^{-1}(a_0) = \kappa \cdot a_0$. Theorem 2.7 on page 11 provides

$$\tilde{a}_N = \arg\max d_T(a, T_{*,N}) \underset{N\to\infty}{\longrightarrow} s(a_0) \text{ almost surely.}$$

Since s^{-1} is continuous, we have immediately

$$\hat{a}_N = \kappa \cdot \tilde{a}_N = \kappa \arg\max_a d_T(a, T_{*,N}) = s^{-1}(\tilde{a}_N) \underset{N\to\infty}{\longrightarrow} s^{-1}(s(a_0)) = a_0$$

almost surely. This holds for all $a_0 > 0$ so that $\kappa \arg\max_a d_T(a, t_{*,N})$ is a strongly consistent estimator for a. □

3.2.3. Uncensored data, shape and scale parameter unknown

Next we consider the case, when both parameters are unknown. It is possible to find the maximum likelihood-depth estimator for b without knowing a but it can not be corrected, if a is unknown. Hence, we base our estimation for a on the biased estimator for b, the median $\tilde{b}_N$. Thus, we have to calculate the correction of the bias of the maximum likelihood-depth estimator for a based on $\tilde{b}_N$. In order to prove consistency of this corrected estimator, the following lemma is used.

Lemma 3.13. *If $\tilde{b}_N$ is the median of $t_{*,N}$, then $\frac{1}{N}\sharp\left\{n; c_1^{\frac{1}{a}}\tilde{b}_N \le t_n \le c_2^{\frac{1}{a}}\tilde{b}_N\right\}$ converges almost surely under* Wei(a_0, b_0) *to* $2^{\left(-c_1^{\frac{a_0}{a}}\right)} - 2^{\left(-c_2^{\frac{a_0}{a}}\right)}$.

Proof: Let be $\varepsilon > 0$. Choose $\delta > 0$ such that

$$\begin{aligned}&\Big|\exp\left(-c_1^{\frac{a_0}{a}}\ln(2)(1-\delta)^{a_0}\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln(2)(1+\delta)^{a_0}\right)\\ &\quad - \left(\exp\left(-c_1^{\frac{a_0}{a}}\ln(2)\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln(2)\right)\right)\Big| < \frac{\varepsilon}{2}\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}&\Big|\exp\left(-c_1^{\frac{a_0}{a}}\ln(2)\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln(2)\right)\\ &\quad - \left(\exp\left(-c_1^{\frac{a_0}{a}}\ln(2)(1+\delta)^{a_0}\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln(2)(1-\delta)^{a_0}\right)\right)\Big| < \frac{\varepsilon}{2.}\end{aligned} \tag{3.4}$$

Since $\tilde{b}_N$ converges almost surely toward the median of the Weibull distribution $b_0(\ln 2)^{\frac{1}{a_0}}$ according to Theorem 3.5, we have $P_{a_0,b_0}(A_\delta) = 1$ for

$$A_\delta := \{\omega; |\tilde{b}_N(T_{*,N}(\omega)) - b_0(\ln 2)^{\frac{1}{a_0}}| < b_0(\ln 2)^{\frac{1}{a_0}}\delta \text{ for almost all } N\}.$$

Let be $\omega \in A_\delta$, then there exists N_0 such that for all $N \ge N_0$ it holds

$$-b_0(\ln 2)^{\frac{1}{a_0}}\delta \le \tilde{b}_N(T_{*,N}(\omega)) - b_0(\ln 2)^{\frac{1}{a_0}} \le b_0(\ln 2)^{\frac{1}{a_0}}\delta,$$

i.e.

$$b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta) \le \tilde{b}_N(T_{*,N}(\omega)) \le b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta).$$

Thus, we have for $\omega \in A_\delta$

$$\begin{aligned}&\frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\}\\ &\le \frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta)\}\\ &= \frac{1}{N}\sum_{n=1}^{N} 1_{[c_1^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta),c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta)]}(T_{n_N}(\omega)),\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\} \\
&\ge \frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta)\} \\
&= \frac{1}{N}\sum_{n=1}^{N} 1_{[c_1^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta), c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta)]}(T_{n_N}(\omega)).
\end{aligned}$$

The strong law of large numbers provides that there exists $N_1 \ge N_0$, such that

$$\begin{aligned}
&\Bigg|\frac{1}{N}\sum_{n=1}^{N} 1_{[c_1^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1\pm\delta), c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1\mp\delta)]}(T_{n_N}(\omega)) \\
&- P_{a_0,b_0}\left(c_1^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1\pm\delta) \le T_1 \le c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1\mp\delta)\right)\Bigg| < \frac{\varepsilon}{2}
\end{aligned}$$

for $N \ge N_1$. Hence, for $N \ge N_1$ we obtain

$$\begin{aligned}
&\frac{1}{N}\sharp\left\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\right\} \\
&\le P_{a_0,b_0}\left(c_1^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta) \le T_1(\omega) \le c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta)\right) + \frac{\varepsilon}{2} \\
&= \exp\left(-\left(\frac{c_1^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1-\delta)}{b_0}\right)^{a_0}\right) - \exp\left(-\left(\frac{c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}}(1+\delta)}{b_0}\right)^{a_0}\right) + \frac{\varepsilon}{2} \\
&\le \exp\left(-c_1^{\frac{a_0}{a}}\ln 2\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln 2\right) + \varepsilon,
\end{aligned}$$

where (3.3) is used. Analogously with (3.4) we obtain

$$\begin{aligned}
&\frac{1}{N}\sharp\left\{c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\right\} \\
&\ge \exp\left(-c_1^{\frac{a_0}{a}}\ln(2)(1+\delta)^{a_0}\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln(2)(1-\delta)^{a_0}\right) - \frac{\varepsilon}{2} \\
&\ge \exp\left(-c_1^{\frac{a_0}{a}}\ln(2)\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln(2)\right) - \varepsilon.
\end{aligned}$$

This implies for $N \ge N_1$

$$\left|\frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\} - \left(2^{-c_1^{\frac{a_0}{a}}} - 2^{-c_2^{\frac{a_0}{a}}}\right)\right| < \varepsilon,$$

so that, with

$$\begin{aligned}
B_\varepsilon := \Bigg\{\omega; \Bigg|&\frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\} \\
&- \left(2^{-c_1^{\frac{a_0}{a}}} - 2^{-c_2^{\frac{a_0}{a}}}\right)\Bigg| < \varepsilon \text{ for } N \text{ large enough}\Bigg\},
\end{aligned}$$

we have $1 = P(A_\delta) \leq P(B_\varepsilon)$ and therefore

$$1 = \lim_{k\to\infty} P_{a_0,b_0}(B_{\frac{1}{k}}) = P_{a_0,b_0}\left(\bigcap_{k=1}^{\infty} B_{\frac{1}{k}}\right)$$
$$= P_{a_0,b_0}\left(\{\omega; \lim_{N\to\infty} \frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}) \leq T_{n_N}(\omega) \leq c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega))\} = 2^{-c_1^{\frac{a_0}{a}}} - 2^{-c_2^{\frac{a_0}{a}}}\right).$$

This proves the claim. □

Lemma 3.14. *Similar to the last lemma, it holds that* $\frac{1}{N}\sharp\left\{n; t_n \leq c_1^{\frac{1}{a}}\tilde{b}_N \vee t_n \geq c_2^{\frac{1}{a}}\tilde{b}_N\right\}$ *converges almost surely under* $\text{Wei}(a_0, b_0)$ *to* $1 - \exp\left(-c_1^{\frac{a_0}{a}} \ln 2\right) + \exp\left(-c_2^{\frac{a_0}{a}} \ln 2\right)$.

The last two lemmas are used to prove consistency of the estimators for the parameters a and b of the Weibull distribution based on likelihood-depth. We can estimate the parameters one after the other. We start with estimating b as the maximum likelihood-depth estimator which is equal to the median. Then we determine the maximum likelihood-depth estimator for a based on the estimator for b. The correction of the bias of the estimator for a is independent of b, so we can correct a. With this corrected estimator we can correct the bias of the estimator for b. The correction formula for a, given in the next procedure, will be proved to be the right one in Theorem 3.16.

Procedure 3.15. *Let be* $t_* = (t_1, \ldots, t_N)$ *realizations of i.i.d.* $T_* = (T_1, \ldots, T_N)$, $T_i \sim \text{Wei}(a, b)$, *$a$ and b unknown. We get unbiased estimators for the parameters of the Weibull distribution based on likelihood-depth (LDE) by following the steps below:*

1. *Determine* $\tilde{b}_N = \text{med}(t_1, \ldots, t_N)$.
2. *Identify* $\tilde{a}_N \in \arg\max_{a>0} d_T^{\tilde{b}_N}(a, t_*)$.
3. *An estimator for a is* $\hat{a}_N = \kappa_1\tilde{a}_N$, *where κ_1 is, analog to the case when b_0 is known, the solution of* $2^{-c_1^{\kappa_1}} - 2^{-c_2^{\kappa_1}} = \frac{1}{2}$, $\kappa_1 \approx 0.757$.
4. *An estimator for b is given by* $\hat{b}_N = \dfrac{\tilde{b}_N}{(\ln 2)^{\frac{1}{\hat{a}_N}}}$.

The algorithm was implemented in R, [R 2009]. The source code can be found in Appendix B.1. This procedure yields a consistent estimator, as the next theorem shows.

Theorem 3.16. *If (a, b) are both unknown, the estimator $(\hat{a}_N, \hat{b}_N)$ given by Procedure 3.15 is a strongly consistent estimator for (a, b).*

Proof: Set

$$\lambda_N^{1,+}(a, t_{*,N}) := \tfrac{1}{N}\sharp\left\{n; c_1^{\frac{1}{a}}\tilde{b}_N \leq t_n \leq c_2^{\frac{1}{a}}\tilde{b}_N\right\},$$
$$\lambda_N^{2,+}(b, t_{*,N}) := \tfrac{1}{N}\sharp\left\{n; t_n \geq b\right\},$$
$$\lambda_N^{1,-}(a, t_{*,N}) := \tfrac{1}{N}\sharp\left\{n; t_n \leq c_1^{\frac{1}{a}}\tilde{b}_N \vee t_n \geq c_2^{\frac{1}{a}}\tilde{b}_N\right\},$$

and

$$\lambda_N^{2,-}(b, t_{*,N}) := \tfrac{1}{N}\sharp\{n; t_n \le b\},$$

further, let be

$$\begin{aligned}
\lambda^{1,+}_{a_0,b_0}(a) &:= 2^{-c_1^{\frac{a_0}{a}}} - 2^{-c_2^{\frac{a_0}{a}}},\\
\lambda^{1,-}_{a_0,b_0}(a) &:= 1 - \lambda^{1,+}_{a_0,b_0}(a),\\
\lambda^{2,+}_{a_0,b_0}(b) &:= \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right),
\end{aligned}$$

and

$$\lambda^{2,-}_{a_0,b_0}(b) := 1 - \lambda^{2,+}_{a_0,b_0}(b).$$

Theorem 3.5 and its proof imply

$$\lambda_N^{2,\pm}(b, T_{*,N}) \underset{N\to\infty}{\longrightarrow} \lambda^{2,\pm}_{a_0,b_0}(b) \text{ almost surely}$$

and Lemma 3.13 and Lemma 3.14 imply

$$\lambda_N^{1,\pm}(a, T_{*,N}) \underset{N\to\infty}{\longrightarrow} \lambda^{1,\pm}_{a_0,b_0}(a) \text{ almost surely.}$$

We showed in Theorem 3.5

$$\lambda^{2,+}_{a_0,b_0}(b) = \frac{1}{2} \Leftrightarrow b = (\ln 2)^{\frac{1}{a_0}} b_0.$$

Hence, $s_2((a_0, b_0)) = (\ln 2)^{\frac{1}{a_0}} b_0$. Moreover, denote the solution of

$$\frac{1}{2} = \lambda^{1,+}_{a_0,b_0}(a) = 2^{-c_1^{\frac{a_0}{a}}} - 2^{-c_2^{\frac{a_0}{a}}}$$

for $\frac{a_0}{a}$ with κ_1. It is $\kappa_1 \approx 0.757$ and $s_1((a_0, b_0)) = \frac{1}{\kappa_1} a_0$.
Using the arguments of the proof of Theorem 2.7 componentwise, we get that

$$(\tilde{a}_N, \tilde{b}_N) \underset{N\to\infty}{\longrightarrow} (s_1(a_0, b_0), s_2(a_0, b_0)) \text{ almost surely.}$$

Set $\Lambda((a_0, b_0), (a, b)) := \left(\lambda^{1,+}_{a_0,b_0}(a) - \frac{1}{2}, \lambda^{2,+}_{a_0,b_0}(b) - \frac{1}{2}\right)$. We have

$$\begin{aligned}
\frac{\partial}{\partial a}\lambda^{1,+}_{a_0,b_0}(a) &= \frac{\partial}{\partial a}\left(\exp\left(-c_1^{\frac{a_0}{a}} \ln 2\right) - \exp\left(-c_2^{\frac{a_0}{a}} \ln 2\right)\right)\\
&= \exp\left(-c_1^{\frac{a_0}{a}} \ln 2\right)\left(-c_1^{\frac{a_0}{a}} \ln 2 \ln c_1\right)\left(-\frac{a_0}{a^2}\right) -\\
&\qquad \exp\left(-c_2^{\frac{a_0}{a}} \ln 2\right)\left(-c_2^{\frac{a_0}{a}} \ln 2 \ln c_2\right)\left(-\frac{a_0}{a^2}\right)\\
&= \exp\left(-c_1^{\frac{a_0}{a}} \ln 2\right) c_1^{\frac{a_0}{a}} \ln 2 \ln c_1 \frac{a_0}{a^2} - \exp\left(-c_2^{\frac{a_0}{a}} \ln 2\right) c_2^{\frac{a_0}{a}} \ln 2 \ln c_2 \frac{a_0}{a^2}\\
&< 0.
\end{aligned}$$

Further, it holds

$$\frac{\partial}{\partial b}\lambda^{1,+}_{a_0,b_0}(a) = 0,\ \frac{\partial}{\partial a}\lambda^{2,+}_{a_0,b_0}(b) = 0$$

and

$$\begin{aligned}\frac{\partial}{\partial b}\lambda^{2,+}_{a_0,b_0}(b) &= \frac{\partial}{\partial b}\exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)\\ &= \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)\left(-a_0\left(\frac{b}{b_0}\right)^{a_0-1}\frac{1}{b_0}\right)\\ &< 0.\end{aligned}$$

Hence, the matrix $\frac{\partial}{\partial(a,b)}\Lambda((a_0,b_0),(a,b))$ is regular. As

$$\begin{aligned}\frac{\partial}{\partial a_0}\lambda^{1,+}_{a_0,b_0}(a) &= \frac{\partial}{\partial a_0}\left(\exp\left(-c_1^{\frac{a_0}{a}}\ln 2\right) - \exp\left(-c_2^{\frac{a_0}{a}}\ln 2\right)\right)\\ &= \exp\left(-c_1^{\frac{a_0}{a}}\ln 2\right)\left(-c_1^{\frac{a_0}{a}}\ln 2\ln c_1\right)\frac{1}{a} - \exp\left(-c_2^{\frac{a_0}{a}}\ln 2\right)\left(-c_2^{\frac{a_0}{a}}\ln 2\ln c_2\right)\frac{1}{a}\\ &= -\exp\left(-c_1^{\frac{a_0}{a}}\ln 2\right)c_1^{\frac{a_0}{a}}\ln 2\ln c_1\frac{1}{a} + \exp\left(-c_2^{\frac{a_0}{a}}\ln 2\right)c_2^{\frac{a_0}{a}}\ln 2\ln c_2\frac{1}{a}\\ &> 0,\end{aligned}$$

besides

$$\frac{\partial}{\partial b_0}\lambda^{1,+}_{a_0,b_0}(a) = 0,\ \frac{\partial}{\partial a_0}\lambda^{2,+}_{a_0,b_0}(b) = -\exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)\left(\frac{b}{b_0}\right)^{a_0}\ln\left(\frac{b}{b_0}\right)$$

and

$$\begin{aligned}\frac{\partial}{\partial b_0}\lambda^{2,+}_{a_0,b_0}(b) &= \frac{\partial}{\partial b_0}\exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)\\ &= \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)\left(-a_0\left(\frac{b}{b_0}\right)^{a_0-1}\right)\left(-\frac{b}{b_0^2}\right)\\ &= \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)a_0\frac{b^{a_0}}{b_0^{a_0+1}}\\ &> 0,\end{aligned}$$

we see that $\frac{\partial}{\partial(a_0,b_0)}\Lambda((a_0,b_0),(a,b))$ is regular, too. Consequently, the proof of Proposition 2.11 on page 16 gives the existence of continuous s^{-1}, such that

$$s^{-1}(\tilde{a}_N(T_{*,N}),\tilde{b}_N(T_{*,N})) \underset{N\to\infty}{\longrightarrow} (a_0,b_0)$$

almost surely. Since

$$s(a_0,b_0) = \left(\frac{1}{\kappa_1}a_0, (\ln(2))^{\frac{1}{a_0}}b_0\right),$$

we have

$$s^{-1}(a,b) = (\kappa_1 a, \frac{b}{(\ln 2)^{\frac{1}{a_0}}}),$$

as

$$s^{-1}(s(a_0,b_0)) = (\kappa_1 \frac{1}{\kappa_1} a_0, \frac{(\ln 2)^{\frac{1}{a_0}}}{(\ln 2)^{\frac{1}{a_0}}} b_0) = (a_0, b_0).$$

Thus, $(\kappa_1 \tilde{a}_N(T_{*,N}), \frac{\tilde{b}_N(T_{*,N})}{(\ln 2)^{\frac{1}{a_0}}})$ is a strongly consistent estimator for (a_0, b_0). □

The power of the LDE was compared to the maximum likelihood estimator (MLE) and the estimator gained by the method of medians (MoM) of He and Fung in a simulation study. The latter is for example proposed in the textbook of Rinne "The Weibull distribution", [Rin 2009], as a robust estimator for the parameters of the Weibull distribution. The maximum likelihood estimator is very sensitive to outliers or contamination. It can be calculated in R, [R 2009], using the method `mle` of the package “stats4”. The MLE can also be obtained by solving

$$\frac{\sum_{i=1}^N t_i^a \ln t_i}{\sum_{i=1}^N t_i^a} - \frac{1}{a} = \frac{1}{N} \sum_{i=1}^N \ln t_i$$

for a and then calculating

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N t_i^{\hat{a}_N},$$

where $\hat{a}_N$ is the solution of the first equation and $\hat{\theta}$ the estimator for $\theta = b^a$. These equations arise directly from the maximum likelihood estimation equations

$$\frac{\partial}{\partial a} \ln(L(t_1, \ldots, t_N; a, \theta)) = 0,$$
$$\frac{\partial}{\partial \theta} \ln(L(t_1, \ldots, t_N; a, \theta)) = 0,$$

see Cohen ([Coh 1965]).

We start the simulation with data without contamination. Table 3.1 shows the mean values of the estimators for Weibull distributed data with varying a and b. We simulate 1000 times 100 data each. Since the likelihood-depth estimator is not unique, there can be more than one parameter with maximum depth, such that we get intervals for a and b, the bounds are indicated with “lb” for lower bound and “ub” for upper bound. These intervals are no confidence intervals. If we want to restrict ourselves to just one point-estimator, we can take the middle of the interval or the upper (lower) bound of the interval, for example. When calculating the mean squared error of the estimators, we use the middle of the interval, when looking at the results we could also have used the lower bound. In Table 3.2 we find the mean squared errors of the estimators for different Weibull distributions, where the mean squared errors are $MSE(\hat{a}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{a}_i - a)^2$ resp. $MSE(\hat{b}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{b}_i - b)^2$, while Table 3.3 gives the mean squared error of both estimators, i.e. $\frac{1}{M^2}\sum_{i=1}^{M}(\hat{a}_i - a)^2 + (\hat{b}_i - b)^2$, here $M = 1000$. The table is visualized in Figure 3.3. Figure 3.4 gives the behavior of the logarithm of the roots of the mean squared errors for growing sample sizes N and data with shape and scale equal to one.

Table 3.1.: MLE, MoM and LDE for different Weibull distributions, number of data $N = 100$ and 1000 repetitions for every parameter.

a	b	$\hat{a}_{MLE}$	$\hat{b}_{MLE}$	$\hat{a}_{MoM}$	$\hat{b}_{MoM}$	$\hat{a}^{lb}_{LDE}$	$\hat{a}^{ub}_{LDE}$	$\hat{b}^{lb}_{LDE}$	$\hat{b}^{ub}_{LDE}$
1	1	1.0153	1.005	1.0255	1.0061	1.0177	1.0434	1.0009	1.0100
1	0.5	1.0158	0.5007	1.0215	0.5009	1.0119	1.0361	0.4988	0.5030
1	10	1.0124	10.0100	1.0199	10.0208	1.0109	1.0347	9.9824	10.0664
0.5	1	0.5061	1.0121	0.5101	1.0102	0.5039	0.5159	1.0051	1.0220
0.5	0.5	0.5062	0.5049	0.5097	0.5085	0.5050	0.5172	0.5047	0.5131
0.5	10	0.5075	10.1104	0.5147	10.2370	0.5078	0.5198	10.1977	10.3663
10	1	10.1604	0.9995	10.2231	0.9997	10.1179	10.3622	0.9993	1.0001
10	10	10.1115	9.9881	10.1757	9.9858	10.0461	10.3018	9.9817	9.9908
10	0.5	10.1673	0.4999	10.278	0.4998	10.1669	10.4197	0.4996	0.5000

Table 3.2.: Mean squared errors of MLE, MoM and LDE for different Weibull samples, number of data $N = 100$ and 1000 repetitions for every parameter.

		Mean squared error					
a	b	$\hat{a}_{MLE}$	$\hat{b}_{MLE}$	$\hat{a}_{MoM}$	$\hat{b}_{MoM}$	$\hat{a}_{LDE}$	$\hat{b}_{LDE}$
1	1	0.0068	0.0113	0.0173	0.0173	0.0182	0.0187
1	0.5	0.0064	0.0028	0.0162	0.0044	0.017	0.0047
1	10	0.007	1.1346	0.0158	1.7983	0.0184	1.9195
0.5	1	0.0017	0.0439	0.0043	0.0675	0.005	0.0697
0.5	0.5	0.0017	0.0115	0.0038	0.0189	0.0043	0.0197
0.5	10	0.0017	4.6620	0.0042	7.3679	0.0044	7.9384
10	1	0.6904	0.0001	1.6574	0.0002	1.8165	0.0002
10	10	0.6944	0.0122	1.6656	0.0183	1.6783	0.0193
10	0.5	0.6724	$< 10^{-4}$	1.7059	$< 10^{-4}$	1.9816	$< 10^{-4}$

Table 3.3.: Mean squared errors of MLE, MoM and LDE for both parameters for different Weibull samples, number of data $N = 100$ and 1000 repetitions for every parameter.

			Mean squared error		
	a	b	MLE	MoM	LDE
a)	1	1	$1.81 \cdot 10^{-5}$	$3.46 \cdot 10^{-5}$	$3.69 \cdot 10^{-5}$
b)	1	0.5	$9.2 \cdot 10^{-6}$	$2.06 \cdot 10-5$	$2.17 \cdot 10^{-5}$
c)	1	10	0.00114	0.00181	0.00194
d)	0.5	1	$4.56 \cdot 10^{-5}$	$7.18 \cdot 10^{-5}$	$7.47 \cdot 10^{-5}$
e)	0.5	0.5	$1.32 \cdot 10^{-5}$	$2.27 \cdot 10^{-5}$	$2.4 \cdot 10^{-5}$
f)	0.5	10	0.00466	0.00737	0.00794
g)	10	1	0.00069	0.00166	0.00182
h)	10	10	0.00707	0.00168	0.00170
i)	10	0.5	0.00067	0.00171	0.00198

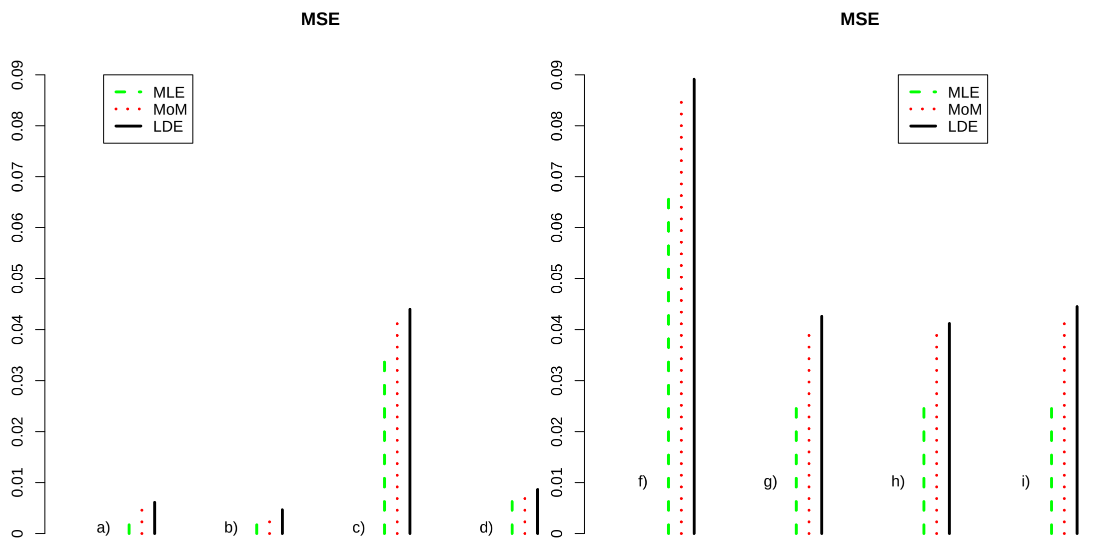

Figure 3.3.: Root of the mean squared errors (MSE) for both parameters, $N = 100$, see Table 3.3.

The tables and the figures show that, as expected, the MLE performs better for uncontaminated data than the LDE and the MoM, the latter ones seem not to differ very much. Especially for smaller sample sizes the mean squared errors of the maximum likelihood estimator are smaller than of the other two methods, see Figure 3.4. Mainly for the estimation of the shape parameter, the differences in the errors of the three estimators, for $N = 10$ and $N = 20$, are quite big. For larger datasets the LDE and the MoM perform

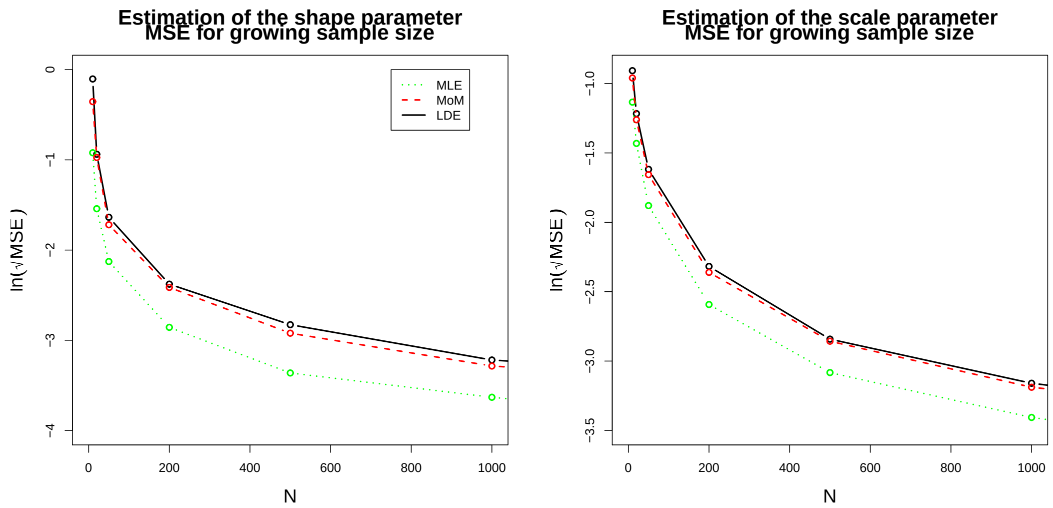

Figure 3.4.: Logarithmic root of the mean squared errors (MSE) for the estimation procedures MLE, MoM and LDE for different sample sizes, $a_0 = b_0 = 1$, 1000 repetitions for every point.

almost as well as the MLE. The asymptotic behavior of the LDE seems to be quite well, because the mean squared error tends to zero, almost as fast as for the MLE.

In the next step we simulate data with contamination. The contamination is given by some data with Weibull distribution with different shape and scale parameters. The ratio of contaminated data is 10% for all examples. We consider the examples from He and Fung, [HeFu 1999], and additionally some more. Each time 1000 times 100 data are simulated. Again, we table the mean squared errors of the estimator for $\theta = (a, b)$, see Table 3.4. The columns a_1 and b_1 give the parameters of the contamination. In order to get an insight into the behavior of the estimators, we depict the roots of the mean squared errors of both estimators in Figure 3.5.

As expected, the MLE performs for some contamination really bad. Especially, when the contaminated data has a small shape parameter, the error of estimation can be quite big. The LDE and the MoM have very similar mean squared errors. Both are more robust against contamination in the shape parameter than in the scale parameter, especially if the contamination parameter is big. But all in all they are not really influenced by the disturbed data.

For uncensored data the new estimator behaves quite similar as the estimator based on the method of medians. For uncontaminated data it is worse than the maximum likelihood estimator, but for contaminated data, where the MLE is very poor, it is robust.

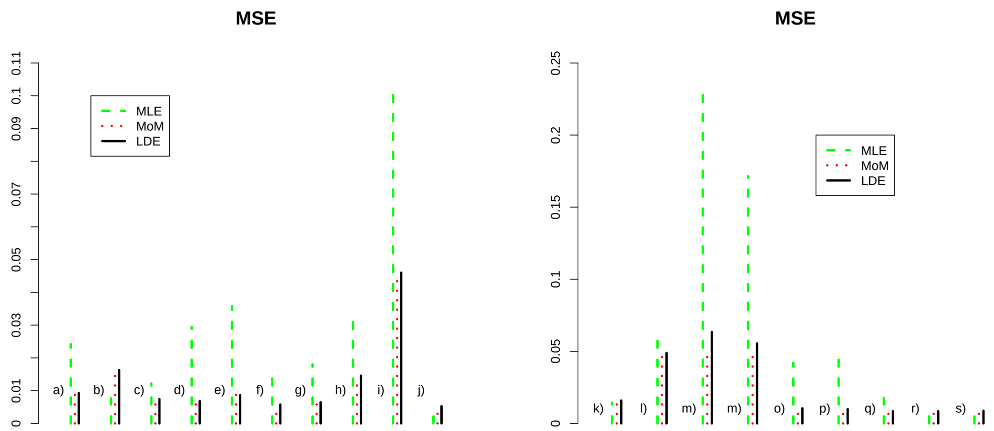

Figure 3.5.: Root of the mean squared errors (MSE) of both parameters for contaminated Weibull data, $N = 100$, values from Table 3.4.

Table 3.4.: MSE of both parameters for contaminated Weibull data, $N = 100$.

					Mean squared error		
	a	b	a_1	b_1	MLE	MoM	LDE
a)	2	1	0.5	1	0.000586	0.00008	0.00009
b)	2	1	10	1	0.00005	0.000228	0.000266
c)	0.7	1	0.2	1	0.000149	0.00005	0.00005
d)	1	1	0.1	0.1	0.000867	0.00004	0.00005
e)	2	0.5	0.5	2	0.001279	0.00008	0.00007
f)	1	0.7	0.2	0.7	0.000273	0.00003	0.00003
g)	1	1	0.2	1	0.000325	0.00004	0.00004
h)	1	3	0.2	3	0.001179	0.000198	0.000212
i)	1	10	0.2	10	0.010352	0.002085	0.002117
j)	1	0.7	1	0.2	0.00001	0.00003	0.00003
k)	1	3	1	0.5	0.000207	0.000247	0.000253
l)	1	10	1	0.5	0.003568	0.002585	0.00239
m)	10	10	1	0.5	0.055147	0.00252	0.004027
n)	10	1	1	0.5	0.029392	0.002177	0.00308
o)	2	1	0.2	0.5	0.001772	0.00009	0.00011
p)	2	1	0.2	2	0.002346	0.00009	0.0001
q)	2	1	1	2	0.000309	0.00007	0.00007
r)	0.5	1	1	1	0.00005	0.00007	0.00007
s)	0.5	1	0.5	0.5	0.00005	0.00007	0.00008

3.2.4. Type-I right-censored data, shape or scale parameter known

In this subsection we will consider type-I right-censored data with fixed censor time c_0. As described in the beginning of this chapter, we examine variables $Y_1, \ldots, Y_N$, $Y_i = \min(T_i, c_0)$, where T_i is the real lifetime and c_0 is the censor time, and the indicator variables Δ_i, $i = 1, \ldots, N$. Assume

$$c_0 > \mathrm{med}(y_1, \ldots, y_N), \tag{3.5}$$

that means less than half the data is censored. This is not a very hard restriction, as estimation for data, where more than half the data is censored, seems very difficult.

The likelihood-function of a data (y_n, δ_n) was described in the beginning of this chapter, namely

$$L(a, b, (y_n, \delta_n)) = f_{a,b}(y_n)^{\delta_n} S_{a,b}(y_n)^{1-\delta_n},\ n = 1, \ldots, N.$$

In the following we consider data $z_n = (y_n, \delta_n)$. The number of uncensored data shall be denoted by k, where $k \leq N$. Because of (3.5) it is $k > \frac{N}{2}$.

We calculate the likelihood-depth for the parameters a and b of the Weibull distribution as in the last section, starting, as before, with the calculation of the likelihood-depth for the scale parameter b. The next theorem shows that, if $b < c_0$, we get the same likelihood-depth for censored data as for uncensored data. If $b \geq c_0$, the likelihood-depth of b in z_* is $\frac{N-k}{N} < \frac{1}{2}$, since $k > \frac{N}{2}$.

Theorem 3.17. *The likelihood-depth of $b > 0$ in $z_* = ((y_1, \delta_1), \ldots, (y_n, \delta_n))$ is*

$$d_T(b, z_*) = \frac{1}{N} \min(\sharp\{n; \delta_n = 1 \text{ and } y_n \geq b\} + (N-k), \sharp\{n; \delta_n = 1 \text{ and } y_n \leq b\}).$$

Especially for $b < c_0$ we have

$$d_T(b, z_*) = \frac{1}{N} \min(\sharp\{n; y_n \geq b\}, \sharp\{n; y_n \leq b\})$$

and for $b \geq c_0$ the likelihood-depth is

$$d_T(b, z_*) = \frac{N-k}{N}.$$

Proof: For $z = (\delta, y)$ we obtain

$$\begin{aligned}
h'_a(b, z) &= \frac{\partial}{\partial b} \ln L(a, b, z) \\
&= \frac{\partial}{\partial b} \left(\delta \ln f_{a,b}(y) + (1-\delta) \ln S_{a,b}(y)\right) \\
&= \frac{\partial}{\partial b} \left(\delta \left(\ln\left(\frac{a}{b}\right) + (a-1)\ln\left(\frac{y}{b}\right) - \left(\frac{y}{b}\right)^a\right) + (1-\delta)\left(-\left(\frac{y}{b}\right)^a\right)\right) \\
&= \delta\left(-\frac{a}{b} - a\left(\frac{y}{b}\right)^{a-1} y \left(-\frac{1}{b^2}\right)\right) + (1-\delta)\left(a y^a \frac{1}{b^{a+1}}\right) \\
&= \delta\left(-\frac{a}{b}\right) + a y^a \frac{1}{b^{a+1}}.
\end{aligned}$$

For uncensored data, i.e. $\delta_n = 1$, we receive $h'_a(b, z_n) \geq 0$ iff $y_n \geq b$. For censored data, i.e. $\delta_n = 0$, we have $h'_a(b, z_n) = a y_n^a \frac{1}{b^{a+1}} > 0$ for all $a > 0$. For instance, the first assertion follows, as $d_T(b, z_*) = \frac{1}{N} \min(\sharp\{n; h'_a(b, z_n) \geq 0\}, \sharp\{n; h'_a(b, z_n) \leq 0\})$.

For $b < c_0$ the likelihood-depth is just

$$d_T(b, z_*) = \frac{1}{N} \min(\sharp\{n; y_n \geq b\}, \sharp\{n; y_n \leq b\}),$$

as for censored data $y_n = c_0 > b$. If $c_0 \leq b$, then $h'_a(b, z_n) \geq 0$ if $\delta_n = 0$ or if $\delta_n = 1$ and $y_n \geq b$. But if $\delta_n = 1$, then $y_n < c_0 \leq b$, so the likelihood-depth is given by the minimum of the number of censored data $N - k$ and the number of uncensored data k. We assumed $N - k < k$, so the minimum is $N - k$ for all $b \geq c_0$. □

Hence, the likelihood-depth of b for censored data is also maximized by the median of the data $\tilde{b}_N$. That means, maximization of the likelihood-depth leads again to a biased estimator. The correction is the same as in the case of uncensored data, as for $b < c_0$ the likelihood-depth is the same for uncensored and censored data and we assumed $c_0 > \tilde{b}_N$. Otherwise the parameter with maximum depth would always be c_0, independent of the underlying distribution, so it would be impossible to find a correction for the estimator.

Corollary 3.18. *Theorem 3.5 on page 39 is also true for type-I right-censored data with fixed censor time, i.e.* $\hat{b}_N = \frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{a_0}}}$ *is a strong consistent estimator for the scale parameter of the Weibull distribution.*

Again the correction depends on the shape parameter. The next theorem gives the likelihood-depth for the shape parameter in dependence of $b = b_0$.

Theorem 3.19. *If* $b = b_0 < c_0$*, the likelihood-depth of* a *in* $z_* = ((y_1, \delta_1), \ldots, (y_N, \delta_N))$ *is*

$$\begin{aligned} d_T^{b_0}(a, z_*) &= \frac{1}{N} \min\Big(\sharp\{n; \delta_n = 1 \text{ and } t_{01}^{a,b_0} \leq y_n \leq t_{02}^{a,b_0}\}, \\ &\qquad \sharp\{n; \delta_n = 1 \text{ and } (y_n \geq t_{02}^{a,b_0} \text{ or } y_n \leq t_{01}^{a,b_0})\} + (N-k)\Big), \end{aligned}$$

else if $b_0 \geq c_0$ *we obtain*

$$d_T^{b_0}(a, z_*) = \frac{1}{N} \min(\sharp\{n; t_{01}^{a,b_0} \leq y_n \leq c_0\}, \sharp\{n; \delta_n = 1 \text{ and } y_n \leq t_{01}^{a,b_0})\}.$$

Thereby t_{01}^{a,b_0} *and* t_{02}^{a,b_0} *are given by Lemma 3.8 on page 41,* $t_{0i}^{a,b} = c_i^{\frac{1}{a}} b$, $i = 1, 2$.

Proof: By definition it is $h'_{b_0}(a, z_n) = \frac{\partial}{\partial a} \ln L(a, b_0, z_n)$, i.e.

$$\begin{aligned} h'_{b_0}(a, z_n) &= \delta_n \left(\frac{1}{a} + \ln\left(\frac{y_n}{b_0}\right) - \left(\frac{y_n}{b_0}\right)^a \ln\left(\frac{y_n}{b_0}\right)\right) + (1 - \delta_n)\left(-\left(\frac{y_n}{b_0}\right)^a \ln\left(\frac{y_n}{b_0}\right)\right) \\ &= \delta_n \left(\frac{1}{a} + \ln\left(\frac{y_n}{b_0}\right)\right) - \left(\frac{y_n}{b_0}\right)^a \ln\left(\frac{y_n}{b_0}\right). \end{aligned}$$

For $\delta_n = 1$, i.e. uncensored data, it is as before, see Lemma 3.8, $h'_{b_0}(a, z_n) \geq 0$ iff $t_{01}^{a,b_0} \leq y_n \leq t_{02}^{a,b_0}$. In the case of $\delta_n = 0$ we have $h'_{b_0}(a, z_n) = -\left(\frac{y_n}{b_0}\right)^a \ln\left(\frac{y_n}{b_0}\right)$. This is negative, iff

$$y_n = c_0 > b_0 (< t_{02}^{a,b_0})$$

and positive or zero, iff $y_n = c_0 \leq b_0$. As

$$d_T^{b_0}(a, z_*) = \tfrac{1}{N} \min(\sharp\{n; h'_{b_0}(a, z_n) \leq 0\}, \sharp\{n; h'_{b_0}(a, z_n) \geq 0\}),$$

the claim is proved. □

We have to consider two cases, as we did in the last section for uncensored data. First we assume $b = b_0$ to be known and in the second step we suppose b to be unknown. Let a be the real parameter and $\tilde{a}_N \in \arg\max_{a'>0} d_T^b(a', z_{*,N})$. The next procedure gives the corrected estimator for a and the derivations of the corrections are given in the subsequent theorem.

Procedure 3.20. *With $\tilde{a}_N$ we denote the parameter with maximum likelihood-depth.*

(1) Let be $b_0 < c_0$.

(a) If $t_{02}^{\tilde{a}_N,b_0} < c_0$, the corrected likelihood-depth estimator for a is (as in the uncensored case) $\hat{a}_N = \kappa \cdot \tilde{a}_N$, with $\kappa \approx 0.691$.

(b) If $t_{02}^{\tilde{a}_N,b_0} \geq c_0$, the correction for the estimator is the solution for a of

$$-\exp\left(-\left(\frac{c_0}{b_0}\right)^a\right) + \exp(-c_1^{\frac{a}{\tilde{a}_N}}) = \frac{1}{2}.$$

(2) Let be $b_0 \geq c_0$. Then the likelihood-depth estimator for a is given by $\hat{a}_N = \kappa_2 \cdot \tilde{a}_N$, where κ_2 is the solution of $\exp(-c_1^{\kappa_2}) = \frac{1}{2}$, $\kappa_2 \approx 0.2715$.

The consistence of this estimator is shown in the next theorem. Recall $t_{01}^{a,b} = c_1^{\frac{1}{a}} b$, $t_{02}^{a,b} = c_2^{\frac{1}{a}} b$, let be $c_0 > b_0$ and define

$$\begin{aligned}
\lambda_N^+(a) &= \frac{1}{N}\sharp\{n; \delta_n = 1, t_{01}^{a,b_0} \leq y_n \leq t_{02}^{a,b_0}\} = \frac{1}{N}\sharp\{n; t_{01}^{a,b_0} \leq t_n \leq \min(c_0, t_{02}^{a,b_0})\}, \\
\lambda_N^-(a) &= \frac{1}{N}\sharp\{n; t_{01}^{a,b_0} \geq t_n \text{ or } t_{02}^{a,b_0} \geq \min(c_0, t_n)\}, \\
\lambda_{a_0,b_0}^+(a) &= -\exp\left(-\left(\frac{\min(c_2^{\frac{1}{a}} b_0, c_0)}{b_0}\right)^{a_0}\right) + \exp\left(-c_1^{\frac{a_0}{a}}\right), \\
\lambda_{a_0,b_0}^-(a) &= 1 - \lambda_{a_0,b_0}^+(a).
\end{aligned}$$

Further recall that with $s(a)$ we denote the solution of $\lambda_{a_0,b_0}^+(s(a)) = \frac{1}{2}$.

Theorem 3.21. *The estimator for the shape parameter of type-I-censored data from the Weibull distribution given in Procedure 3.20 is strongly consistent, if $c_2^{\frac{1}{s(a)}} b_0 \neq c_0$.*

Proof: (1) Consider $c_0 > b_0$. With the strong law of large numbers we have $\lambda_N^{\pm}(a) \to_{N\to\infty} \lambda_{(a_0,b_0)}^{\pm}(a)$ almost surely under $\mathrm{Wei}(a_0, b_0)$. We already proved that λ_N^+ is decreasing and that λ_N^- is increasing for uncensored data. As $c_2^{\frac{1}{a}}$ is decreasing, it is

$$-\exp\left(-\left(\frac{\min(c_2^{\frac{1}{a}}b_0, c_0)}{b_0}\right)^{a_0}\right)$$

decreasing for a and as $\exp(-c_1^{\frac{a_0}{a}})$ is strictly decreasing for a, it is $\lambda_{(a_0,b_0)}^+(\cdot)$ strictly decreasing and $\lambda_{(a_0,b_0)}^-(\cdot)$ strictly increasing.

Therefore, $\tilde{a}_N$ is, with Theorem 2.7 on page 11, a strong consistent estimator for $s(a_0)$, the solution of $\lambda_{(a_0,b_0)}^+(s(a_0)) = \frac{1}{2}$.

Further, it holds

$$\begin{aligned}
\frac{\partial}{\partial a_0}\lambda_{(a_0,b_0)}^+(a) &= -\exp\left(-c_1^{\frac{a_0}{a}}\right)\ln(c_1)c_1^{\frac{a_0}{a}}\frac{1}{a} \\
&\quad + \left\{\begin{array}{ll} \exp\left(-c_2^{\frac{a_0}{a}}\right)\ln(c_2)c_2^{\frac{a_0}{a}}\frac{1}{a}, & c_0 > c_2^{\frac{1}{a}}b_0 \\ \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right)\ln\left(\frac{c_0}{b_0}\right)\left(\frac{c_0}{b_0}\right)^{a_0}, & c_0 < c_2^{\frac{1}{a}}b_0 \end{array}\right\} > 0, \\
\frac{\partial}{\partial a}\lambda_{(a_0,b_0)}^+(a) &= \exp\left(-c_1^{\frac{a_0}{a}}\right)\ln(c_1)c_1^{\frac{a_0}{a}}\frac{a_0}{a^2} \\
&\quad - \left\{\begin{array}{ll} \exp\left(-c_2^{\frac{a_0}{a}}\right)\ln(c_2)c_2^{\frac{a_0}{a}}\frac{a_0}{a^2}, & c_0 > c_2^{\frac{1}{a}}b_0 \\ 0, & c_0 < c_2^{\frac{1}{a}}b_0 \end{array}\right\} < 0.
\end{aligned}$$

Hence, Proposition 2.11 on page 16 yields that s^{-1} exists and that it is continuous for $c_2^{\frac{1}{s(a)}}b_0 \neq c_0$. The solutions for a_0 of $\lambda_{(a_0,b_0)}^+(a) = \frac{1}{2}$ are given in Procedure 3.20.

(2) Now consider $b_0 \geq c_0$. Then $t_{02}^{\tilde{a}_N,b_0} = c_2^{s(a)}b_0 > b_0$ and we already showed in the proof of Theorem 3.19 that $T_{pos}^{\tilde{a}_N,b_0} = \{z = (\delta, y); y \geq t_{01}^{\tilde{a}_N,b_0}\}$. This leads to

$$\begin{aligned}
P_{a,b_0}(T_{pos}^{\tilde{a}_N,b_0}) &= P_{a,b_0}(Y \geq t_{01}^{\tilde{a}_N,b_0}) \\
&= P_{a,b_0}(Y \geq c_1^{\frac{1}{\tilde{a}_N}}b_0) \\
&= \exp\left(-\left(\frac{c_1^{\frac{1}{\tilde{a}_N}}b_0}{b_0}\right)^a\right) \\
&= \exp(-c_1^{\frac{a}{\tilde{a}_N}}).
\end{aligned}$$

Using this to solve $P_{a,b_0}(T_{pos}^{s(a),b_0}) = \frac{1}{2}$, we arrive at

$$\begin{aligned}
&\exp(-c_1^{\frac{a}{\tilde{a}_N}}) = \frac{1}{2} \Leftrightarrow c_1^{\frac{a}{\tilde{a}_N}} = \ln 2 \\
\Leftrightarrow\ &\frac{a}{\tilde{a}_N} = \frac{\ln(\ln 2)}{\ln c_1} \Leftrightarrow a = \underbrace{\frac{\ln(\ln 2)}{\ln c_1}}_{:=\kappa_2}\tilde{a}_N.
\end{aligned}$$

κ_2 is approximately 0.2715. With

$$\lambda_N^+(a) = \frac{1}{N}\sharp\{n; t_{01}^{a,b_0} \le y_n \le c_0\},$$
$$\lambda_N^-(a) = \frac{1}{N}\sharp\{n; \delta_n = 1, y_n \le t_{01}^{a,b_0}\},$$
$$\lambda_{a_0,b_0}^+(a) = P_{a_0,b_0}(Y \ge t_{01}^{a,b_0}) = \exp\left(-c_1^{\frac{a_0}{a}}\right),$$

and

$$\lambda_{a_0,b_0}^-(a) = 1 - \exp\left(-c_1^{\frac{a_0}{a}}\right),$$

it is λ_N^+ decreasing, λ_N^- increasing, $\lambda_{a_0,b_0}^+(\cdot)$ strictly decreasing, and $\lambda_{a_0,b_0}^-(\cdot)$ strictly increasing. Thus, Theorem 2.7 on page 11 provides $\tilde{a}_N \rightarrow_{N\to\infty} s(a_0)$ almost surely and as $s^{-1}(a) = \kappa_2 a$ is continuous, we also have $s^{-1}(\tilde{a}_N) \rightarrow_{N\to\infty} a_0$ almost surely. □

3.2.5. Type-I right-censored data, shape and scale parameter unknown

In the next case, when we assume the scale parameter b to be unknown and estimate it by $\tilde{b}_N$, we know $\tilde{b}_N < c_0$, see (3.5). Therefore, we only have to consider two cases for the correction of the maximum depth estimator for the shape parameter. Before the correction of the maximum depth estimator is determined, we state the following

Lemma 3.22. *Let be $\tilde{b}_N$ the median of $y_{*,N}$, then*

$$\frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N \le t_n \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N)\}$$

converges almost surely to

$$-\exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})}{b_0}\right)^{a_0}\right) + 2^{-c_1^{\frac{a_0}{a}}}$$

under Wei(a_0, b_0).

Proof: The proof works analog to the proof of Lemma 3.13. We will only discuss

$$\frac{1}{N}\sharp\{n; T_n \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N)\} \rightarrow_{N\to\infty} P_{(a_0,b_0)}(T_1 \le \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})).$$

Since $\tilde{b}_N$ converges almost surely under Wei(a_0, b_0) to $b_0(\ln 2)^{\frac{1}{a_0}}$ as $N \to \infty$, and since $\min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})$ is continuous in b_0, also $\min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N)$ converges almost surely under Wei(a_0, b_0) to $\min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})$ as $N \to \infty$.

Let be $\varepsilon > 0$. Then choose δ such that

$$\begin{aligned}&\Bigg|\exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0)}{b_0}\right)^{a_0}\right)\\&\quad-\exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0)(1\pm\delta)}{b_0}\right)^{a_0}\right)\Bigg| < \frac{\varepsilon}{2}.\end{aligned} \tag{3.6}$$

As $\tilde{b}_N \to b_0(\ln 2)^{\frac{1}{a_0}}$ almost surely as $N \to \infty$, it holds for

$A_\delta := \left\{\omega; |\min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N) - \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})| < \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})\delta \text{ for almost all } N\right\}$,

$P(A_\delta) = 1$.

Hence, it holds for $\omega \in A_\delta$ and almost all N

$$\begin{aligned}&\tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\}\\&\le \tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1+\delta)\}\\&= \tfrac{1}{N}\sum_{n=1}^{N} 1_{(-\infty,\min(c_0,c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1+\delta)]}(T_{n_N}(\omega))\end{aligned}$$

and

$$\begin{aligned}&\tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\}\\&\ge \tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1-\delta)\}\\&= \tfrac{1}{N}\sum_{n=1}^{N} 1_{(-\infty,\min(c_0,c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1-\delta)]}(T_{n_N}(\omega)).\end{aligned}$$

Now, the strong law of large numbers provides

$$\begin{aligned}&\Bigg|\tfrac{1}{N}\sum_{n=1}^{N} 1_{(-\infty,\min(c_0,c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1\pm\delta)]}(T_{n_N}(\omega))\\&\quad -P_{a_0,b_0}\left(T_1 \le \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1\pm\delta)\right)\Bigg| < \frac{\varepsilon}{2},\end{aligned}$$

for almost all N. Thus, it holds

$$\begin{aligned}&\tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\}\\&\le P_{(a_0,b_0)}(T_1(\omega) \le \min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1+\delta)) + \frac{\varepsilon}{2}\\&= \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})(1+\delta)}{b_0}\right)^{a_0}\right) + \frac{\varepsilon}{2}\\&= \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}})}{b_0}\right)^{a_0}\right) + \varepsilon,\end{aligned}$$

and

$$
\begin{aligned}
&\tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\}\\
&\ge P_{(a_0,b_0)}(T_1(\omega) \le \min(c_0, c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}})(1-\delta)) - \frac{\varepsilon}{2}\\
&= \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}})(1-\delta)}{b_0}\right)^{a_0}\right) - \frac{\varepsilon}{2}\\
&= \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}})}{b_0}\right)^{a_0}\right) - \varepsilon,
\end{aligned}
$$

where (3.6) is used. This implies for almost all N

$$
|\tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\} - \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}})}{b_0}\right)\right)| < \varepsilon.
$$

Consequently, with

$$
\begin{aligned}
B_\varepsilon := \Bigg\{\omega; \Bigg|&\frac{1}{N}\sharp\{n; c_1^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)) \le T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\}\\
&- \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}})}{b_0}\right)\right)\Bigg| < \varepsilon \text{ for } N \text{ large enough}\Bigg\},
\end{aligned}
$$

it holds $1 = P(A_\delta) \le P(B_\varepsilon)$ and therefore

$$
\begin{aligned}
&1 = \lim_{k\to\infty} P_{a_0,b_0}(B_{\frac{1}{k}}) = P_{a_0,b_0}\left(\bigcap_{k=1}^{\infty} B_{\frac{1}{k}}\right) =\\
&P_{a_0,b_0}\left(\{\omega; \tfrac{1}{N}\sharp\{n; T_{n_N}(\omega) \le \min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N(T_{*,N}(\omega)))\} = \exp(-(\tfrac{\min(c_0, c_2^{\frac{1}{a}} b_0(\ln 2)^{\frac{1}{a_0}})}{b_0}))\}\right).
\end{aligned}
$$

□

We are able to give an estimation procedure for the type-I right-censored data with Weibull distribution for both scale and shape parameter.

Procedure 3.23. *Let be $Z_1, \dots, Z_N$ i.i.d., $Z_i = (Y_i, \Delta_i)$, and $Y_i = \min(T_i, c_0)$, $i = 1, \dots, N$. Suppose $T_i \sim \mathrm{Wei}(a, b)$, $i = 1, \dots, N$, and let be $z_* = ((y_1, \delta_1), \dots, (y_N, \delta_N))$ with $c_0 > \mathrm{med}(y_1, \dots, y_N)$. Then the two parameters a and b of the Weibull distribution can be estimated with the help of the likelihood-depth as follows:*

1. *Calculate $\tilde{b}_N = \mathrm{med}(y_1, \dots, y_N)$.*
2. *Determine $\tilde{a}_N \in \arg\max d_T^{\tilde{b}_N}(a, z_*)$.*

3.A *If $c_2^{\frac{1}{\tilde{a}_N}}\tilde{b}_N < c_0$, then we correct the estimator for a_0 as in the uncensored case, i.e.*

$$\hat{a}_N = \kappa_1 \cdot \tilde{a}_N$$

and

$$\hat{b}_N = \frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{\hat{a}_N}}}.$$

3.B If $c_2^{\frac{1}{\hat{a}_N}} \tilde{b}_N \geq c_0$, then $\hat{a}_N$ is the solution of

$$-2^{-\left(\frac{c_0}{\tilde{b}_N}\right)^a} + 2^{-c_1^{\frac{a}{\hat{a}_N}}} = \frac{1}{2}$$

for a and

$$\hat{b}_N = \frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{\hat{a}_N}}}.$$

We define

$$\begin{aligned}
\lambda_N^{1,+}(a, z_{*,N}) &:= \tfrac{1}{N}\sharp\left\{n; \delta_n = 1 \text{ and } c_1^{\frac{1}{a}}\tilde{b}_N \leq y_n \leq c_2^{\frac{1}{a}}\tilde{b}_N\right\},\\
\lambda_N^{2,+}(b, z_{*,N}) &:= \tfrac{1}{N}\sharp\{n; y_n \geq b\},\\
\lambda_N^{1,-}(a, z_{*,N}) &:= \tfrac{1}{N}\sharp\left\{n; y_n \leq c_1^{\frac{1}{a}}\tilde{b}_N \vee y_n \geq \min(c_2^{\frac{1}{a}}\tilde{b}_N, c_0)\right\},\\
\lambda_N^{2,-}(b, z_{*,N}) &:= \tfrac{1}{N}\sharp\{n; y_n \leq b\},
\end{aligned}$$

and further

$$\begin{aligned}
\lambda_{a_0,b_0}^{1,+}(a) &:= \exp\left(-c_1^{\frac{a_0}{a}} \ln 2\right) - \exp\left(-\left(\frac{\min(c_2^{\frac{1}{a}} b_0 (\ln 2)^{\frac{1}{a_0}}, c_0)}{b_0}\right)^{a_0}\right),\\
\lambda_{a_0,b_0}^{1,-}(a) &:= 1 - \lambda_{a_0,b_0}^{1,+}(a),\\
\lambda_{a_0,b_0}^{2,+}(b) &:= \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right),\\
\lambda_{a_0,b_0}^{2,-}(b) &:= 1 - \lambda_{a_0,b_0}^{2,+}(b).
\end{aligned}$$

Let be $s(a,b) = (s_1(a_0,b_0), s_2(a_0,b_0))$ such that

$$\begin{cases} \lambda_{(a_0,b_0)}^{1,+}(s_1(a_0,b_0)) = \frac{1}{2} \\ \lambda_{(a_0,b_0)}^{2,+}(s_2(a_0,b_0)) = \frac{1}{2} \end{cases},$$

we already discussed $s_2(a_0,b_0) = (\ln 2)^{\frac{1}{a_0}} b_0 =: \tilde{b}$.

The consistence of the estimators in the censored case, shows

Theorem 3.24. *If $c_0 > b_0(\ln 2)^{\frac{1}{a_0}}$ and $c_2^{\frac{1}{s_1(a_0,b_0)}}(\ln 2)^{\frac{1}{a_0}} b_0 \neq c_0$, then the estimator $(\hat{a}_N, \hat{b}_N)$ given by Procedure 3.23 is a strongly consistent estimator for (a_0, b_0).*

Proof: Using the results from before, we have $\lambda_N^{1,+}(\cdot)$, $\lambda_N^{2,+}(\cdot)$ being decreasing, $\lambda_N^{1,-}(\cdot)$, $\lambda_N^{2,-}(\cdot)$ increasing, $\lambda_{(a_0,b_0)}^{1,+}(\cdot), \lambda_{(a_0,b_0)}^{2,+}(\cdot)$ strictly decreasing and so $\lambda_{(a_0,b_0)}^{1,-}(\cdot), \lambda_{(a_0,b_0)}^{2,-}(\cdot)$ strictly increasing. Also we already discussed for the different cases that

$$\lim_{N\to\infty} \lambda_N^{1,\pm}(a) = \lambda_{(a_0,b_0)}^{1,\pm}(a),$$

and

$$\lim_{N\to\infty} \lambda_N^{2,\pm}(b) = \lambda_{(a_0,b_0)}^{2,\pm}(b)$$

almost surely. Hence, as it is $s(a,b) = (s_1(a_0,b_0), s_2(a_0,b_0))$ such that

$$\begin{cases} \lambda_{(a_0,b_0)}^{1,+}(s_1(a_0,b_0)) = \frac{1}{2} \\ \lambda_{(a_0,b_0)}^{2,+}(s_2(a_0,b_0)) = \frac{1}{2} \end{cases},$$

Using the arguments of the proof of Theorem 2.7 on page 11 componentwise gives $(\tilde{a}_N(Z_{*,N}), \tilde{b}_N(Z_{*,N})) \to_{N\to\infty} s(a_0,b_0)$ almost surely.

Now we prove that the inverse s^{-1} exists and that it is continuous. Therefore, we use arguments from the proof of Proposition 2.11 on page 16 and define

$$\Lambda((a_0,b_0),(a,b)) = \left(\lambda_{(a_0,b_0)}^{1,+}(a) - \frac{1}{2}, \lambda_{(a_0,b_0)}^{2,+}(b) - \frac{1}{2}\right).$$

It holds

$$\begin{aligned}
\frac{\partial}{\partial a_0}\lambda_{(a_0,b_0)}^{1,+}(a) &= -2^{-c_1^{\frac{a_0}{a}}}\ln(2)\ln(c_1)c_1^{\frac{a_0}{a}}\frac{1}{a} \\
&\quad + \begin{Bmatrix} 2^{-c_2^{\frac{a_0}{a}}}\ln(2)\ln(c_2)c_2^{\frac{a_0}{a}}\frac{1}{a}, & c_0 > c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0 \\ \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right)\ln\left(\frac{c_0}{b_0}\right)\left(\frac{c_0}{b_0}\right)^{a_0}, & c_0 < c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0 \end{Bmatrix} > 0, \\
\frac{\partial}{\partial a}\lambda_{(a_0,b_0)}^{1,+}(a) &= 2^{-c_1^{\frac{a_0}{a}}}\ln(2)\ln(c_1)c_1^{\frac{a_0}{a}}\frac{a_0}{a^2} \\
&\quad - \begin{Bmatrix} 2^{-c_2^{\frac{a_0}{a}}}\ln(2)\ln(c_2)c_2^{\frac{a_0}{a}}\frac{a_0}{a^2}, & c_0 > c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0 \\ 0, & c_0 < c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0 \end{Bmatrix} < 0, \\
\frac{\partial}{\partial b_0}\lambda_{(a_0,b_0)}^{1,+}(a) &= \begin{cases} 0, & c_0 > c_2^{\frac{1}{a}}(\ln 2)^{\frac{1}{a_0}}b_0 \\ \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right)\left(-a_0\left(\frac{c_0}{b_0}\right)^{a_0-1}\frac{1}{b_0}\right) < 0, & c_0 < c_2^{\frac{1}{a_0}}(\ln 2)^{\frac{1}{a}}b_0 \end{cases}, \\
\frac{\partial}{\partial b}\lambda_{(a_0,b_0)}^{1,+}(a) &= 0, \\
\frac{\partial}{\partial a_0}\lambda_{(a_0,b_0)}^{2,+}(b) &= -\exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)\ln\left(\frac{b}{b_0}\right)\left(\frac{b}{b_0}\right)^{a_0}, \\
\frac{\partial}{\partial a}\lambda_{(a_0,b_0)}^{2,+}(b) &= 0, \\
\frac{\partial}{\partial b_0}\lambda_{(a_0,b_0)}^{2,+}(b) &= \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)a_0\frac{b^{a_0}}{b_0^{a_0+1}} > 0, \\
\frac{\partial}{\partial b}\lambda_{(a_0,b_0)}^{2,+}(b) &= -\exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)a_0\frac{b^{a_0-1}}{b_0^{a_0}} < 0.
\end{aligned}$$

Thus, $\frac{\partial}{\partial(a,b)}\Lambda((a,b),s(a,b))|_{(a,b)=(a_0,b_0)}$ and $\frac{\partial}{\partial(a,b)}\Lambda((a_0,b_0),(a,b))|_{(a,b)=s(a_0,b_0)}$ are regular matrices and with the same arguments as in the proof of Proposition 2.11 there exists continuous s^{-1}, such that

$$s^{-1}((\tilde{a}_N(Z_{*,N}),\tilde{b}_N(Z_{*,N}))) \rightarrow_{N\rightarrow\infty} (a_0,b_0)$$

almost surely. For $c_2^{\frac{1}{a}}b_0(\ln 2)^{\frac{1}{a_0}} > c_0$, $s^{-1}(a,b)$ is the solution for (a_0,b_0) of

$$\begin{cases} 2^{-c_1^{\frac{a_0}{a}}} - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) = \frac{1}{2} \\ \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right) = \frac{1}{2} \end{cases}. \tag{3.7}$$

Since the second equation is only true, iff $b_0 = \frac{b}{(\ln 2)^{\frac{1}{a_0}}}$, (3.7) is equivalent to

$$\begin{cases} 2^{-c_1^{\frac{a_0}{a}}} - \exp\left(-\left(\frac{c_0}{b}\right)^{a_0}\ln 2\right) = \frac{1}{2} \\ b_0 = \frac{b}{(\ln 2)^{\frac{1}{a_0}}} \end{cases},$$

what completes the proof. □

Before we examine the behavior of this new estimator in simulations studies, the shift-function is studied a little more precisely. In the uncensored case we showed that the expected parameter with maximum depth for the shape parameter was always greater than the real parameter, i.e. $s(a) > a$. But in the censored case this is not true for all a.

Lemma 3.25. *If b_0 is known and $c_0 > b_0$, $t_{02}^{s(a),b_0} = c_2^{\frac{1}{s(a)}}b_0 \geq c_0$ and*

$$a \leq \frac{\ln\left(-\ln\left(e^{-c_1}-\frac{1}{2}\right)\right)}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)},$$

then $s(a) \leq a$ (where $s(a) = a$, if $a = \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$), else $s(a) > a$ holds, where $s(a)$ such that

$$\lambda^{1,+}_{(a,b_0)}(s(a)) = -\exp\left(-\left(\frac{\min(c_2^{\frac{1}{s(a)}}b_0, c_0)}{b_0}\right)^a\right) + \exp\left(-c_1^{\frac{a}{s(a)}}\right) = \frac{1}{2}.$$

If b_0 is unknown, then let be $\tilde{b} = (\ln 2)^{\frac{1}{a}}b_0$, if $\tilde{b} < c_0$. If $t_{02}^{s(a),\tilde{b}} = c_2^{\frac{1}{s(a)}}\tilde{b} \geq c_0$ and

$$a \leq \frac{\ln\left(-\frac{\ln\left(2^{-c_1}-\frac{1}{2}\right)}{\ln(2)}\right)}{\ln\left(\frac{c_0}{\tilde{b}}\right)} \approx \frac{0.455}{\ln\left(\frac{c_0}{\tilde{b}}\right)},$$

then $s(a) \leq a$ (where $s(a) = a$ if $a = \frac{0.455}{\ln\left(\frac{c_0}{\tilde{b}}\right)}$), else $s(a) > a$ holds, where $s(a)$ such that

$$\lambda^{1,+}_{(a,b_0)}(s(a)) = \exp\left(-c_1^{\frac{a}{s(a)}}\ln 2\right) - \exp\left(-\left(\frac{\min(c_2^{\frac{1}{s(a)}}b_0(\ln 2)^{\frac{1}{a}}, c_0)}{b_0}\right)^a\right) = \frac{1}{2}.$$

Proof: Consider b_0 to be known.

- For $b_0 \geq c_0$ we showed in the proof of Theorem 3.21 that $s(a) = \frac{1}{\kappa_2} a$, with $\frac{1}{\kappa_2} \approx 3.683$, i.e. $s(a) > a$.
- For $b_0 < c_0$ and $t_{02}^{s(a),b_0} < c_0$ we can use the results from the uncensored case, there we proved $s(a) > a$.
- Now let be $b_0 < c_0$ and $t_{02}^{s(a),b_0} \geq c_0$ then $s(a)$ is the solution of

$$- \exp\left(-\left(\frac{c_0}{b_0}\right)^a\right) + \exp\left(-c_1^{\frac{a}{s(a)}}\right) = \frac{1}{2}. \tag{3.8}$$

If b_0 is unknown we consider $\tilde{b} = b_0(\ln 2)^{\frac{1}{a_0}}$. We assume $\tilde{b} < c_0$, as $\tilde{b}$ is the limit of the median and this is supposed to be always smaller than the censoring time, see (3.5) on page 54.

- For $t_{02}^{a,\tilde{b}} < c_0$ we can use the results from the uncensored case where we proved that $s(a) > a$.
- For $t_{02}^{a,\tilde{b}} \geq c_0$ ($\tilde{b} < c_0$) $s(a)$ is the solution of

$$\begin{aligned}
& \exp\left(-c_1^{\frac{a}{s(a)}} \ln 2\right) - \exp\left(-\left(\frac{c_0}{b_0}\right)^a\right) = \frac{1}{2} \\
\Leftrightarrow \quad & \exp\left(-c_1^{\frac{a}{s(a)}} \ln 2\right) - \exp\left(-\left(\frac{c_0}{\frac{\tilde{b}}{(\ln 2)^{\frac{1}{a}}}}\right)^a\right) = \frac{1}{2} \\
\Leftrightarrow \quad & 2^{-c_1^{\frac{a}{s(a)}}} - 2^{\left(\frac{c_0}{\tilde{b}}\right)^a} = \frac{1}{2}.
\end{aligned}$$

Thus, in both situations, b_0 known and unknown, the only case where $s(a)$ can be smaller than a is

$$t_{02}^{s(a),b} \geq c_0, \text{ while } b < c_0.$$

Here it is $b = b_0$ if b_0 is known and $b = \tilde{b} = b_0(\ln 2)^{\frac{1}{a}}$, else. In order to consider both cases b_0 known and b_0 unknown at once, we write x instead of $\exp(1)$ in the first case and instead of 2 in the second case. Then $s(a)$ is in both situations given by solving

$$-x^{-\left(\frac{c_0}{b}\right)^a} + x^{-c_1^{\frac{a}{s(a)}}} = \frac{1}{2}. \tag{3.9}$$

This yields

$$s(a) = \frac{\ln(c_1)}{\ln\left(-\frac{\ln\left(\frac{1}{2}+x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}\right)} \cdot a, \tag{3.10}$$

as the next lines show:

$$\begin{aligned}
& -x^{-\left(\frac{c_0}{b}\right)^a} + x^{-c_1^{\frac{a}{s(a)}}} = \frac{1}{2} \\
\Leftrightarrow\ & \ln(x)(-c_1^{\frac{a}{s(a)}}) = \ln\left(\frac{1}{2} + x^{-\left(\frac{c_0}{b}\right)^a}\right) < 0, \text{ as } x^{-\left(\frac{c_0}{b}\right)^a} < x^{-1} \leq \frac{1}{2} \\
\Leftrightarrow\ & \frac{a}{s(a)} = \ln\left(-\ln\left(\frac{1}{2} + x^{-\left(\frac{c_0}{b}\right)^a}\right)\right)\frac{1}{\ln(c_1)} \\
\Leftrightarrow\ & s(a) = \frac{\ln(c_1)}{\ln\left(-\frac{\ln\left(\frac{1}{2}+x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}\right)} \cdot a.
\end{aligned}$$

It is $\ln(c_1) < 0$ and $\ln\left(-\frac{\ln\left(\frac{1}{2}+x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}\right) < 0$, so

$$\ln(c_1)\left(\ln\left(-\frac{\ln\left(\frac{1}{2} + x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}\right)\right)^{-1} > 0.$$

The second claim can be seen as follows:

$$\ln\left(\frac{1}{2}\right) < \ln\left(\frac{1}{2} + x^{\overbrace{-\left(\frac{c_0}{b}\right)^a}^{<-1}}\right) < \ln(1) = 0, \tag{3.11}$$

as $x \geq 2$ it is $\ln\left(\frac{1}{2}\right) \geq -\ln(x)$ and thus (3.11) is equivalent to

$$\begin{aligned}
& -1 \leq \frac{\ln\left(\frac{1}{2}\right)}{\ln(x)} < \frac{\ln\left(\frac{1}{2} + x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)} < 0, \text{ as } \ln(x) > 0 \\
\Leftrightarrow\ & 0 < -\frac{\ln\left(\frac{1}{2} + x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)} < 1 \\
\Leftrightarrow\ & \ln\left(-\frac{\ln\left(\frac{1}{2} + x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}\right) < 0.
\end{aligned}$$

Now we determine a, such that $\ln(c_1) > \ln\left(-\frac{\ln\left(\frac{1}{2}+x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}\right)$, i.e. $c_1 > -\frac{\ln\left(\frac{1}{2}+x^{-\left(\frac{c_0}{b}\right)^a}\right)}{\ln(x)}$.

This leads to

$$
\begin{aligned}
& -c_1\ln(x) < \ln\left(\frac{1}{2}+x^{-\left(\frac{c_0}{b}\right)^a}\right) \\
\Leftrightarrow\quad & \exp\left(-c_1\ln(x)\right)-\frac{1}{2} < x^{-\left(\frac{c_0}{b}\right)^a} \\
\Leftrightarrow\quad & \ln\left(x^{-c_1}-\frac{1}{2}\right) < \ln(x)\left(-\left(\frac{c_0}{b}\right)^a\right) \\
\Leftrightarrow\quad & \left(\frac{c_0}{b}\right)^a < -\frac{\ln\left(x^{-c_1}-\frac{1}{2}\right)}{\ln(x)} \\
\Leftrightarrow\quad & a < \ln\left(-\frac{\ln\left(x^{-c_1}-\frac{1}{2}\right)}{\ln(x)}\right)\frac{1}{\ln\left(\frac{c_0}{b}\right)}.
\end{aligned}
$$

For $b = b_0$, i.e. $x = \exp(1)$, it is $\ln\left(-\frac{\ln\left(x^{-c_1}-\frac{1}{2}\right)}{\ln(x)}\right) \approx 0.265$ and for $b = \tilde{b}$, i.e. $x = 2$, it is $\ln\left(-\frac{\ln\left(x^{-c_1}-\frac{1}{2}\right)}{\ln(x)}\right) \approx 0.455$. □

When we construct tests, the conclusion of this lemma will be important.

We compare the LDE for censored data to the maximum likelihood estimator (MLE) for censored data. The latter can be calculated by first solving

$$\frac{\sum_{i=1}^N y_i^a \ln y_i}{\sum_{i=1}^N y_i^a} - \frac{1}{a} = \frac{1}{k}\sum_{i=1}^k \ln y_i,$$

and then determining $\theta = b^a$ as

$$\hat{\theta} = \frac{1}{k}\sum_{i=1}^N y_i^{\hat{a}},$$

where $\hat{a}$ is the solution of the first equation, i.e. the MLE for the shape parameter, and k the number of uncensored data. These equations can be found in Cohen [Coh 1965]. Also, we compare our new estimator (LDE) again to the estimator based on the method of median (MoM) for censored data as a robust estimator. In their paper [HeFu 1999] He and Fung wrote, that the estimator is not affected by censoring, as far as less than 16 percent of the largest observations are right censored. The source code of the estimators can be found in the Appendix B.

In the first study, every time 16% of the biggest data are censored. The sample size is fixed at $N = 100$ and we repeat every simulation 1000 times. The mean squared errors of the estimators can be found in Tables 3.5 and 3.6. We display Table 3.6 in Figure 3.6 on the left.

Table 3.5.: Mean squared errors for censored Weibull data, with 16% right-censored data, $N = 100$, 1000 repetitions each.

	a	b	Mean squared errors $\hat{a}_{MLE}$	$\hat{b}_{MLE}$	$\hat{a}_{MoM}$	$\hat{b}_{MoM}$	$\hat{a}_{LDE}$	$\hat{b}_{LDE}$
a)	1	1	0.0082	0.0125	0.0257	0.0148	0.017	0.0182
b)	2	1	0.0363	0.0030	0.1089	0.0038	0.0685	0.0045
c)	1	2	0.0091	0.0546	0.0284	0.0637	0.0187	0.0775
d)	2	2	0.0372	0.0127	0.1163	0.0159	0.0644	0.0192
e)	0.5	1	0.0023	0.0500	0.0073	0.0579	0.8440	0.0745
f)	1	0.5	0.0090	0.0029	0.0273	0.0037	0.0178	0.0043
g)	0.5	0.5	0.0021	0.0142	0.0069	0.0155	0.5811	0.0212
h)	0.5	10	0.0023	4.8857	0.0067	6.0464	1.2506	7.9931
i)	0.5	100	0.0022	521.2956	0.0067	625.6615	1.1530	839.1851

Table 3.6.: Mean squared errors for both parameters for censored Weibull data, with 16% right-censored data, $N = 100$, 1000 repetitions each.

	a	b	MSE for both parameter MLE	MoM	LDE
a)	1	1	2e-05	4e-05	4e-05
b)	2	1	4e-05	0.00011	7e-05
c)	1	2	6e-05	9e-05	1e-04
d)	2	2	5e-05	0.00013	8e-05
e)	0.5	1	5e-05	7e-05	9e-05
f)	1	0.5	1e-05	3e-05	2e-05
g)	0.5	0.5	2e-05	2e-05	2e-05
h)	0.5	10	0.0049	0.0060	0.0092
i)	0.5	100	0.5213	0.62567	0.68474

In order to consider heavier censoring we simulate, once more, various Weibull samples and introduce a censoring, such that 40% of the data is censored. Table 3.7 gives the mean errors for the comparison of all three estimators. Again, the sample size is fixed at $N = 100$ and we simulated 1000 repetitions each. For a clearer view, we display the values of Table 3.7 in Figure 3.6 on the right. In Figure 3.7, the evolution of the simulated mean squared error for growing sample sizes is depicted. Here we simulated data with $a = b = 1$ and censored 20% of the data.

Table 3.7.: Mean squared errors for censored Weibull data, 40% of the data censored, $N = 100$, 1000 repetitions each.

			MSE for both parameter		
	a	b	MLE	MoM	LDE
a)	1	1	0.00003	0.01236	0.00116
b)	2	1	0.00006	0.04689	0.00414
c)	1	2	0.00009	0.01223	0.00126
d)	2	2	0.00008	0.04663	0.00407
e)	0.5	1	0.00008	0.00301	0.00038
f)	1	0.5	0.00002	0.01147	0.00109
g)	0.5	0.5	0.00003	0.00290	0.00028
h)	0.5	10	0.00841	0.02087	0.01111
i)	0.5	100	0.96366	1.78749	1.10283

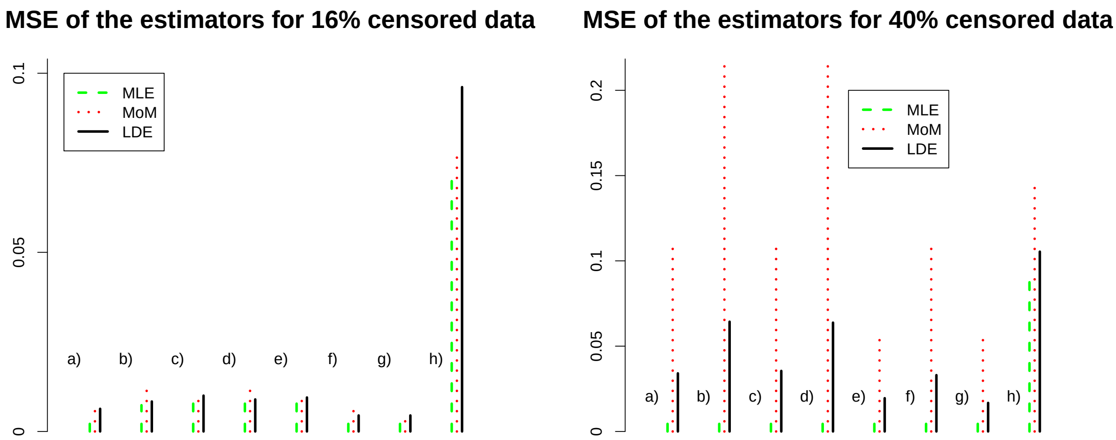

Figure 3.6.: Root of MSE of both estimators for the different procedures, see Table 3.6 and Table 3.7.

We see that for uncontaminated censored data in the first case, where 16% of the data is censored, the MLE performs best, concerning the MSEs. In most cases the MoM and the LDE are quite close. If we consider data where 40% is censored, the MoM behaves worst, while MLE again has the smallest mean squared errors. For censored data the new estimator seems to be better than the one based on the method of medians but worse than the MLE. For small sample sizes the MLE has smaller mean squared errors than the LDE but with N growing, the MSEs of the LDE are also shrinking. The mean squared errors of the estimator based on the method of medians are for large sample sizes still the biggest ones.

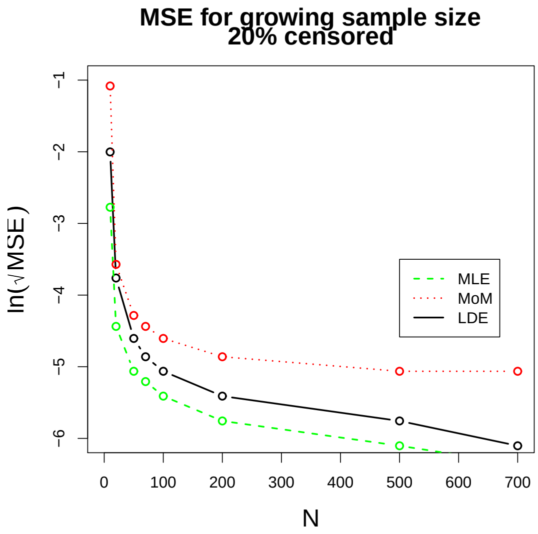

Figure 3.7.: Development of the logarithmic root of MSE for censored data, $a_0 = b_0 = 1$, 20% censored, 1000 repetitions each.

We also simulate different ε-contaminated and censored data. Here a contamination of 10% with data from $\text{Wei}(a_1, b_1)$ and 20% right-censoring is considered. The mean squared errors of both estimators for the three different procedures are given in Table 3.8. We only considered contamination distribution such that the probability that the contaminated data is censored is quite small, i.e. the contamination distribution has smaller scale and/ or shape parameter.

Table 3.8.: Mean squared errors for contaminated and censored Weibull data, 10% contaminated and 20% of the data censored, $N = 100$, 1000 repetitions each.

					MSE for both parameter		
	a	b	a_1	b_1	MLE	MoM	LDE
a)	10	10	0.1	0.1	0.08918	0.00471	0.00508
b)	2	2	0.5	1	0.00028	0.00022	0.00001
c)	2	100	0.5	10	0.06862	0.10058	0.06714
d)	2	100	1	10	0.08228	0.11017	0.06773
e)	5	1	1	0.1	0.00843	0.00102	0.00079
f)	5	1	2	1	5e-04	0.00139	0.00044

For contaminated data the MLE behaves worst, as expected. In most cases the likelihood-depth estimator is quite robust and has the smallest mean squared errors. See also Figure 3.8.

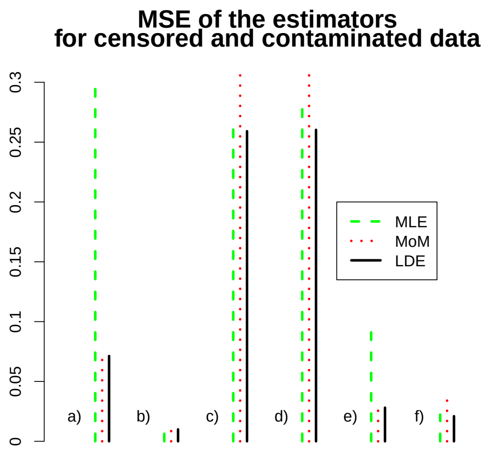

Figure 3.8.: Root of the mean squared errors of the estimators in contaminated (10%) and censored data (20%).

As a last consideration we examine a real data example with very little data and some modifications to analyze the robustness of the new estimator.

Example 3.26. In order to make predictions about the lifetime of some sort of steel, specimens were taken from the material and loaded. The lifetime is given in load cycles until the specimen broke. Let be y_* the dataset of lifetimes of steel specimens,

$$y_* = (4030, 4680, 4860, 5750, 7170, 34100, 51000).$$

We consider various artificial censoring, addition of some new fake specimen and replacement of data by extreme values, to examine the robustness of the estimators. The first column of Table 3.9 gives the data, the next columns show the values of the different estimators. For the LDE we took the mean of the estimators, if there were more than one.

For a better view we display the estimated values in Figures 3.9 and 3.10, where the first value represents always the MLE, the second one the MoM, and the third one the LDE. We see that the new estimator seems to be quite robust. It takes about the same values as the estimator based on the method of medians. If the two or three largest values are censored, the LDE and MoM do not really change in contrast to the MLE. Even if we only censor two observations, the maximum likelihood estimator for b falls from about 16000 to 8000. Also in the estimation of a the MLE differs much more than the LDE. Especially it changes from $a < 1$ to $a > 1$, if a censoring is included. The MoM also changes more than the LDE. We observe, that the estimator of the shape parameter is smaller and sometimes greater than one. Hence it changes from a decreasing to an increasing hazard function. For the LDE we have almost every time an $a > 1$. The interpretation of the

Table 3.9.: Estimated a and b for real data and artificially modified data of Example 3.26.

	data	$\hat{a}_{MLE}$	$\hat{b}_{MLE}$	$\hat{a}_{MoM}$	$\hat{b}_{MoM}$	$\hat{a}_{LDE}$	$\hat{b}_{LDE}$
a)	y_*	0.98	15795.48	2.47	6670.44	3.86	6377.83
censoring							
b)	$c_0 = 7000$	3.48	7088.82	4.96	6191.23	3.05	6485.44
c)	$c_0 = 10000$	2.28	8275.86	2.47	6670.45	3.86	6377.83
new fake specimen							
d)	$(500, 500, y_*)$	0.74	10212.88	2.51	5624.85	2.04	8180.28
e)	$(500, 500, 500, y_*)$	0.69	8547.79	0.57	9074.21	0.98	8400.45
f)	$(50, 50, y_*)$	0.58	8430.9	2.51	5624.88	1.97	10517.12
g)	$(50, 50, 50, y_*)$	0.51	6386.8	0.57	9076.72	0.9	10796.48
h)	$(5, 5, y_*)$	0.47	7153.81	2.51	5624.91	1.97	10517.12
i)	$(5, 5, 5, y_*)$	0.4	4921.67	0.57	9075.51	0.9	10796.48
j)	$(y_*, 10^5, 10^5)$	0.82	30970.88	0.88	10863.38	1.39	13310.11
k)	$(y_*, 10^5, 10^5, 10^5)$	0.84	37629.94	0.58	38745.07	0.68	35336.45
l)	$(y_*, 10^6, 10^6)$	0.44	82672.41	0.88	10863.38	1.39	13310.11
m)	$(y_*, 10^6, 10^6, 10^6)$	0.45	125807.33	0.35	59342.81	0.68	35336.71
replacement							
n)	$y_1 = 5$	0.59	11312.8	2.47	6670.45	2.65	11450.76
o)	$y_1 = y_2 = 5$	0.4	6998.44	0.55	11229.15	1.52	13927.62
p)	$y_1 = 500$	0.82	13775.27	2.47	6670.45	2.69	9998.11
q)	$y_1 = y_2 = 500$	0.69	11560.87	0.55	11226.96	1.59	10185.17
r)	$y_7 = 10^5$	0.79	19334.97	2.47	6670.44	3.86	6377.83
s)	$y_6 = y_7 = 10^5$	0.71	24948.66	2.47	6670.44	3.86	6377.83
t)	$y_7 = 10^6$	0.44	41140.82	2.47	6670.44	3.86	6377.83
u)	$y_6 = y_7 = 10^6$	0.4	84350.04	2.47	6670.44	3.86	6377.83

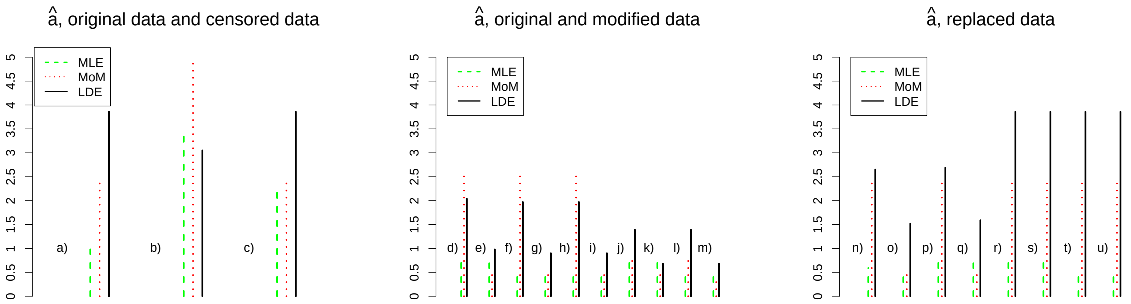

Figure 3.9.: Estimation of the shape parameter for the data of Example 3.26, see also Table 3.9.

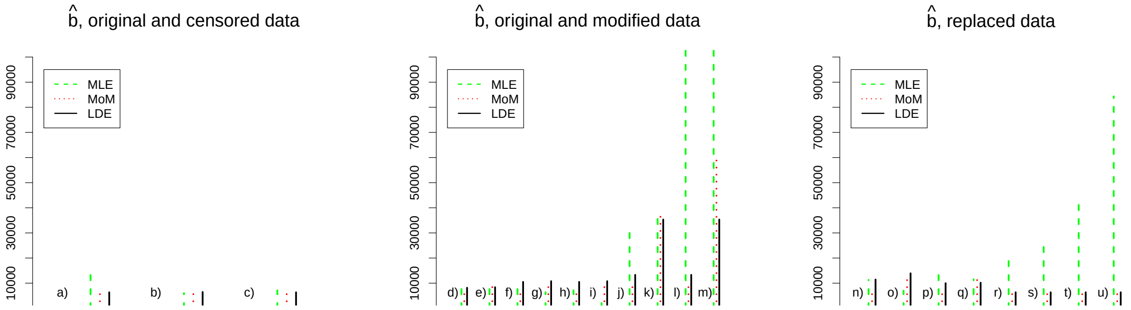

Figure 3.10.: Estimation of the scale parameter for the data of Example 3.26, see also Table 3.9.

estimators would therefore be very different. If one ore two observations are adulterated to a high value, the estimation of b by the MLE also seems to increase without limit.

Thus, the LDE is a robust estimator that can also be used in censored data.

3.3. Tests and confidence intervals for the shape parameter

In this section the theory of Section 2.2 is used to give tests for the shape parameter a of the Weibull distribution. We start with uncensored data and consider both cases, $b = b_0$ to be known and b unknown and estimated with the help of the median.

3.3.1. Uncensored data with known scale parameter

We stick to Lemma 2.14, see page 17, and define the test statistic to test hypotheses about the shape parameter for the case of uncensored data t_* with known scale parameter.

Definition 3.27. *Let the test statistic be defined as*

$$T(a, t_*) := \sqrt{N} \frac{d_S^{b_0}(a, t_*) - 2p_{shape}(1 - p_{shape})}{2\sqrt{p_{shape}(1 - p_{shape})(1 - 2p_{shape})^2}},$$

where $p_{shape} := \exp(-c_1) - \exp(-c_2) \approx 0.665$ *and* $d_S^{b_0}(a, t_*)$ *denotes the simplicial likelihood-depth of* a *in* t_* *depending on* b_0*. Further,* c_1, c_2 *with* $c_1 < 1 < c_2$*, shall denote the solutions of* $\ln c = \frac{1}{c-1}$.

We start with testing the null hypothesis $H_0 : a \leq a_0$. Therefore, the theory of Chapter 2.2 to find tests with good asymptotic power functions will be adapted.

Theorem 3.28. *The test*

$$\varphi_{a_0}^{\leq,0}(t_*) := 1_{\{\sup_{a \leq a_0} T(a,t_*) < \Phi^{-1}(\alpha)\}}(t_*)$$

is asymptotically an α*-level test for* $H_0 : a \leq a_0$ *against* $H_1 : a > a_0$*.*

Proof: Using Corollary 2.17 on page 19, the only thing to prove is that $p_{shape} = P_{a,b_0}(T_{pos}^{a,b_0})$. We already showed in the last section in Lemma 3.8 on page 41 and Corollary 3.9 on page 41 that $T_{pos}^{a,b_0} = [c_1^{\frac{1}{a}} b_0, c_2^{\frac{1}{a}} b_0]$ and in the proof of Lemma 3.10 on page 42 we showed $P_{a,b_0}(T_{pos}^{a,b_0}) = \exp(-c_1) - \exp(-c_2)$. □

We compare this new test based on likelihood-depth to a test based on the maximum likelihood estimator (MLE). It can be found in the textbook of Rinne, [Rin 2009]. There it is shown that the maximum likelihood estimator for a is asymptotically normal distributed with mean a and variance $\frac{0.6079a^2}{N}$. So $H_0 : a \leq a_0$ is rejected, if $\hat{a}_{MLE} > a_0(1 + \Phi^{-1}(1 - \alpha)\sqrt{\frac{0.6079}{N}})$, where $\hat{a}_{MLE}$ is the maximum likelihood estimator for the shape parameter of the Weibull distribution, see [Coh 1965] and the section about estimators for the Weibull distribution, Section 3.2.

The graphics in Figure 3.11 show the estimated power functions for various a_0. We simulate data with Weibull distribution Wei(a, b_0) and count how often $H_0 : a \leq a_0$ is rejected. Some of the results for $N = 100$ data and 1000 repetitions for every a are displayed. We show the simulated power for $a_0 = 0.5, 1, 1.5$ and 2, where $b_0 = 1$. In the last row the simulated power, if b_0 is not 1 but 0.5 and 2 respectively, for $H_0 : a \leq 1$ is depicted. The source code for the test can be found in the Appendix B.1.

We see in Figure 3.11 that the power does not change, if the scale is varying. Also we note that for $N = 100$ data the level is not kept by both tests. The test based on the MLE seems to be a little bit more powerful for these uncontaminated data.

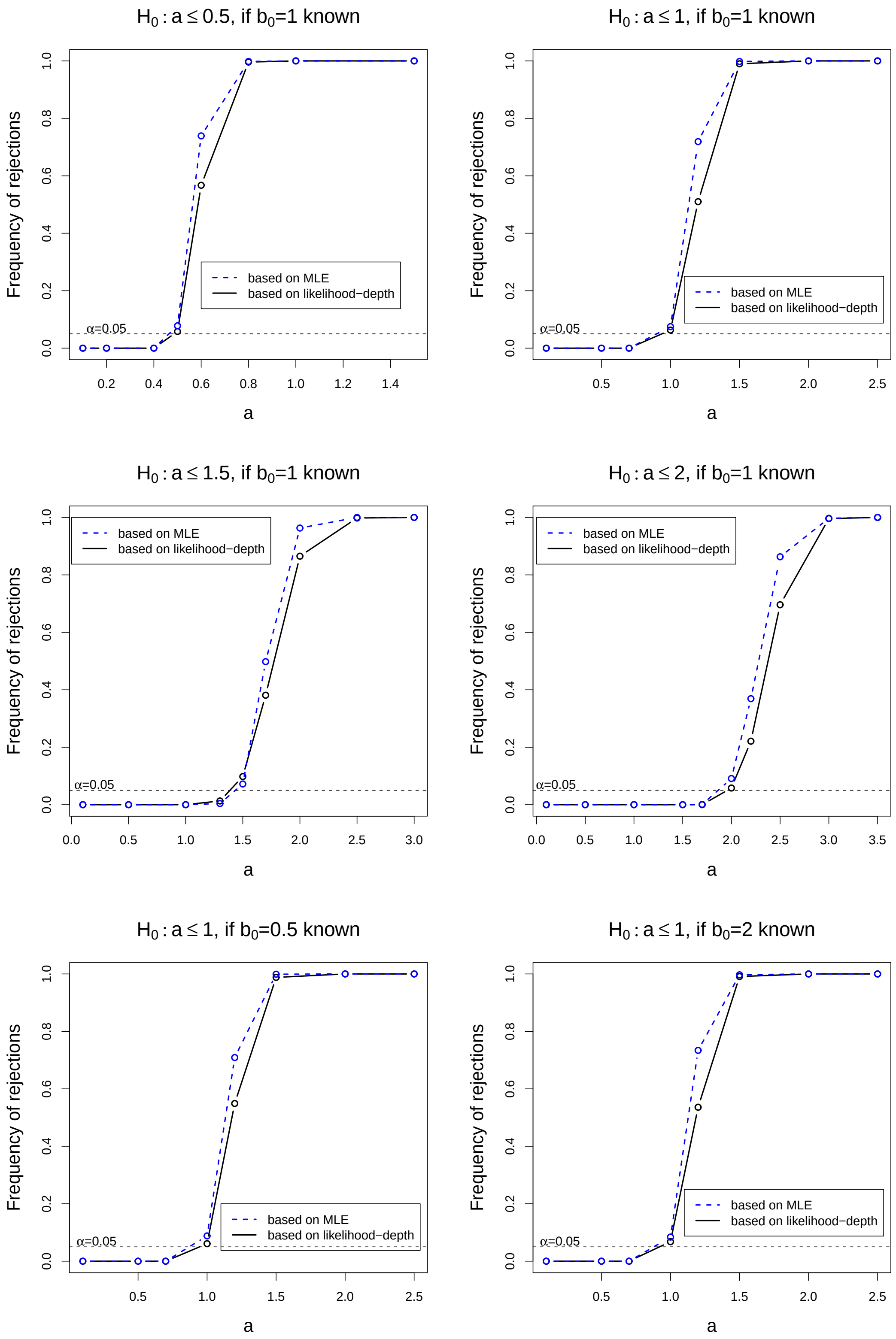

Figure 3.11.: Simulated power of the tests for $H_0 : a \leq a_0$, b_0 the scale parameter known.

We also consider contaminated data, as the new test is supposed to be robust against contamination. In a next study we examine data where some part is given by another distribution, here $\text{Wei}(a_1, b_0)$. We simulate data with Weibull distribution $\text{Wei}(a, b_0)$ for different a and mix with data coming from the contamination distribution $\text{Wei}(a_1, b_0)$, then count how often $H_0 : a \leq a_0$ is rejected. In Figure 3.12 the simulated power-functions for $H_0 : a \leq 1$, where the contaminated data has a shape parameter $a_1 = 0.5$ ($a_1 = 10$), are pictured. The ratio of the contaminated data is 10%.

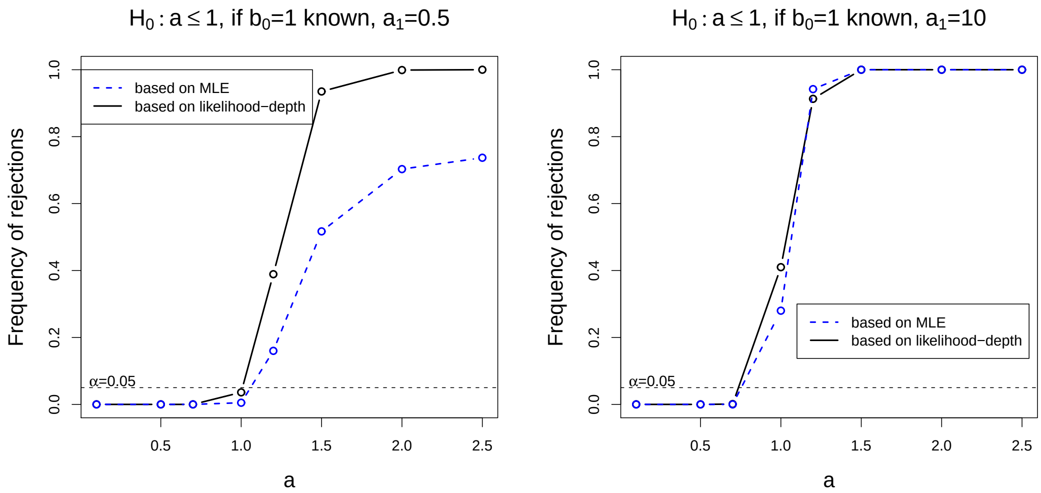

Figure 3.12.: Simulated power of the tests for $H_0 : a \leq a_0$, b_0 known, for ε-contaminated data with $\text{Wei}(a_1, b_0)$, $\varepsilon = 0.1$.

For ε-contamination with a small shape parameter, the test based on the maximum likelihood estimator has a very bad power in contrast to the test based on likelihood-depth, that is not very much affected by the contamination with $a_1 = 0.5$. Considering contamination with a higher shape parameter, both tests do not keep the level.

To define the test for hypotheses of type $H_0 : a \geq a_0$ against $H_1 : a < a_0$ and to prove consistency of both tests, we have to work little more. Let the test statistic be defined as in Definition 3.27. Since $p_{shape} > 0.5$ and the parameter with maximum likelihood-depth ($s(a) \approx 1.32 \cdot a$) is expected to be greater than the real parameter, we have to determine

$$c^1_\alpha(a_0) := \max\{a; \lim_{N \to \infty} P_{a_0,b_0}(T(a, T_*) < \Phi^{-1}(\alpha)) \leq \alpha\},$$

see Definition 2.18 on page 19 to correct the test, see Definition 2.19 on page 20.

Before we do this, we state two lemmas, that will be used in the proof of the subsequent theorems.

Lemma 3.29. *The probability that for $T \sim \text{Wei}(a_0, b_0)$ it holds $T \in T^{a,b_0}_{pos}$, is determined by*

$$p_{a_0,a} \;:=\; P_{a_0,b_0}(T^{a,b_0}_{pos}) = \exp(-c_1^{\frac{a_0}{a}}) - \exp\left(-c_2^{\frac{a_0}{a}}\right).$$

Proof: It holds

$$\begin{aligned}
p_{a_0,a} &= P_{a_0,b_0}(T^{a,b_0}_{pos}) = P_{a_0,b_0}(c_1^{\frac{1}{a}} b \le T \le c_2^{\frac{1}{a}} b) \\
&= \exp\left(-\left(\frac{c_1^{\frac{1}{a}} b}{b_0}\right)^{a_0}\right) - \exp\left(-\left(\frac{c_2^{\frac{1}{a}} b}{b_0}\right)^{a_0}\right) \\
&= \exp(-c_1^{\frac{a_0}{a}}) - \exp\left(-c_2^{\frac{a_0}{a}}\right).
\end{aligned}$$

□

Lemma 3.30. *It is $p_{a_0,(\cdot)}$ strictly decreasing and $p_{(\cdot),a_0}$ is strictly increasing.*

Proof: We showed in the last lemma $p_{a_0,a} = \exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-c_2^{\frac{a_0}{a}}\right)$. Let be $a_1 < a_2$, then $\frac{a_0}{a_1} > \frac{a_0}{a_2}$ and therefore, since $c_1 < 1$, $c_1^{\frac{a_0}{a_1}} < c_1^{\frac{a_0}{a_2}}$ so $-c_1^{\frac{a_0}{a_1}} > -c_1^{\frac{a_0}{a_2}}$, i.e. $e^{-c_1^{\frac{a_0}{a_1}}} > e^{-c_1^{\frac{a_0}{a_2}}}$. With the same arguments it is, since $c_2 > 1$, $c_2^{\frac{a_0}{a_1}} > c_2^{\frac{a_0}{a_2}}$ so $-c_2^{\frac{a_0}{a_1}} < -c_2^{\frac{a_0}{a_2}}$, i.e. $-e^{-c_2^{\frac{a_0}{a_1}}} > -e^{-c_2^{\frac{a_0}{a_2}}}$. All in all we get

$$\begin{aligned}
\exp\left(-c_1^{\frac{a_0}{a_1}}\right) - \exp\left(-c_2^{\frac{a_0}{a_1}}\right) &> \exp\left(-c_1^{\frac{a_0}{a_2}}\right) - \exp\left(-c_2^{\frac{a_0}{a_1}}\right) \\
&> \exp\left(-c_1^{\frac{a_0}{a_2}}\right) - \exp\left(-c_2^{\frac{a_0}{a_2}}\right).
\end{aligned}$$

With analog arguments we can show that $p_{(\cdot),a_0}$ is strictly increasing. □

Now we use Lemma 2.25 on page 23 to determine $c^1_\alpha(a_0)$.

Lemma 3.31. *Let be $\alpha < 0.5$. It holds*

$$c^1_\alpha(a_0) = k_0 \cdot a_0,$$

with $k_0 \approx 2.275$.

Especially $c^1_\alpha(a_0)$ exists for all $a_0 > 0$, it is $c^1_\alpha(a_0) > a_0$ for all $a_0 > 0$ and $c^1_\alpha(\cdot)$ strictly increasing.

Proof: In Lemma 3.29 we showed $p_{a_0,a} = \exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-c_2^{\frac{a_0}{a}}\right)$. It is $0.5 < p_{shape} \approx 0.665 < \frac{1}{2} + \frac{1}{\sqrt{8}} \approx 0.85$ and Lemma 3.30 gives that $p_{a_0,(\cdot)}$ is strictly decreasing and $p_{(\cdot),a_0}$ is strictly increasing. The assumptions of Lemma 2.25 are fulfilled and we obtain $c^1_\alpha(a_0)$ as that value a, which fulfills $1 - p_{a,a} = p_{a_0,a}$. As $p_{a,a} = p_a = p_{shape}$, this means we are looking for $a > a_0$ such that $1 - p_{shape} = \exp(-c_1^{\frac{a_0}{a}}) - \exp(-c_2^{\frac{a_0}{a}})$, i.e. $0.439 = \frac{a_0}{a}$, so

$$a = 2.275 a_0.$$

This proves the claim. □

Using the last lemma and the theory of Section 2.2, we prove that $\varphi^{0,\le}$ is a consistent test.

Theorem 3.32. *Let be $\alpha < 0.5$. The test $\varphi_{a_0}^{0,\leq}$ is a consistent test with asymptotic level α for $H_0 : a \leq a_0$.*

Proof: With Lemma 3.29, 3.30 and 3.31, the assumptions of Theorem 2.33 on page 28 are fulfilled, since also $p_{shape} = \exp(-c_1) - \exp(-c_2)$ is constant and thereby continuous.

□

Analogously we deduce from Theorem 2.34 on page 29 the following

Theorem 3.33. *Let be $\alpha < 0.5$. We use the test statistic $T(a, t_*)$ given by Definition 3.27 and get a consistent test $\varphi_{a_0}^{\geq}$ for $H_0 : a \geq a_0$ with asymptotic level α, by rejecting H_0, if $\sup\limits_{a \geq c_\alpha^1(a_0)} T(a, t_*) < \Phi^{-1}(\alpha)$, where c_α^1 is given by Lemma 3.31.*

Proof: Use the same arguments as in the last proof in order to conclude that the assumptions of Theorem 2.34 are fulfilled. □

To estimate the power of the new test for finite sample size, we simulate datasets with various shape and scale parameters a and b_0 and count for different a_0 how often the hypothesis $H_0 : a \geq a_0$ is rejected. Again, this test is compared to the test based on the maximum likelihood estimator (MLE). Here $H_0 : a \geq a_0$ is rejected, if $\hat{a}_{MLE} < a_0 \left(1 + \Phi^{-1}(\alpha)\sqrt{\frac{0.6079}{N}}\right)$, see [Rin 2009].

The results can be found in Figure 3.13. For each point in each graphic we simulate 1000 times 100 data.

We see that compared to the case of testing $H_0 : a \leq a_0$, the level is better kept for $N = 100$ data. The power of the test based on the maximum likelihood estimator is better than the power of the new test for uncontaminated data. Consider now data that are contaminated with some data coming from another distribution with different shape parameter a_1. The simulated power for $H_0 : a \geq 1$, with $a_1 = 0.5$ and $a_1 = 10$, is displayed in Figure 3.14. The ratio of contamination is 10%.

Figure 3.14 indicates that both tests are infected by the contamination with a small shape parameter, as both do not keep the level anymore. The test based on the MLE behaves worse than the test based on likelihood-depth. The contamination with shape parameter $a_1 = 10$ has only a little influence on the power of both tests.

Finally, we use Theorem 2.32 on page 27 and the results from above to construct a consistent test for $H_0 : a = a_0$. Further, we give confidence intervals for the shape parameter of the Weibull distribution.

Theorem 3.34. *Let be $\alpha < 0.5$, $T_1, \ldots, T_N$ i.i.d., $T_i \sim \text{Wei}(a, b_0)$, $i = 1, \ldots, N$. With $t_* = (t_1, \ldots, t_N)$ we denote the realizations of $T_1, \ldots, T_N$. A consistent test with asymptotic level α for*

$$H_0 : a = a_0$$

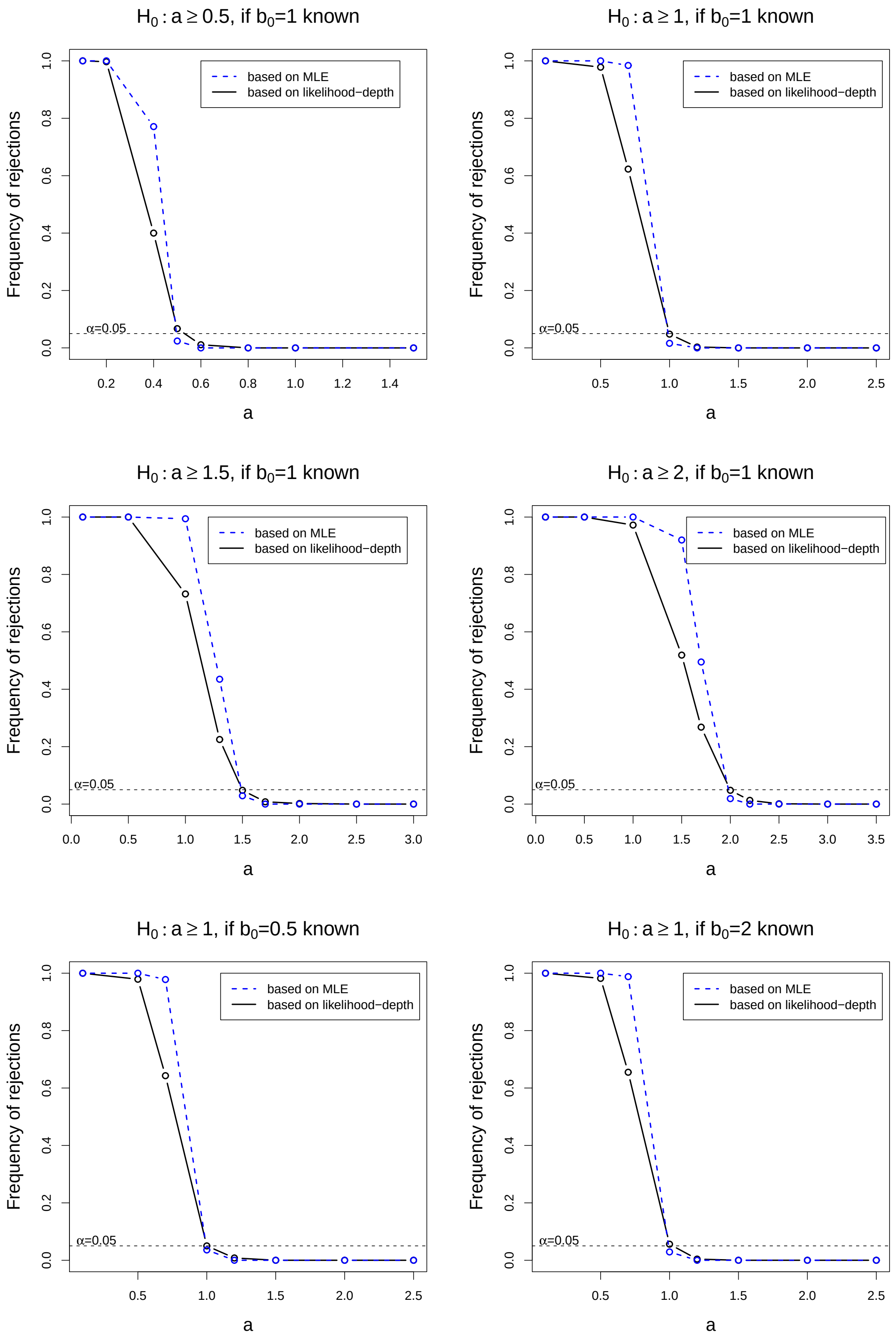

Figure 3.13.: Simulated power of the tests for $H_0 : a \geq a_0$, b_0 known.

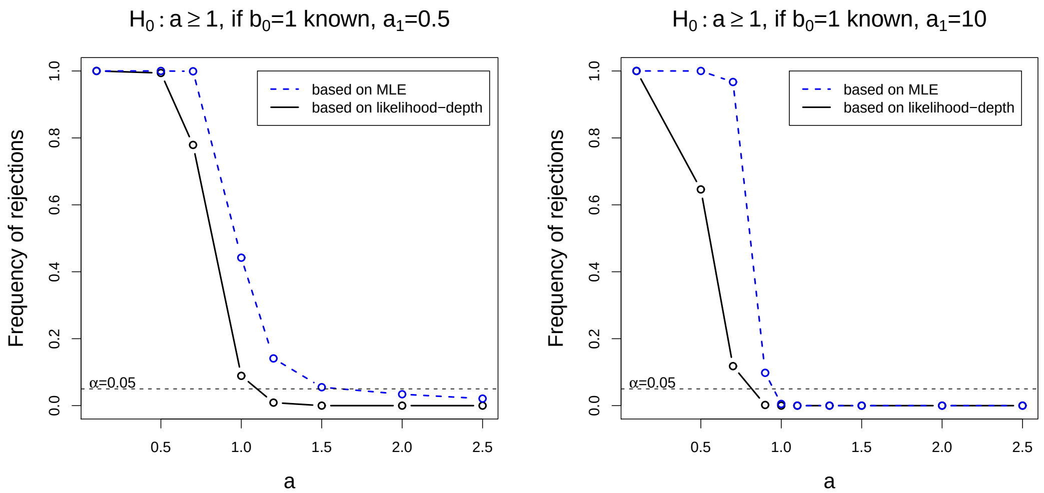

Figure 3.14.: Simulated power of $H_0 : a \geq 1$, b_0 known, with ε-contaminated data from $\text{Wei}(a_1, b_0)$, $\varepsilon = 0.1$.

against $H_1 : a \neq a_0$ *is given by*

$$\varphi^{=}_{a_0}(t_*) = \max\left(1_{\{T(a_0,t_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(t_*), 1_{\{T(c^1_{\frac{\alpha}{2}}(a_0),t_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(t_*)\right).$$

A confidence interval with level $\gamma = 1 - \alpha$ *for the shape parameter* a *is given by*

$$\{a_0 > 0; \varphi^{=}_{a_0}(t_*) = 0\}.$$

Proof: As the assumptions of Theorem 2.32 on page 27 are fulfilled, the claim is true. □

The estimated power-function for finite samples of the test for $H_0 : a = a_0$ based on likelihood-depth is compared to the power of the test based on the MLE and the test based on the method of medians (MoM) in a simulation study. There we use the results from He and Fung in [HeFu 1999]. They give a confidence interval as

$$(\hat{a}_{MoM} \exp\left(\frac{1.2 \cdot \Phi^{-1}(\frac{\alpha}{2})}{\sqrt{N}}\right), \hat{a}_{MoM} \exp\left(-\frac{1.2 \cdot \Phi^{-1}(\frac{\alpha}{2})}{\sqrt{N}}\right)),$$

where $\hat{a}_{MoM}$ is the estimator based on the method of medians, see also the section about estimation. In the mentioned article lower and upper bound are given the other way around, but as $\Phi^{-1}(\alpha)$ is negative for $\alpha < 0.5$, we think it is a typing error. Based on this confidence interval we get a test for $H_0 : a = a_0$, by rejecting H_0, if a_0 is not lying in the confidence interval with level $\gamma = 1 - \alpha$.

As before we consider various a_0 and b_0. The results are displayed in Figure 3.15. We see that the new test, and sometimes also the other two tests, does not keep the level.

Here too, we also consider ε-contaminated data, the results are displayed in Figure 3.16 and show the robustness of the new test and the test based on the median. It is $\varepsilon =$

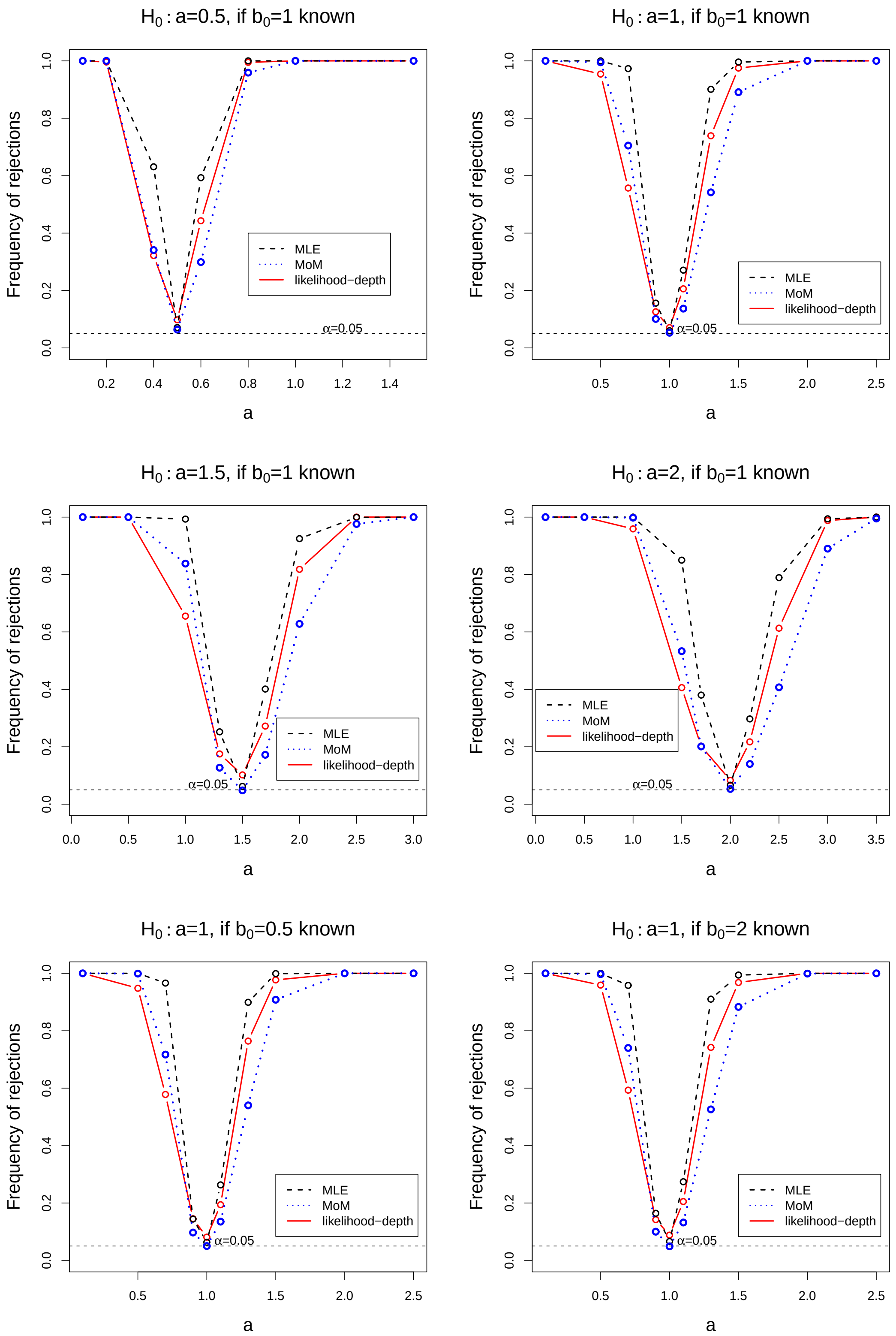

Figure 3.15.: Simulation of the power-function of the tests for $H_0 : a = a_0$, known scale.

0.1, $N = 100$ and $M = 100$. In all cases, uncontaminated and contaminated data, the power of the test based on likelihood-depth is better than the power of the test based on the MoM for $a > a_0$. For $a < a_0$ it is the other way around. Both tests are more robust against ε-contamination than the test based on the MLE.

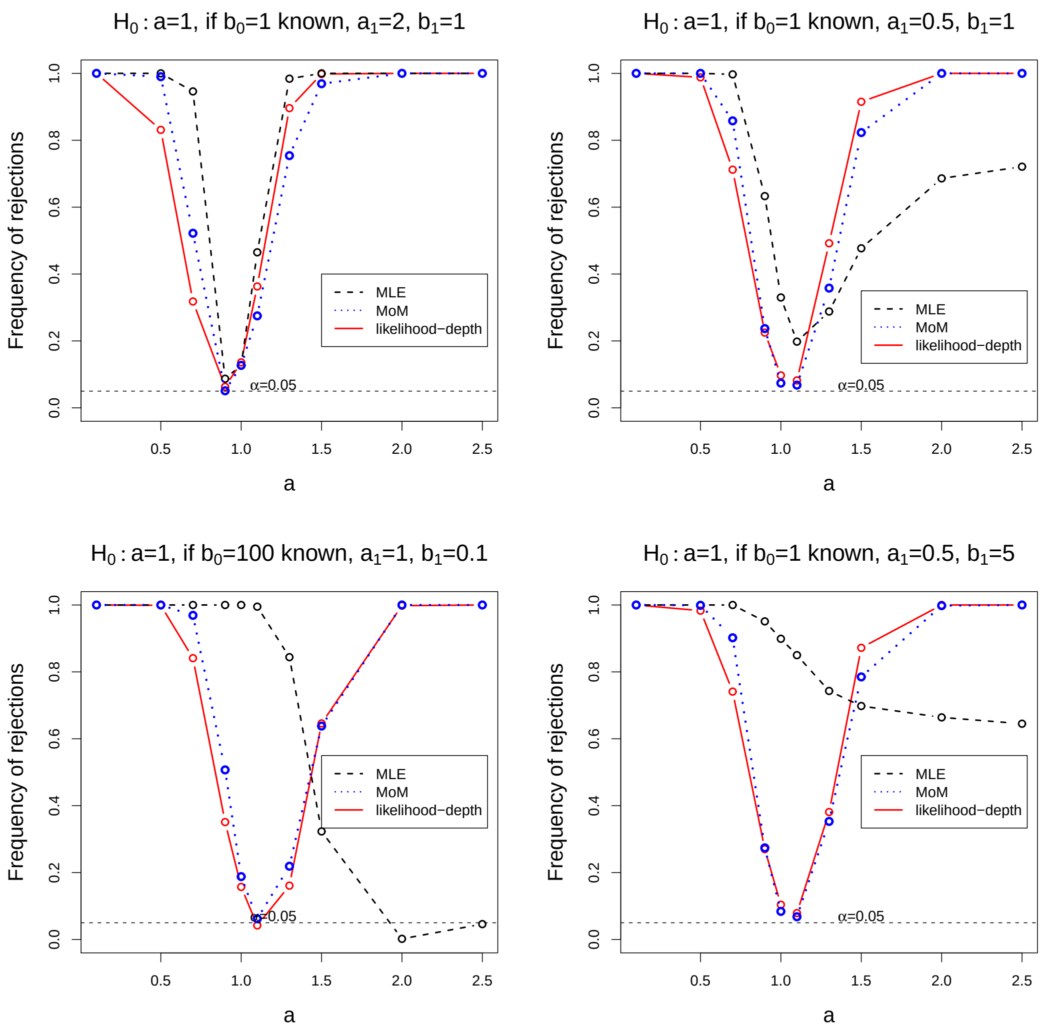

Figure 3.16.: Simulated power-function of the tests for $H_0 : a = a_0$, known scale, for ε-contaminated data from Wei(a_1, b_1).

There are some more methods to determine confidence intervals for the shape parameter of the Weibull distribution. For example in Lawless' textbook “Statistical Models and Methods for Lifetime Data”, see [Law 2003] on page 211-229, the Wald-type and likelihood-ratio procedures are described in Chapter 5 “Inference Procedures for Log-Location-Scale Procedures”. As the accuracy of the likelihood-ratio procedure is superior to that of Wald-type procedures, according to Lawless, we will consider them here. Let be $y_n = \ln t_n$, $n = 1, \ldots, N$. Then the density function of $Y = \ln T$ becomes

$$f_{u,v}(y) = \frac{1}{v} e^{\frac{y-u}{v}} \exp\left(-e^{\frac{y-u}{v}}\right),$$

where $u = \ln b$ and $v = a^{-1}$. The log-likelihood function of the sample $y_* = (y_1, \ldots, y_N)$ has the form

$$l(u, v) = -N \ln b + \sum_{i=1}^{N} (z_i - e^{z_i}),$$

where $z_i = \frac{y_i - u}{v}$. The likelihood-ratio statistic is given by

$$\Lambda(v_0) = 2l(\hat{u}, \hat{v}) - 2l(\hat{u}(v_0), v_0),$$

where $\hat{u}, \hat{v}$ are the MLEs for u, v and $\hat{u}(v_0)$ is the parameter that maximizes $l(u, v_0)$. The latter is determined as $\hat{u}(v_0) = v_0 \ln \left(\frac{1}{N} \sum_{i=1}^{N} e^{\frac{y_i}{v_0}} \right)$. Under the hypothesis $v = v_0$, the test statistic is asymptotically χ^2-distributed with one degree of freedom. Let $\chi^2_{1,q}$ denote the q-quantile of the χ^2_1-distribution. Then a confidence interval with level q for v_0 is given by

$$\{v > 0; \Lambda(v) \leq \chi^2_{1,q}\},$$

which can be transformed to receive a confidence interval for $a = v^{-1}$.

We compare the confidence intervals for the shape parameter based on likelihood-depth (lik-depth) to three other methods. The first is, as before, the method based on the MLE, the second the above mentioned method based on likelihood-ratio statistics (LRS), and as a robust method we consider confidence intervals based on the method of medians (MoM), see [HeFu 1999].

Table 3.10 shows the mean length and the coverage rate of the confidence intervals of the different methods for various Weibull distributions. We simulate 1000 times 100 data each and calculate the confidence intervals with level 0.95. Figure 3.17 displays the coverage rate and mean length.

Table 3.10.: Confidence intervals for the shape parameter, known scale parameter, $N = 100$, 1000 repetitions each.

			MLE		LRS		MoM		lik-depth	
	a	b	coverage	length	cov.	length	cov.	length	cov.	length
a)	1	1	0.910	0.319	0.968	0.353	0.943	0.485	0.911	0.515
b)	1	0.5	0.936	0.318	0.979	0.347	0.959	0.485	0.939	0.512
c)	0.5	1	0.911	0.159	0.972	0.176	0.944	0.243	0.897	0.257
d)	0.5	0.5	0.915	0.159	0.966	0.174	0.947	0.243	0.897	0.259
e)	3	1	0.926	0.950	0.961	1.028	0.946	1.45	0.919	1.566
f)	1	3	0.920	0.317	0.963	0.345	0.945	0.484	0.909	0.509
g)	3	3	0.930	0.953	0.966	1.019	0.946	1.461	0.915	1.547
h)	3	0.5	0.944	0.95	0.964	1.019	0.957	1.451	0.916	1.517
i)	0.5	3	0.914	0.159	0.970	0.174	0.938	0.243	0.905	0.264

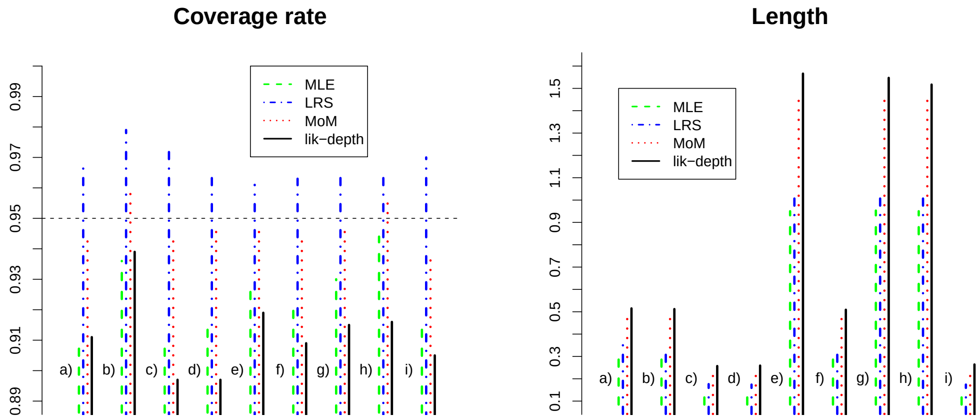

Figure 3.17.: Diagrammed values of the simulated coverage rate and length of the confidence intervals for the shape parameter, when the scale parameter is supposed to be known. For the values see Table 3.10.

We see that for uncontaminated data the confidence intervals based on likelihood-depth perform worst. The rate of coverage is twice even smaller than 0.9. The reason for this is maybe that the test for $H_0 : a \leq a_0$ does not keep the level.

If we consider contaminated data too, we see that the confidence intervals based on the method of medians and the ones based on likelihood-depth are quite robust in contrast to the ones based on the MLE. Table 3.11 shows some results for ε-contaminated data. Again we simulate 1000 times 100 data, where 10% of the data come from another Weibull distribution with parameters a_1, b_1. We also graph the results, see Figure 3.18.

Table 3.11.: Confidence intervals for the shape parameter for ε-contaminated data from $\text{Wei}(a_1, b_1)$, known scale parameter b_0, $N = 100$, 1000 repetitions each.

					MLE		LRS		MoM		lik-depth	
	a	b	a_1	b_1	cov.	length	cov.	l.	cov.	l.	cov.	l.
a)	2	0.1	0.5	10	$< 10^{-2}$	0.14	$< 10^{-2}$	0.28	0.85	0.85	0.85	1.04
b)	0.5	1	2	10	0.93	0.16	0.98	0.22	0.93	0.22	0.87	0.22
c)	0.5	1	2	1	0.76	0.17	0.93	0.18	0.78	0.27	0.80	0.40
d)	2	1	0.5	1	0.07	0.40	0.06	0.41	0.88	0.88	0.89	1.00
e)	1	10^2	1	10^6	0.05	0.23	0.13	0.34	0.88	0.44	0.90	0.52
f)	1	10^2	1	10	0.83	0.28	0.83	0.35	0.85	0.44	0.88	0.48

Table 3.11 and Figure 3.18 show that the method of medians and the method based on likelihood-depth are quite robust, especially when the shape parameter of the contamination distribution is smaller than the main parameter. The confidence intervals based on the maximum likelihood estimator are very bad. The method based on likelihood-depth

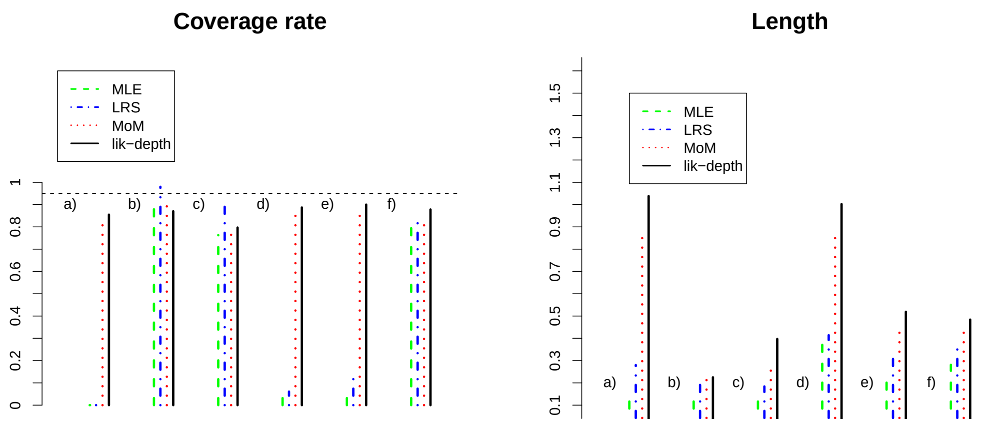

Figure 3.18.: Diagrammed values of the simulated coverage rate and length of the confidence intervals for the shape parameter in contaminated data, when the scale parameter is supposed to be known. For the values see Table 3.11.

gives the best coverage rates in four of the cases regarded. Only in one case it is worse than the method based on the method of medians.

In this subsection tests for the shape parameter, under the assumption that the scale parameter is known, were given. It turns out that these have good asymptotic power, are quite robust and also have a good power for finite sample size. The resulting confidence intervals are also robust against contamination and most times in case of contaminated data superior to the ones given by the other methods considered here.

3.3.2. Uncensored data with unknown scale parameter

If the scale parameter b_0 is unknown and has to be estimated, the depth of a can only be calculated based on the median of the data. To calculate the test statistic, we plug $\tilde{b}_N$ into the simplicial depth instead of b_0. Thus, $d_S^{\tilde{b}_N}(T_*)$ is not a U-statistic any more. We can not use the theorem of Hoeffding to get the asymptotic distribution of the test statistic. Anyway, we develop how the quantities would look like, if we still could use the same theory as before and show in simulations studies that the power is quite good for these disturbed cases. We determine $\tilde{p}_{shape}$ as the asymptotic value for the part of observations lying in $T_{pos}^{a_0,\tilde{b}}$, see also the section about estimation in the uncensored case, Section 3.2.3.

Analog to the case, where the scale parameter is known, we define the test statistic as

$$\tilde{T}(a,t_*) := \sqrt{N}\frac{d_S^{\tilde{b}_N}(a,t_*) - 2\tilde{p}_{shape}(1-\tilde{p}_{shape})}{2\sqrt{\tilde{p}_{shape}(1-\tilde{p}_{shape})(1-2\tilde{p}_{shape})^2}},$$

where here $\tilde{p}_{shape} = 2^{-c_1} - 2^{-c_2} \approx 0.624$, since $\tilde{p}_{shape} := \frac{1}{N} \lim_{N\to\infty} \sharp\{n; y_n \in T_{pos}^{a,\tilde{b}_N}\}$ and with Lemma 3.8, page 41, we get that $T_{pos}^{a,\tilde{b}_N} = [c_1^{\frac{1}{a}} \tilde{b}_N, c_2^{\frac{1}{a}} \tilde{b}_N]$ holds, and Lemma 3.13, page 44, shows $\lim_{N\to\infty} \sharp\{n; y_n \in T_{pos}^{a,\tilde{b}_N}\} = 2^{-c_1} - 2^{-c_2}$.

Thus, we test $H_0 : a \leq a_0$ with $\tilde{\varphi}_{a0}^{0,\leq}(t_*) = 1_{\{\sup_{a\leq a_0} T(a,t_*)<\Phi^{-1}(\alpha)\}}(t_*)$.

The power of this test is, as before, compared to the power of the test based on the MLE. The graphics in Figure 3.19 show the estimated power-functions for various a_0. We simulate data with Weibull distribution Wei(a, b_0) and count how often $H_0 : a \leq a_0$ is rejected. Some of the results for $N = 100$ data and 1000 repetitions for every a can be found in Figure 3.19. We see the estimated power for $a_0 = 0.5, 1$ and 2, where $b_0 = 1$. In the last row on the right the estimated power, if b_0 is not 1 but 0.5, for $H_0 : a \leq 1$, is depicted.

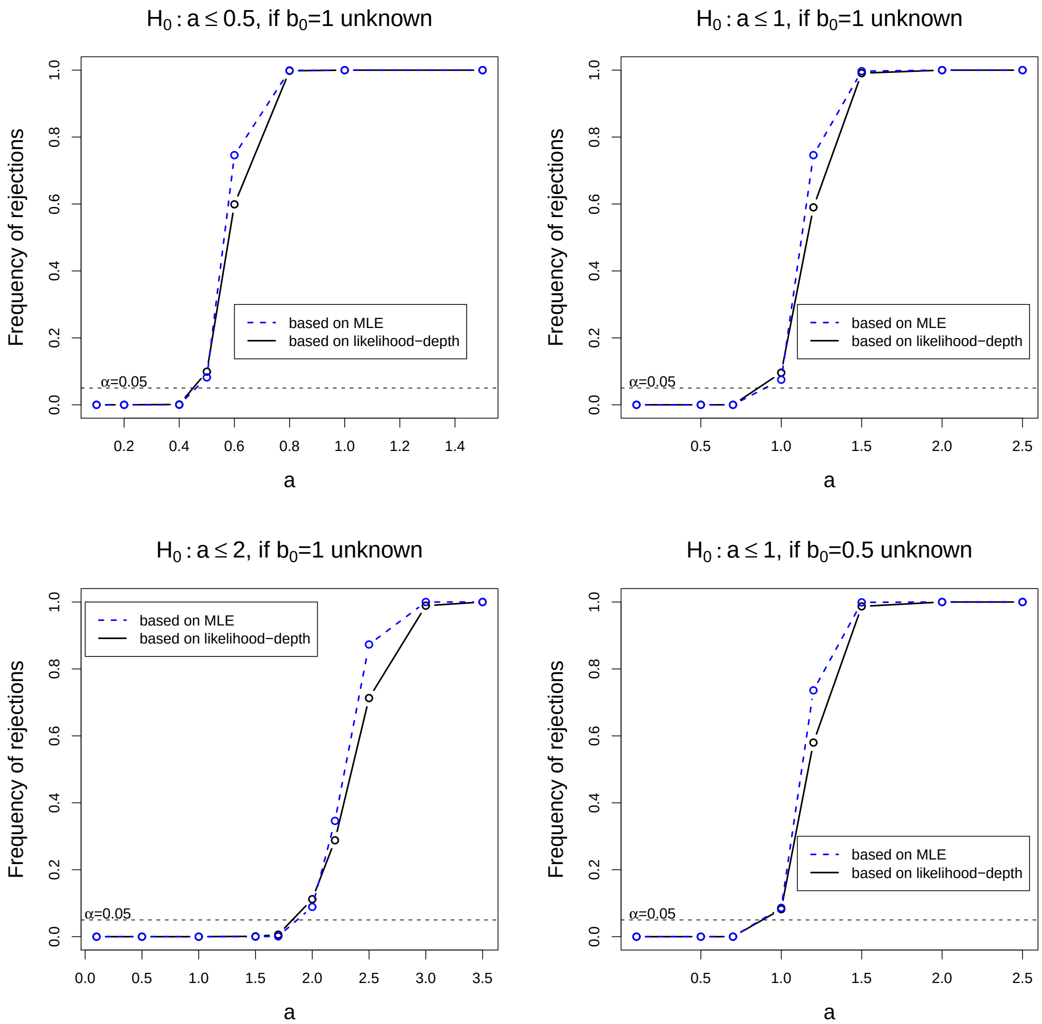

Figure 3.19.: Power of the tests for $H_0 : a \leq a_0$, unknown scale parameter.

If we take a look at the plots of the estimated power-functions in Figure 3.19 and compare them to the ones in Figure 3.11, we realize that the power has not really changed compared to the case, when b_0 is supposed to be known. Still both tests do not keep the level, the only difference is that for b_0 known the ratio of rejections for $a = a_0$ was smaller for the test based on likelihood-depth than for the one based on MLE. For b_0 unknown this fact changes, but the differences are very small.

Once more we consider ε-contaminated data that has a shape parameter $a_1 = 0.5$ and in a second study $a_1 = 10$. The resulting estimated power-functions are displayed in Figure 3.20. The ratio of contamination is 10%.

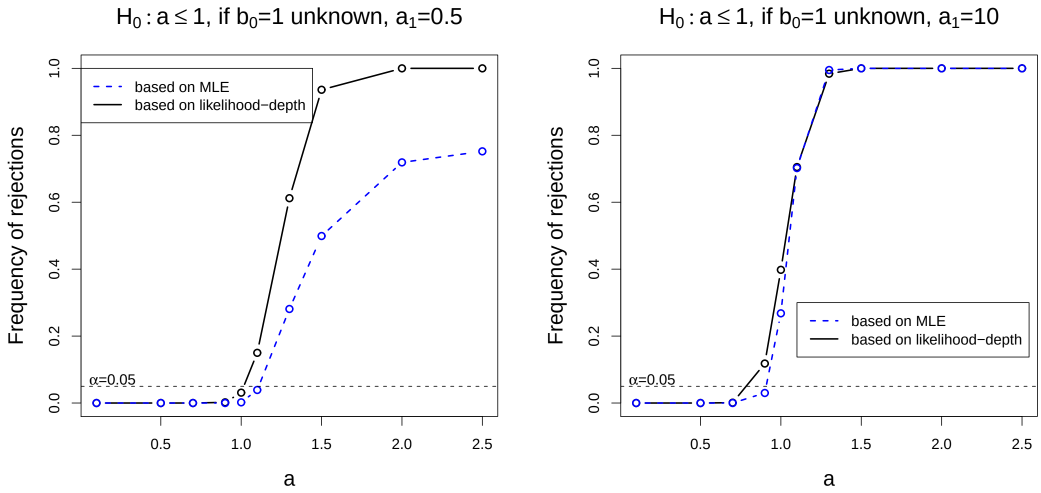

Figure 3.20.: Simulated power of the tests for $H_0 : a \leq a_0$ for ε-contaminated data $\mathrm{Wei}(a_1, b_0)$, $\varepsilon = 0.1$, unknown scale parameter b_0.

Here too, we get almost the same results as in the case, where the scale parameter is known, see Figure 3.12. All in all we see that also when the scale parameter is unknown, we do get a robust test with a good power for $H_0 : a \leq a_0$.

As $\tilde{p}_{shape} \approx 0.624 > 0.5$, a correction for the test $H_0 : a \geq a_0$ is needed. Determining $\tilde{c}^1_\alpha(a_0)$ analog to the case of known scale parameter would lead to

$$\tilde{c}^1_\alpha(a_0) = \tilde{k}_0 a_0,$$

with $\tilde{k}_0 \approx 1.835$: It is $\tilde{c}^1_\alpha(a_0)$ that value a such that $\tilde{p}_{a_0,a} = 1 - \tilde{p}_{shape}$, i.e. $2^{-c_1^{\frac{a_0}{a}}} - 2^{-c_2^{\frac{a_0}{a}}} = 0.376$, what leads to $a = 1.835a_0$. We test $H_0 : a \geq a_0$ against $H_1 : a < a_0$ with

$$\tilde{\varphi}^{\geq}_{a_0} := 1_{\{\sup_{a \geq \tilde{c}^1_\alpha(a_0)} T(a,t_*) \leq \Phi^{-1}(\alpha)\}}(t_*).$$

The graphics of Figure 3.21 show the estimated power of this test for different a_0. For each point in each graphic we do 1000 repetitions with 100 data each.

We realize that the changes in the power of the test with b_0 unknown are very small compared to the test with known scale parameter, which can be found in Figure 3.13.

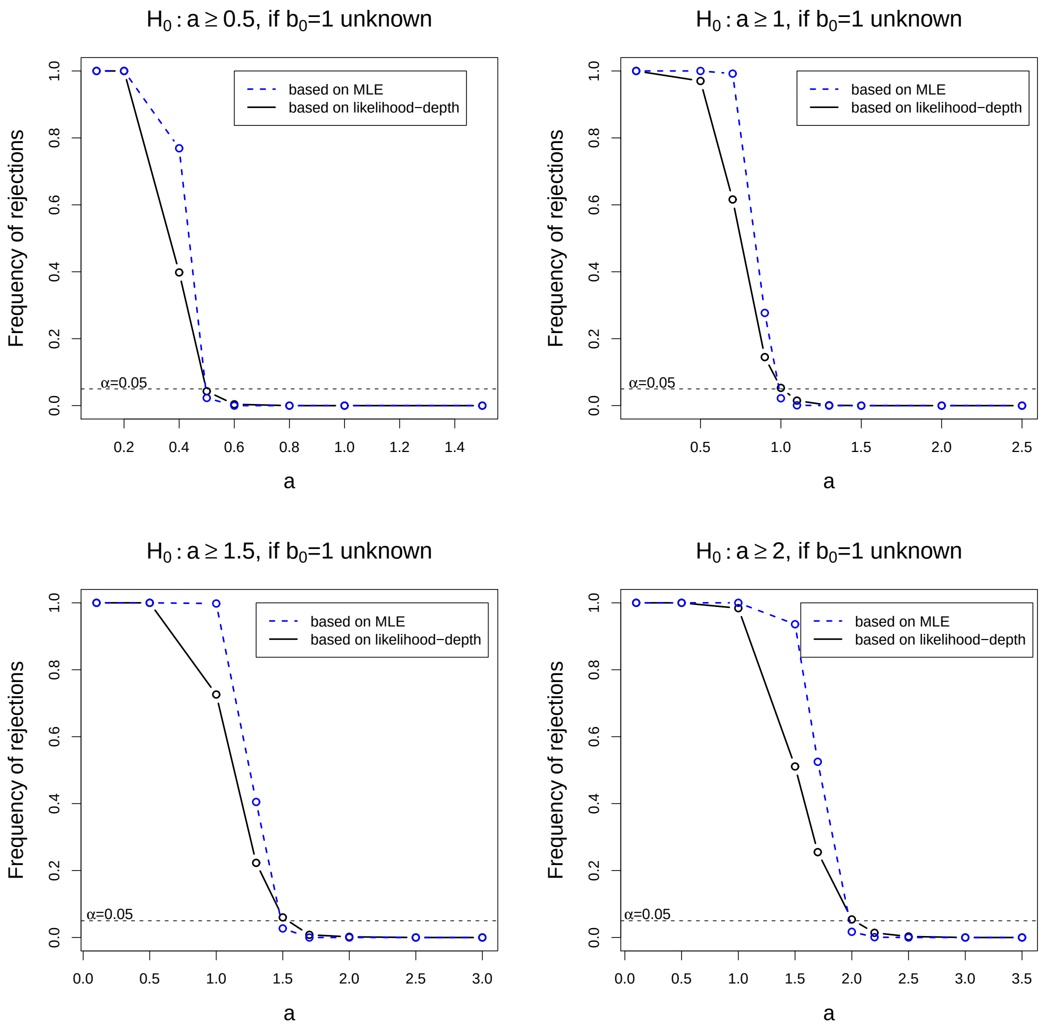

Figure 3.21.: Power of the tests for $H_0 : a \geq a_0$, unknown scale parameter b_0.

We consider also contaminated data, the results of two simulations can be found in Figure 3.22. Here 10% of the data is distributed with $\text{Wei}(a_1, b_1)$. 1000 times 100 data are simulated for each point in the graphics. Comparing this results to the results where the scale parameter is supposed to be known, see Figure 3.14, we see that the test based on likelihood-depth still seems to give the same results.

Using the tests for $H_0 : a \leq a_0$ and $H_0 : a \geq a_0$, we can also define a test for the hypothesis $H_0 : a = a_0$ against $H_1 : a \neq a_0$ as

$$\tilde{\varphi}^{=}_{a_0}(t_*) = \max(1_{\{T(a_0,t_*) \leq \Phi^{-1}(\frac{\alpha}{2})\}}, 1_{\{T(\tilde{c}^1_{\frac{\alpha}{2}}(a_0),t_*) \leq \Phi^{-1}(\frac{\alpha}{2})\}}).$$

The simulated power for this test compared to the simulated power of the test based on MLE and the test based on the method of medians is displayed in Figure 3.23.

The graphics suggest that the power does not really change compared to the case with

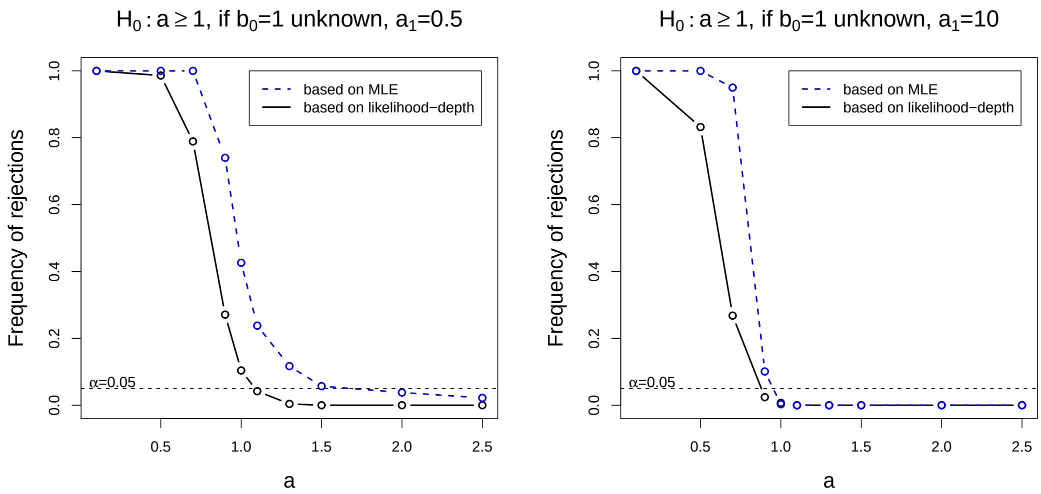

Figure 3.22.: Power of the tests for $H_0 : a \geq a_0$, b_0 unknown, ε-contamination with $\mathrm{Wei}(a_1, b_0)$, $\varepsilon = 0.1$.

known scale parameter in Figure 3.15. Also, when we consider contaminated data, see Figure 3.24, the results do not seem to change in comparison to Figure 3.16.

Using $\tilde{\varphi}^{=}_{a_0}$ we can introduce confidence intervals. We consider them here only for an example. As a real data example we take a look at the steel-lifetime data from Example 3.26.

Example 3.35. We test for the lifetime-data from Example 3.26, $y_* = (4030, 4680, 4860, 5750, 7170, 34100, 51000)$ for exponential distribution, i.e. $H_0 : a = 1$. This is not rejected with level $\alpha = 0.05$, neither for the test based on likelihood-depth nor for the test based on the MLE, but the test based on the method of medians by He and Fung does reject H_0. The confidence intervals for the shape parameter we get by the methods from above are given in Table 3.12.

Table 3.12.: 95%-confidence interval for the shape parameter for the steel data from Example 3.26.

MLE	LRS	MoM	lik-depth
[0.63,2.33]	[0.513,1.587]	[1.015,6.004]	[0.46,4.38]

We see that only for the method of medians the confidence interval does not include 1.

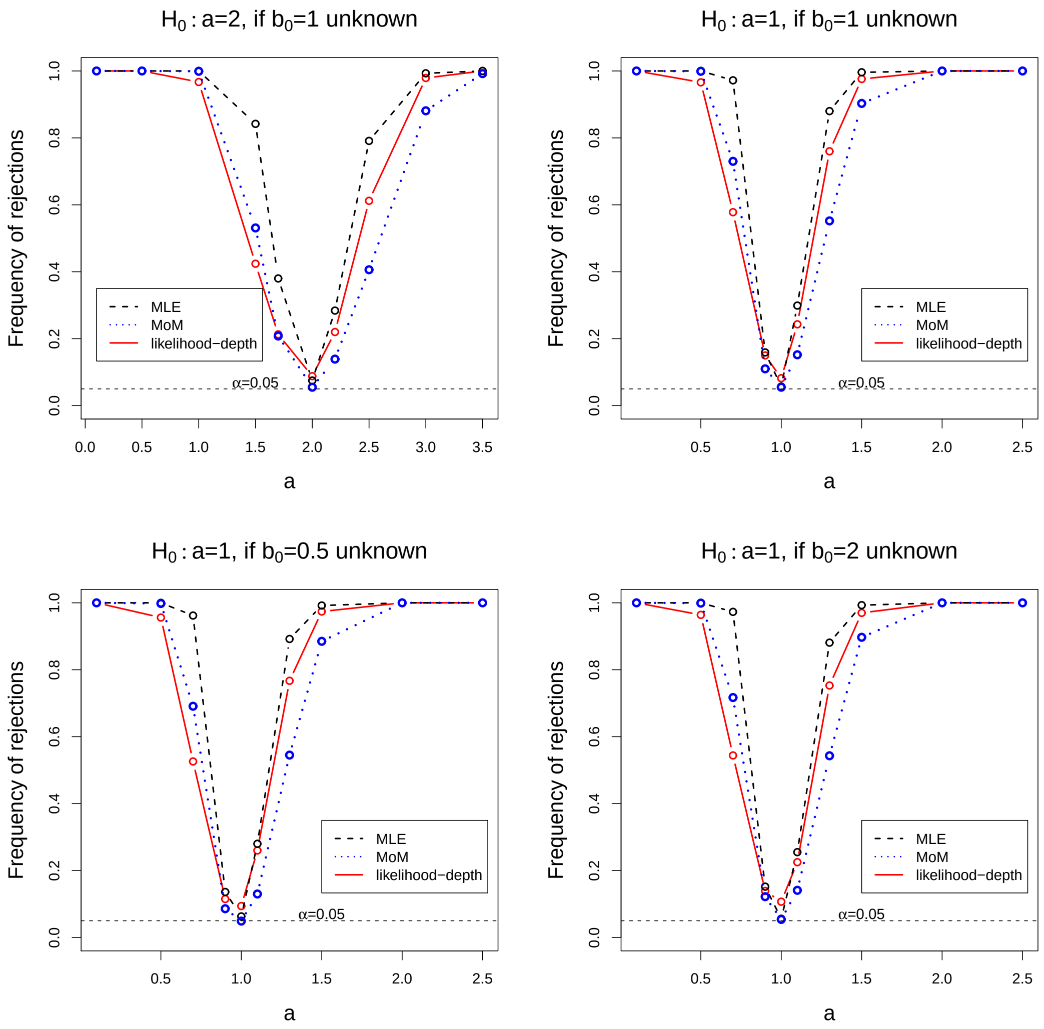

Figure 3.23.: Power of the test for $H_0 : a = a_0$, unknown scale parameter b_0.

3.3.3. Type-I right-censored data with known scale parameter

We consider type-I right-censored data with fixed censor time c_0. We use the notations as described in the beginning of this chapter, i.e. we examine data $z_* = (z_1, \ldots, z_N) = ((y_1, \delta_1), \ldots, (y_N, \delta_N))$, where $y_i = \min(t_i, c_0)$, $i = 1, \ldots, N$, $\delta_i = 0$, if $y_i = c_0$ and $\delta_i = 1$ if $y_i = t_i$. Before we define the test statistic for censored data, we determine $P_{a_0,b_0}(h'_{b_0}(a_0, Y) \geq 0)$. In this subsection, we will assume b_0 to be known.

Lemma 3.36. *For* $Y = \min(T, c_0)$ *and* $T \sim \mathrm{Wei}(a_0, b_0)$*, it holds:*

(a)

$$p^{b_0}_{a_0,c_0} := P_{a_0,b_0}(h'_{b_0}(a_0, Y) \geq 0) = \begin{cases} \exp(-c_1) - \exp(-c_2), & b_0 < c_0 \wedge t_{02}^{a_0,b_0} < c_0 \\ \exp(-c_1) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right), & b_0 < c_0 \wedge t_{02}^{a_0,b_0} \geq c_0 \\ \exp(-c_1), & b_0 \geq c_0 \end{cases},$$

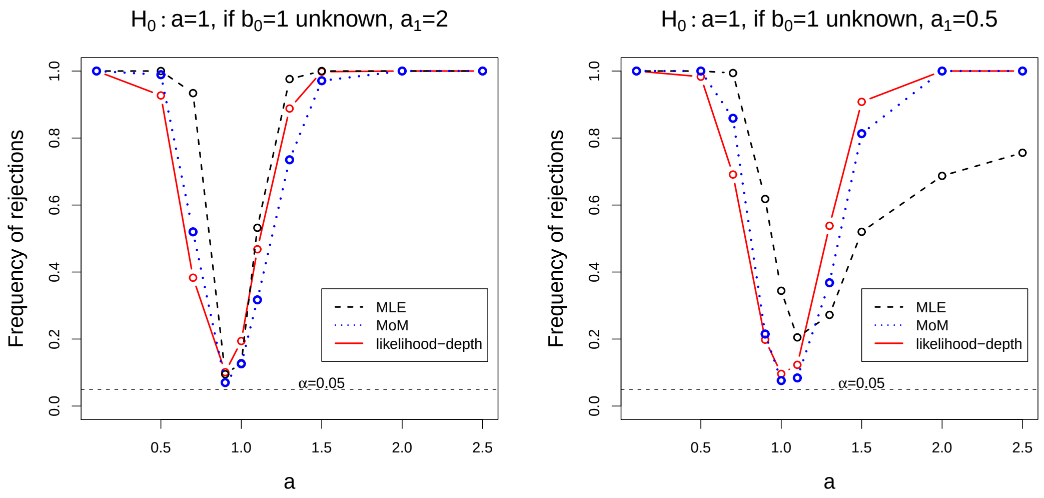

Figure 3.24.: Power of the test for $H_0 : a = a_0$, b_0 unknown, with ε-contamination from Wei(a_1, b_0), $\varepsilon = 0.1$.

where $c_1 < c_2$ *are the solutions of* $\ln c = \frac{1}{c-1}$ *and* $t_{02}^{a_0,b_0} = c_2^{\frac{1}{a_0}} b_0$.

(b) Especially, $p_{a_0,c_0}^{b_0}$ *is continuous.*

Proof: We start with b_0 such that $c_0 > b_0$ and $t_{02}^{a_0,b_0} = c_2^{\frac{1}{a_0}} b_0 < c_0$. Then the likelihood-depth of a_0 in $(z_1, z_2) = ((y_1, \delta_1), (y_2, \delta_2))$ is

$$d_T(a_0, (z_1, z_2)) = 1_{T_{pos}^{a_0,b_0}}(y_1) 1_{T_{neg}^{a_0,b_0}}(y_2) + 1_{T_{neg}^{a_0,b_0}}(y_1) 1_{T_{pos}^{a_0,b_0}}(y_2),$$

where

$$T_{pos}^{a_0,b_0} = [t_{01}^{a_0,b_0}, t_{02}^{a_0,b_0}]$$

as in the uncensored case and $T_{neg}^{a_0,b_0} = (0, t_{01}^{a_0,b_0}] \cup [t_{02}^{a_0,b_0}, c_0]$. Thus, we can use the results from the last subsection for uncensored data, see also Theorem 3.19 on page 55, and get

$$p_{a_0,c_0}^{b_0} = P_{a_0,b_0}(T_{pos}^{a_0,b_0}) = P_{a_0,b_0}(c_1^{\frac{1}{a_0}} b_0 \leq Y \leq c_2^{\frac{1}{a_0}} b_0) = p_{shape} = \exp(-c_1) - \exp(-c_2).$$

Considering b_0 such that $t_{02}^{a_0,b_0} \geq c_0$, we get for the likelihood-depth of a_0 in (z_1, z_2), $d_T(a_0, (z_1, z_2)) = 1_{T_{pos}^{a_0,b_0}}(y_1) 1_{T_{neg}^{a_0,b_0}}(y_2) + 1_{T_{neg}^{a_0,b_0}}(y_1) 1_{T_{pos}^{a_0,b_0}}(y_2)$. Here it is

$$T_{pos}^{a_0,b_0} = \{z = (y, \delta); \delta = 1, t_{01}^{a_0,b_0} \leq y \leq c_0\} = \{z = (y, \delta); t_{01}^{a_0,b_0} \leq y \leq c_0\}$$

and

$$T_{neg}^{a_0,b_0} = \{z = (y, \delta); \delta = 0 \vee y \leq t_{01}^{a_0,b_0}\},$$

see Theorem 3.19. Hence,

$$\begin{aligned} p^{b_0}_{a_0,c_0} &= P_{a_0,b_0}(T^{a_0,b_0}_{pos}) = P_{a_0,b_0}(t^{a_0,b_0}_{01} \le Y < c_0) \\ &= \exp\left(-\left(\frac{c_1^{\frac{1}{a_0}} b_0}{b_0}\right)^{a_0}\right) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) \\ &= \exp(-c_1) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right). \end{aligned}$$

Now only the case $b_0 \ge c_0$ is left. Then the likelihood-depth of a_0 in (z_1, z_2) is given by $d_T(a_0,(z_1,z_2)) = 1_{T^{a_0,b_0}_{pos}}(y_1)1_{T^{a_0,b_0}_{neg}}(y_2) + 1_{T^{a_0,b_0}_{neg}}(y_1)1_{T^{a_0,b_0}_{pos}}(y_2)$, where this time

$$T^{a_0,b_0}_{pos} = [t^{a_0,b_0}_{01}, c_0],$$

see Theorem 3.19. Therefore, the probability that one data lies inside T^{a_0,b_0}_{pos} is given by

$$\begin{aligned} P_{a_0,b_0}(T^{a_0,b_0}_{pos}) &= P_{a_0,b_0}(t^{a_0,b_0}_{01} \le Y \le c_0) = P(T \ge t^{a_0,b_0}_{01}) \\ &= \exp\left(-\left(\frac{c_1^{\frac{1}{a_0}} b_0}{b_0}\right)^{a_0}\right) = \exp(-c_1). \end{aligned}$$

□

Recall that $t^{a_0,b_0}_{01} = c_1^{\frac{1}{a_0}} b_0 < b_0$ and $t^{a_0,b_0}_{02} = c_2^{\frac{1}{a_0}} b_0 > b_0$ since $c_1 < 1 < c_2$. We analyze for which a_0 it is $p^{b_0}_{a_0,c_0} > 0.5$, for which $p^{b_0}_{a_0,c_0} = 0.5$ and for which a_0 $p^{b_0}_{a_0,c_0} < 0.5$. This is done in the next lemma, but see also the Section 3.2.5 about estimation for censored data.

Lemma 3.37. *It is $p^{b_0}_{a_0,c_0} > 0.5$, if*

- *$b_0 \ge c_0$, or*
- *$b_0 < t^{a_0,b_0}_{02} < c_0$, or*
- *$b_0 < c_0 \le t^{a_0,b_0}_{02}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$.*

It holds $p^{b_0}_{a_0,c_0} < 0.5$, if $b_0 < c_0 \le t^{a_0,b_0}_{02}$ and $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$. And $p^{b_0}_{a_0,c_0} = 0.5$, if $b_0 < c_0 \le t^{a_0,b_0}_{02}$ and $a_0 = \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$.

Proof: We already showed in Lemma 3.36 that

$$p^{b_0}_{a_0,c_0} = P_{a_0,b_0}(h'_b(a_0,Y) \ge 0) = \begin{cases} \exp(-c_1) - \exp(-c_2) \approx 0.66, & b_0 < c_0 \text{ and } t^{a_0,b_0}_{02} < c_0 \\ \exp(-c_1) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right), & b_0 < c_0 \text{ and } t^{a_0,b_0}_{02} \ge c_0 \\ \exp(-c_1) \approx 0.77, & b_0 \ge c_0 \end{cases}.$$

For $b_0 > c_0$ and $b_0 < t_{02}^{a_0,b_0} < c_0$ the claim is true. Now let be $b_0 < c_0 \leq t_{02}^{a_0,b_0}$. Then it holds $p_{a_0,c_0}^{b_0} = \exp(-c_1) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right)$, i.e.

$$\begin{aligned}
&p_{a_0,c_0}^{b_0} > 0.5 \\
&\Leftrightarrow \underbrace{\exp(-c_1) - 0.5}_{\approx 0.22<1} > \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) \\
&\Leftrightarrow -\ln(\exp(-c_1) - 0.5) < \left(\frac{c_0}{b_0}\right)^{a_0} \\
&\Leftrightarrow \underbrace{\ln(-\ln(\exp(-c_1) - 0.5))}_{\approx 0.265}(\ln\left(\frac{c_0}{b_0}\right))^{-1} < a_0.
\end{aligned}$$

With this we also get $p_{a_0,c_0}^{b_0} = 0.5$, if $a_0 = \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, and $p_{a_0,c_0}^{b_0} < 0.5$, if $a_0 < \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$. □

Thus, in case of $b_0 < c_0 \leq t_{02}^{a_0,b_0}$ and $a_0 = \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$, the simplicial likelihood depth of a_0 is a degenerated U-statistic, as $p_{a_0,c_0}^{b_0} = 0.5$. In these cases we can not give α-level tests with the methods presented in Section 2.2.

Lemma 3.38. *For $P_{a_0,b_0}(T_{pos}^{a,b_0})$ the following results hold: If $c_0 > b_0$,*

$$p_{a_0,a,c_0}^{b_0} := P_{a_0,b_0}(T_{pos}^{a,b_0}) = \exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0)}{b_0}\right)^{a_0}\right),$$

if $b_0 \geq c_0$,

$$p_{a_0,a,c_0}^{b_0} = \exp\left(-c_1^{\frac{a_0}{a}}\right).$$

Further, $p_{a_0,(\cdot),c_0}^{b_0}$ is strictly decreasing and $p_{(\cdot),a,c_0}^{b_0}$ strictly increasing.

Proof: We already discussed in the proof of Theorem 3.19 on page 55 that

$$\begin{aligned}
T_{pos}^{a,b_0,c_0} &= \{z = (y,\delta); y \in [t_{01}^{a,b_0}, t_{02}^{a,b_0}],\ \delta = 1\},\ \text{if } b_0 < t_{02}^{a,b_0} < c_0, \\
T_{pos}^{a,b_0,c_0} &= \{z = (y,\delta); y \in [t_{01}^{a,b_0}, c_0],\ \delta = 1\},\ \text{if } b_0 < c_0 \leq t_{02}^{a,b_0}, \\
T_{pos}^{a,b_0,c_0} &= \{z = (y,\delta); y \in [t_{01}^{a,b_0}, c_0],\ \delta = 1 \text{ or } y = c_0,\ \delta = 0\},\ \text{if } b_0 \geq c_0.
\end{aligned}$$

Hence, we obtain $p_{a_0,a,c_0}^{b_0} = P_{a_0,b_0}(T_{pos}^{a,b_0}) = \exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{\min(c_0,c_2^{\frac{1}{a}} b_0)}{b_0}\right)^{a_0}\right)$, if $b_0 < c_0$, else $p_{a_0,a,c_0}^{b_0} = \exp\left(-c_1^{\frac{a_0}{a}}\right)$ holds.

Now consider $a' > a$, then $\frac{1}{a'} < \frac{1}{a}$ and therefore $\exp\left(-c_1^{\frac{a_0}{a'}}\right) < \exp\left(-c_1^{\frac{a_0}{a}}\right)$. Further, for $b_0 < c_0$, we have $c_2^{\frac{1}{a'}} < c_2^{\frac{1}{a}}$ and therefore $\frac{\min(c_2^{\frac{1}{a'}} b_0, c_0)}{b_0} \leq \frac{\min(c_2^{\frac{1}{a}} b_0, c_0)}{b_0}$, consequently

$$-\exp\left(-\left(\frac{\min(c_2^{\frac{1}{a'}} b_0, c_0)}{b_0}\right)^{a_0}\right) \leq -\exp\left(-\left(\frac{\min(c_2^{\frac{1}{a}} b_0, c_0)}{b_0}\right)^{a_0}\right).$$

Altogether, this means

$$\exp(-c_1^{\frac{a_0}{a'}}) - \exp(-\left(\frac{\min(c_2^{\frac{1}{a'}} b_0, c_0)}{b_0}\right)^{a_0}) < \exp(-c_1^{\frac{a_0}{a}}) - \exp(-\left(\frac{\min(c_2^{\frac{1}{a}} b_0, c_0)}{b_0}\right)^{a_0}).$$

What proves $p^{b_0}_{a_0,(\cdot),c_0}$ strictly decreasing.

If $a_1 < a_2$, it is $\exp\left(-c_1^{\frac{a_1}{a}}\right) < \exp\left(-c_1^{\frac{a_2}{a}}\right)$ and if $b_0 < c_0$

$$-\exp\left(-\left(\frac{\min(c_2^{\frac{1}{a}} b_0, c_0)}{b_0}\right)^{a_1}\right) < -\exp\left(-\left(\frac{\min(c_2^{\frac{1}{a}} b_0, c_0)}{b_0}\right)^{a_2}\right),$$

so $p^{b_0}_{(\cdot),a,c_0}$ is strictly increasing in both cases. □

Now we determine c^1_α and c^2_α for censored samples.

Lemma 3.39. *(a) If $b_0 < t_{02}^{a_0,b_0} = c_2^{\frac{1}{a_0}} b_0 < c_0$, it holds $c^1_\alpha(a_0) = k_0 \cdot a_0$, with $k_0 \approx 2.275$ as in the uncensored case.*

(b) If $b_0 < c_0 \leq t_{02}^{a_0,b_0} = c_2^{\frac{1}{a_0}} b_0$ and $a_0 < \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, $c^2_\alpha(a_0)$ can be determined as the solution of

$$\exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right),$$

for $a < a_0$ (especially $c^2_\alpha(a_0) < a_0$).

(c) If $b_0 < c_0 \leq t_{02}^{a_0,b_0}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, it is $c^1_\alpha(a_0)$ the solution for a of

$$\exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0)}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right).$$

Further, it holds $c^1_\alpha(a_0) > a_0$.

(d) If $b_0 \geq c_0$, $c^1_\alpha(a_0)$ does not exist.

Proof: (a) If $b_0 < t_{02}^{a_0,b_0} < c_0$, we can use the results from the uncensored case, see Lemma 3.31 on page 76.

(b) Let be $b_0 < c_0 < t_{02}^{a_0,b_0}$ and $a_0 < \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)}$. According to Lemma 3.37, it holds $p^{b_0}_{a_0,c_0} < 0.5$ and in Lemma 3.25 on page 63 we proved that in this situation $s(a_0) < a_0$. Recall that $s(a)$ is that value such that $p^{b_0}_{a,s(a),c_0} = \frac{1}{2}$. So we have to determine c^2_α

to improve the power of the test for $H_0 : a \leq a_0$. For the calculation of c_α^2 we use Lemma 2.27, see page 25. Therefore we check the conditions of this Lemma: It holds $\frac{1}{2} < 1 - p^{b_0}_{a_0,a_0,c_0} < \frac{1}{2} + \frac{1}{\sqrt{8}}$, since

$$\begin{aligned} 1 - p^{b_0}_{a_0,a_0,c_0} &= 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) \\ &< 1 - \exp(-c_1) + \exp(-1) \approx 0.596 \\ &< \frac{1}{2} + \frac{1}{\sqrt{8}} \approx 0.853 \end{aligned}$$

We proved in Lemma 3.38 that $p^{b_0}_{a_0,(\cdot),c_0}$ is strictly decreasing. According to Lemma 2.27, $c_\alpha^2(a_0)$ is calculated as the value a for that

$$p^{b_0}_{a_0,a,c_0} = 1 - p^{b_0}_{a,c_0}$$

holds. As $p^{b_0}_{a_0,a_0,c_0} < 0.5$ and $p^{b_0}_{a_0,(\cdot),c_0}$ strictly decreasing, the solution a must be smaller than a_0. Solving $p^{b_0}_{a_0,a,c_0} = 1 - p^{b_0}_{a,c_0}$ for $a < a_0$ leads to solving

$$\exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right).$$

(c) Assume $b_0 < c_0 \leq t_{02}^{a,b_0}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$. We proved in Lemma 3.37 that $p^{b_0}_{a_0,c_0} > 0.5$ and in Lemma 3.38 we showed $p^{b_0}_{a_0,(\cdot),c_0}$ being strictly decreasing. As $p^{b_0}_{a_0,c_0} < \exp(-c_1) \approx 0.78 < 0.85 \approx \frac{1}{2} + \frac{1}{\sqrt{8}}$, we can use Lemma 2.25, see page 23, to determine $c_\alpha^1(a_0)$ and get it as the value a, such that $1 - p^{b_0}_{a,c_0} = p^{b_0}_{a_0,a,c_0}$. I.e., $a = c_\alpha^1(a_0)$ is the solution of

$$1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right) = \exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0)}{b_0}\right)^{a_0}\right).$$

As $p^{b_0}_{a_0,(\cdot),c_0}$ is strictly decreasing and $1 - p^{b_0}_{a_0,a_0,c_0} < p^{b_0}_{a_0,a_0,c_0}$, we have $c_\alpha^1(a_0) > a_0$.

(d) If $b_0 \geq c_0$ holds,

$$1 - p^{b_0}_{a,c_0} = 1 - \exp(-c_1) = \exp\left(-c_1^{\frac{a_0}{a}}\right) = p^{b_0}_{a_0,a,c_0}$$

has no solution $a > 0$, as $1 - \exp(-c_1) = \exp(-c_1^{\frac{a_0}{a}})$ is equivalent to

$$\ln\left(1 - \exp(-c_1)\right) = -c_1^{\frac{a_0}{a}}. \tag{3.12}$$

Since $-\ln(1 - \exp(-c_1)) \approx 1.477 > 1$ holds and $c_1 < 1$, (3.12) is never true for $a, a_0 > 0$.

□

The proof of consistency of the tests needs c_α^1 and c_α^2 being strictly increasing. But in case of $b_0 < c_0 < c_2^{\frac{1}{a_0}} b_0 = t_{02}^{a_0,b_0}$, this is not easy to see. We only give two examples here where we fix b_0 and c_0 and determine $c_\alpha^1(a_0)$ for $a_0 > \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, resp. $c_\alpha^2(a_0)$ for $a_0 < \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$. The results for $b_0 = 1, c_0 = 2$ and $b_0 = 2, c_0 = 3$ are displayed in Figure 3.25.

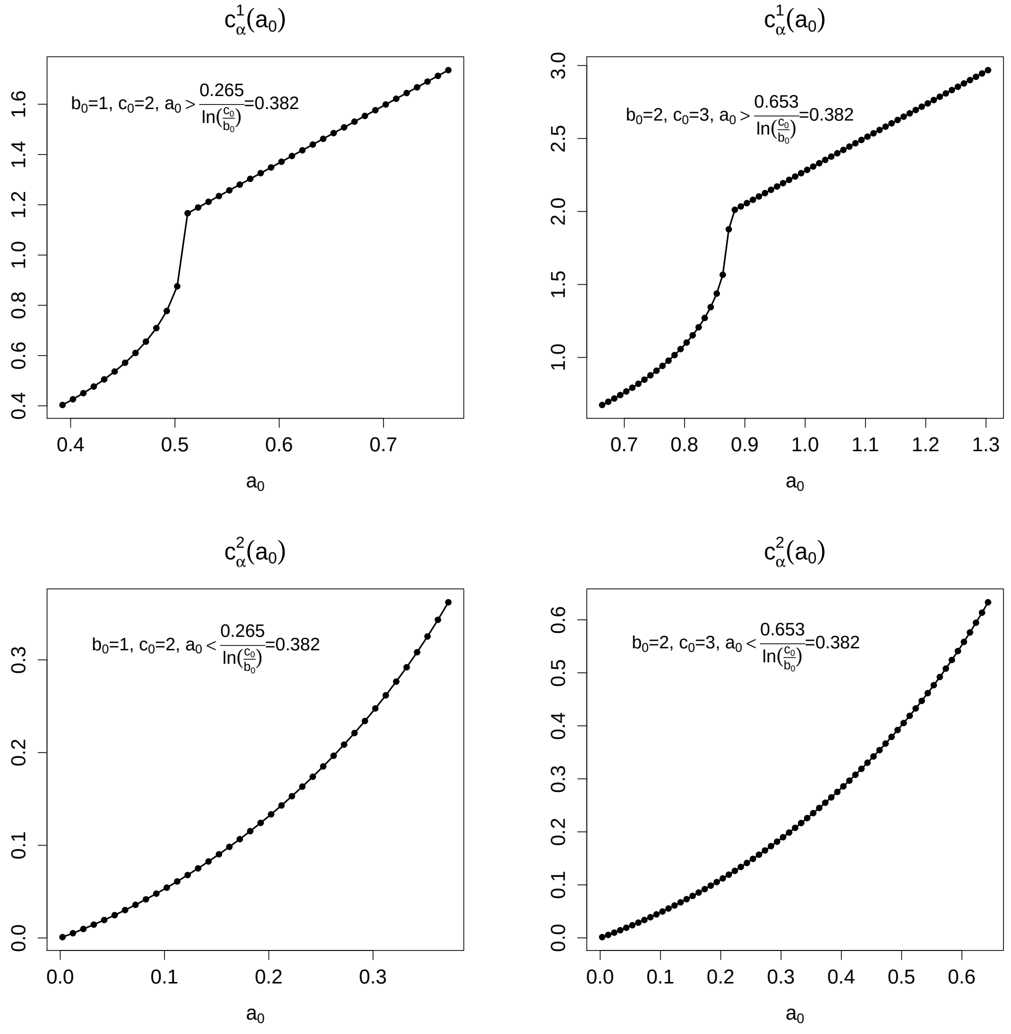

Figure 3.25.: Development of c_α^1 (first row) and c_α^2 (second row) for $b_0 = 1$, $c_0 = 2$ (left) and $b_0 = 2$, $c_0 = 3$ (right).

The graphics provide the assumption, that c_α^1 and c_α^2 are strictly increasing for $a_0 > \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$ resp. $a_0 < \frac{\ln(-\ln(\exp(-c_1)-\frac{1}{2}))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, at least for $c_0 = 2, b_0 = 1$ and $c_0 = 3, b_0 = 2$.

The remaining assumptions of Theorem 2.33 on page 28 resp. Theorem 2.38 on page 31 are true, though, as $p^{b_0}_{(\cdot),a,c_0}$ is strictly increasing, $p^{b_0}_{a,(\cdot),c_0}$ strictly decreasing, $\frac{1}{2} < p^{b_0}_{a_0,c_0} < \frac{1}{2} + \frac{1}{\sqrt{8}}$ and $c^1_\alpha(a_0) > a_0$ resp. $\frac{1}{2} < 1 - p^{b_0}_{a_0,c_0} < \frac{1}{2} + \frac{1}{\sqrt{8}}$ and $c^2_\alpha(a_0) < a_0$ hold.

We define a test statistic analog to Lemma 2.14, see page 17.

Definition 3.40. *We define the test statistic as*

$$T(a_0, z_*) := \sqrt{N} \frac{d^{b_0}_S(a_0, z_*) - 2p^{b_0}_{a_0,c_0}(1 - p^{b_0}_{a_0,c_0})}{2\sqrt{p^{b_0}_{a_0,c_0}(1 - p^{b_0}_{a_0,c_0})(1 - 2p^{b_0}_{a_0,c_0})^2}},$$

with $p^{b_0}_{a_0,c_0}$ given by Lemma 3.36.

We use the theory of the second chapter, especially Definition 2.19 on page 20, Corollary 2.17 on page 19 and Theorem 2.21 on page 20 to construct a test for the hypothesis $H_0 : a \leq a_0$. We already discussed that, as $p^{b_0}_{a_0,c_0}$ can take values smaller than one-half and greater than one-half, we have to distinguish these cases, when determining a test.

Theorem 3.41. *Let be $z_* = (z_1, \dots, z_N)$ realizations of $Z_* = Z_1, \dots, Z_N$ with $Z_i = (Y_i, \Delta_i)$, $Y_i = \min(T_i, c_0)$ and $T_i \sim \mathrm{Wei}(a, b_0)$, $i = 1, \dots, N$.*

(a) If $b_0 < t^{a_0,b_0}_{02} < c_0$, then the test

$$\varphi^{\leq}_{a_0}(z_*) = 1_{\{\sup_{a \leq a_0} T(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*)$$

is a consistent test with asymptotic level α for $H_0 : a \leq a_0$ against $H_1 : a > a_0$.

(b) If $b_0 < c_0 \leq t^{a_0,b_0}_{02}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, then the test

$$\varphi^{\leq}_{a_0}(z_*) = 1_{\{\sup_{a \leq a_0} T(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*)$$

is a test with asymptotic level α for $H_0 : a \leq a_0$ against $H_1 : a > a_0$.

(c) If $b_0 < c_0 \leq t^{a_0,b_0}_{02}$ and $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, then the test

$$\varphi^{\leq}_{a_0}(z_*) = 1_{\{\sup_{a \leq c^2_\alpha(a_0)} T(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*),$$

with $c^2_\alpha(a_0)$ being the solution of

$$\exp(-c_1^{\frac{a_0}{a}}) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right),$$

for $a < a_0$, is a test with asymptotic level α for $H_0 : a \leq a_0$ against $H_1 : a > a_0$.

Proof: If $b_0 < t^{a_0,b_0}_{02} < c_0$, we can use the results from the uncensored case, see Theorem 3.32 on page 77, so (a) is proved. If $b_0 < c_0 \leq t^{a_0,b_0}_{02}$, we have to distinguish the cases $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$, as in the first case it holds $s(a_0) < a_0$

and in the second $s(a_0) > a_0$, see Lemma 3.25 on page 63. Thus, according to the second chapter, we have to correct the test for $H_0 : a \leq a_0$, if $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$. The correction function c_α^2 is given in Lemma 3.39. □

For c_α^2 from Lemma 3.39 we could not prove that it is strictly increasing, hence, we could not prove that the assumptions of Theorem 2.38 on page 31, resp. Theorem 2.33 on page 28, are fulfilled and thereby not prove consistency in both cases $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)}$.

To analyze the behavior of the power-function for finite sample size ($N = 100$) we simulate data with different shape parameter, censor the biggest 20% of the data in each simulation and count how often $H_0 : a \leq a_0$ is rejected for different a_0. The power-function of this new test is compared to the power-function of the test based on the maximum likelihood estimator (MLE). We display the results in the graphics of Figure 3.26.

We see, if one fifth of the data is censored, the new test does not keep the level. The differences between the new test and the test based on the maximum likelihood estimator seem not to be very large for uncontaminated data. We consider later on also contaminated data, when simulating confidence intervals. There we will see that the new test is robust in contrast to the test based on the MLE.

The next aim is to define a test for the hypothesis $H_0 : a \geq a_0$ against $H_1 : a < a_0$. This is given by the next theorem. Again we have to distinguish the cases when $p_{a_0,c_0}^{b_0} > 0.5$ and $p_{a_0,c_0}^{b_0} < 0.5$.

Theorem 3.42. *Let be $\alpha < 0.5$.*

(a) If $c_0 > b_0$ and $c_0 > t_{02}^{a_0,b_0}$, then we get the same correction for the power, c_α^1, as in the uncensored case, i.e.

$$c_\alpha^1(a_0) = k_0 \cdot a_0, \text{ with } k_0 \approx 2.275$$

and a consistent asymptotic α-level test for $H_0 : a \geq a_0$ is given by

$$\varphi_{a_0}^{\geq}(z_*) := 1_{\{\sup_{a \geq c_\alpha^1(a_0)} T(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*).$$

(b) If $b_0 < c_0 \leq t_{02}^{a_0,b_0}$ and $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, then

$$\varphi_{a_0}^{\geq}(z_*) := 1_{\{\sup_{a \geq a_0} T(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*)$$

is a test with asymptotic level α for $H_0 : a \geq a_0$.

(c) If $b_0 < c_0 \leq t_{02}^{a_0,b_0}$ and $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, then

$$\varphi_{a_0}^{\geq}(z_*) = 1_{\{\sup_{a \geq c_\alpha^1(a_0)} T(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*),$$

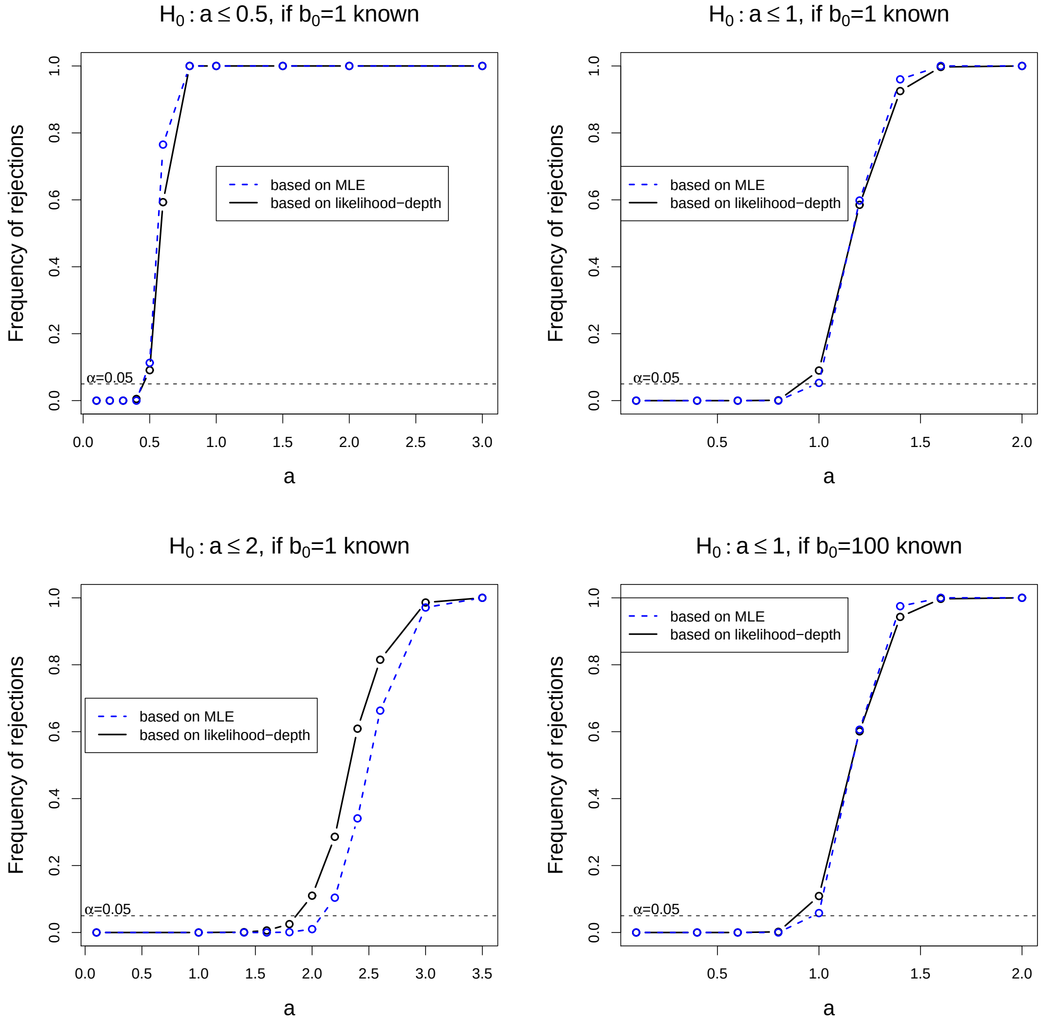

Figure 3.26.: Simulated power of the tests for $H_0 : a \leq a_0$ for 20% right-censored data with known scale parameter b_0.

with $c^1_\alpha(a_0)$ *being the solution for* $a > a_0$ *of*

$$\exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0)}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^a\right),$$

is a test with asymptotic level α *for* $H_0 : a \geq a_0$.

Proof: As we discussed in the proof of Theorem 3.41, we have to distinguish the cases $c_0 > t_{02}^{a_0,b_0}$, where we are in the same situation as for uncensored data and can use the proofs from that section, and $b_0 < c_0 \leq t_{02}^{a_0,b_0}$. In the second situation the test has to be corrected by c^1_α, if $a_0 > \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$. The correction was determined in Lemma 3.39. If $a_0 < \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$, no correction is needed. □

Again we can not proof consistency in case (b) and (c), because we can not show the monotonicity of c_α^1 and c_α^2.

We simulate the power-function of this new test for finite sample size ($N = 100$). As before, we consider 20% of the data to be censored. Here too, we compare the new test with the test based on the MLE. Figure 3.27 shows some of the results for different a_0 and different scale parameter b_0. We see that the test does almost keep the level and that the power does not change as the scale parameter changes. The power of the test based on the MLE seems to be a little better than the power of the new test.

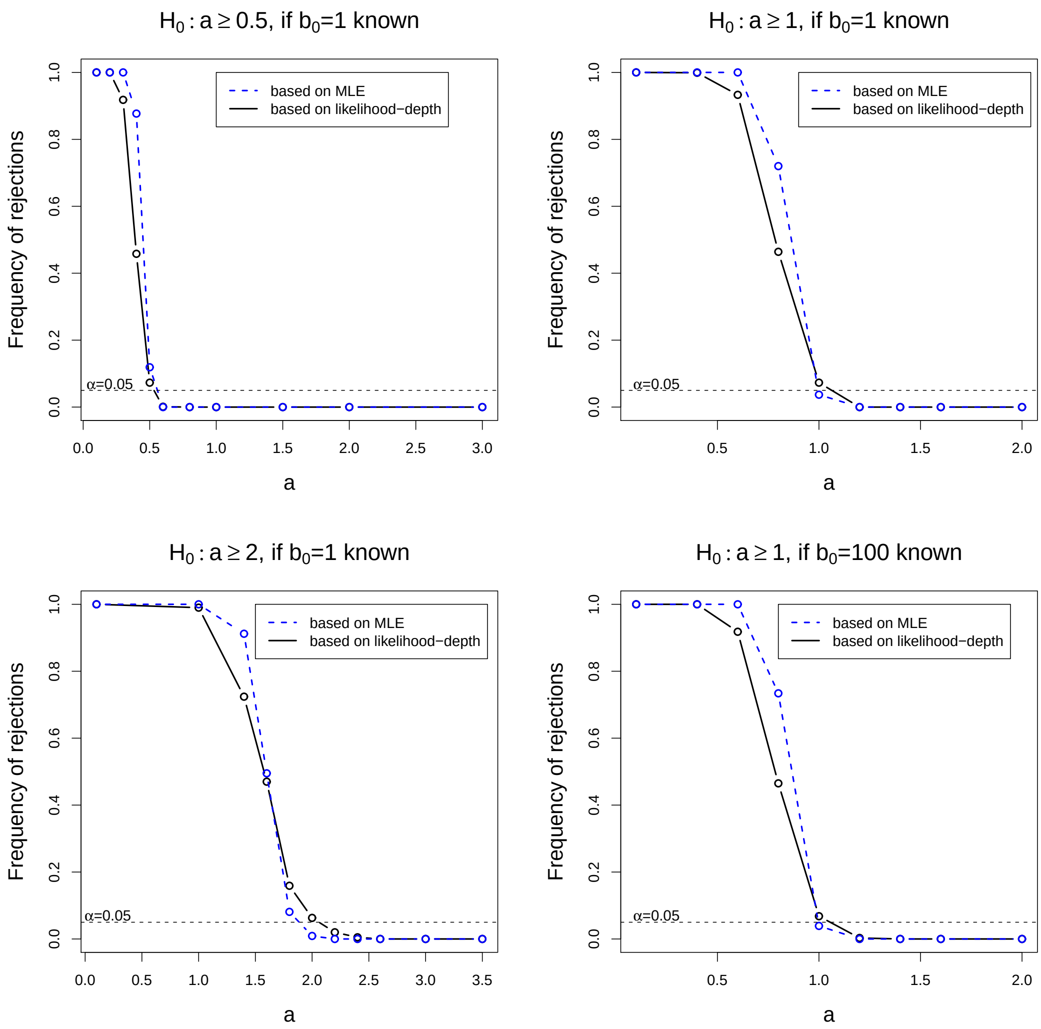

Figure 3.27.: Simulated power of the test for $H_0 : a \geq a_0$ for 20%-right-censored data with known scale parameter b_0.

Up to now, we just censored the biggest 20% of the data. But to force the situation $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln(\frac{c_0}{b_0})}$ resp. $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln(\frac{c_0}{b_0})}$, we will now fix a censor time and a_0, b_0, such that the inequalities are fulfilled. For the test $H_0 : a \leq a_0$ we fix $b_0 = 10$, $c_0 = 13$ and choose $a_0 = 1$, as $1 < 1.01 \approx \frac{0.265}{\ln(\frac{c_0}{b_0})}$ and for $H_0 : a \leq a_0$, we choose $a_0 = 2 > 1.01 \approx \frac{0.265}{\ln(\frac{c_0}{b_0})}$, in both cases it holds $c_2^{\frac{1}{a_0}} b_0 > c_0$. The results of simulations are displayed in Figure 3.28.

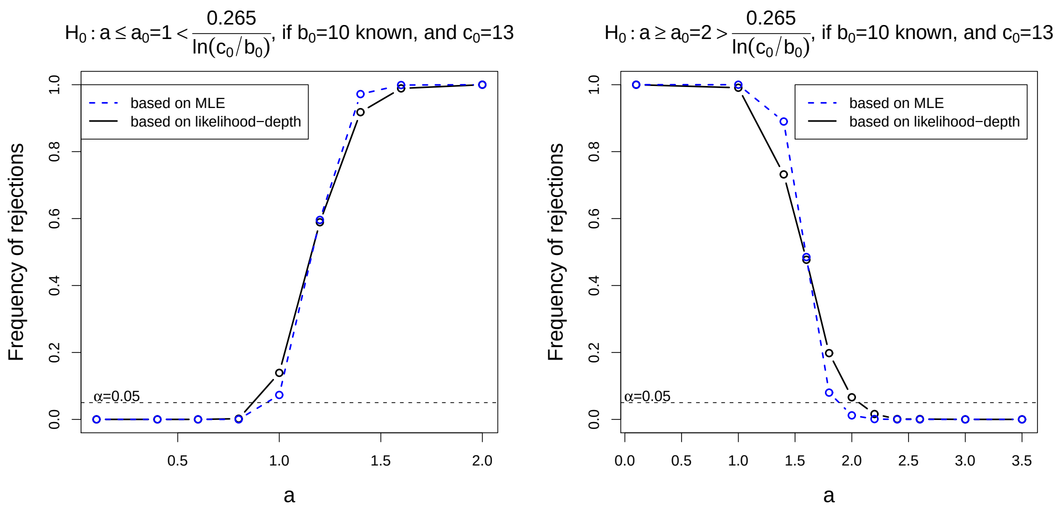

Figure 3.28.: Simulated power of the one-sided tests if $c_2^{\frac{1}{a_0}} b_0 > c_0$ and $a_0 < \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$ (left) and $a_0 > \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$ (right).

The graphics show that the new test and the test based on the MLE do not seem to behave differently in this situation. So here too, the new test is nearly as good as the test based on the MLE.

We use Theorem 2.32 on page 27 resp. Theorem 2.35 on page 30, Lemma 3.38 and Lemma 3.39 to give also a test for the hypothesis $H_0 : a = a_0$ against $H_1 : a \neq a_0$.

Theorem 3.43. *Let be $\alpha < 0.5$, $Z_* = (Z_1, \ldots, Z_N)$, $Z_n = (Y_n, \Delta_n)$, with $Y_n = \min(c_0, T_n)$, $T_n \sim \mathrm{Wei}(a, b_0)$, $n = 1, \ldots, N$.*

(a) *If $c_0 > b_0$ and $c_0 > t_{02}^{a_0,b_0}$, then we can use the same test as in the uncensored case, i.e.*

$$\varphi_{a_0}^{=}(z_*) := \max(1_{\{T(a_0,z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{T(c_{\frac{\alpha}{2}}^1(a_0),z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*)),$$

where

$$c_{\frac{\alpha}{2}}^1(a_0) = k_0 \cdot a_0, \text{ with } k_0 \approx 2.275.$$

$\varphi_{a_0}^{=}$ is a consistent asymptotic α-level test for $H_0 : a = a_0$ against $H_1 : a \neq a_0$.

(b) If $b_0 < c_0 < t_{02}^{a_0,b_0}$ *and* $a_0 < \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$*, then*

$$\varphi_{a_0}^{=}(z_*) := \max(1_{\{T(c^2_{\frac{\alpha}{2}}(a_0),z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{T(a_0,z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*)),$$

where $c^2_{\frac{\alpha}{2}}(a_0)$ *is the solution of*

$$\exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{c_0}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right),$$

and $\varphi_{a_0}^{=}$ *is an asymptotic* α*-level test for* $H_0 : a = a_0$ *against* $H_1 : a \neq a_0$*.*

(c) If $b_0 < c_0 < t_{02}^{a_0,b_0}$ *and* $a_0 > \frac{\ln(-\ln(\exp(-c_1)-0.5))}{\ln\left(\frac{c_0}{b_0}\right)} \approx \frac{0.265}{\ln\left(\frac{c_0}{b_0}\right)}$*, then*

$$\varphi_{a_0}^{=}(z_*) := \max(1_{\{T(a_0,z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{T(c^1_{\frac{\alpha}{2}}(a_0),z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*)),$$

with $c^1_{\alpha}(a_0)$ *being the solution for* $a > a_0$ *of*

$$\exp\left(-c_1^{\frac{a_0}{a}}\right) - \exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0)}{b_0}\right)^{a_0}\right) = 1 - \exp(-c_1) + \exp\left(-\left(\frac{c_0}{b_0}\right)^{a}\right),$$

and $\varphi_{a_0}^{=}$ *is an asymptotic* α*-level test for* $H_0 : a = a_0$ *against* $H_1 : a \neq a_0$*.*

Consequently, in all cases a confidence interval for a is given by

$$\{a_0; \varphi_{a_0}^{=}(z_*) = 0\}.$$

Proof: Use analog arguments as in the proofs of Theorem 3.41 and Theorem 3.42. □

We examine the power-function of the test for $H_0 : a = a_0$ in comparison to the tests based on the maximum likelihood estimator (MLE), see e.g. the textbook of Lawless [Law 2003], and the test based on the method of medians (MoM), see He and Fung [HeFu 1999], in a simulation study. Again the ratio of censored data is 20% and we consider different shapes and scales.

The graphics in Figure 3.29 show that for a sample size of $N = 100$ the level is not kept by the tests based on the MoM and the new test. The test based on the MoM behaves worst, its power function seems to be shifted to the left. For the MLE the power function takes very low values in a_0, for $a > a_0$ it is always below the power function of the new test based on likelihood-depth. Again, we see that changes in the scale do not affect the power of the new test and that it is good for all a_0 considered here.

Also, we simulate 95%-confidence intervals for the shape parameter with the new test (lik-depth) and compare the results to the confidence intervals we get by the method based on the maximum likelihood estimator (MLE) and the one based on the method of medians (MoM). For uncontaminated censored data (again the biggest 20% of the data was censored), the method based on the MLE gives the highest coverage rates and the

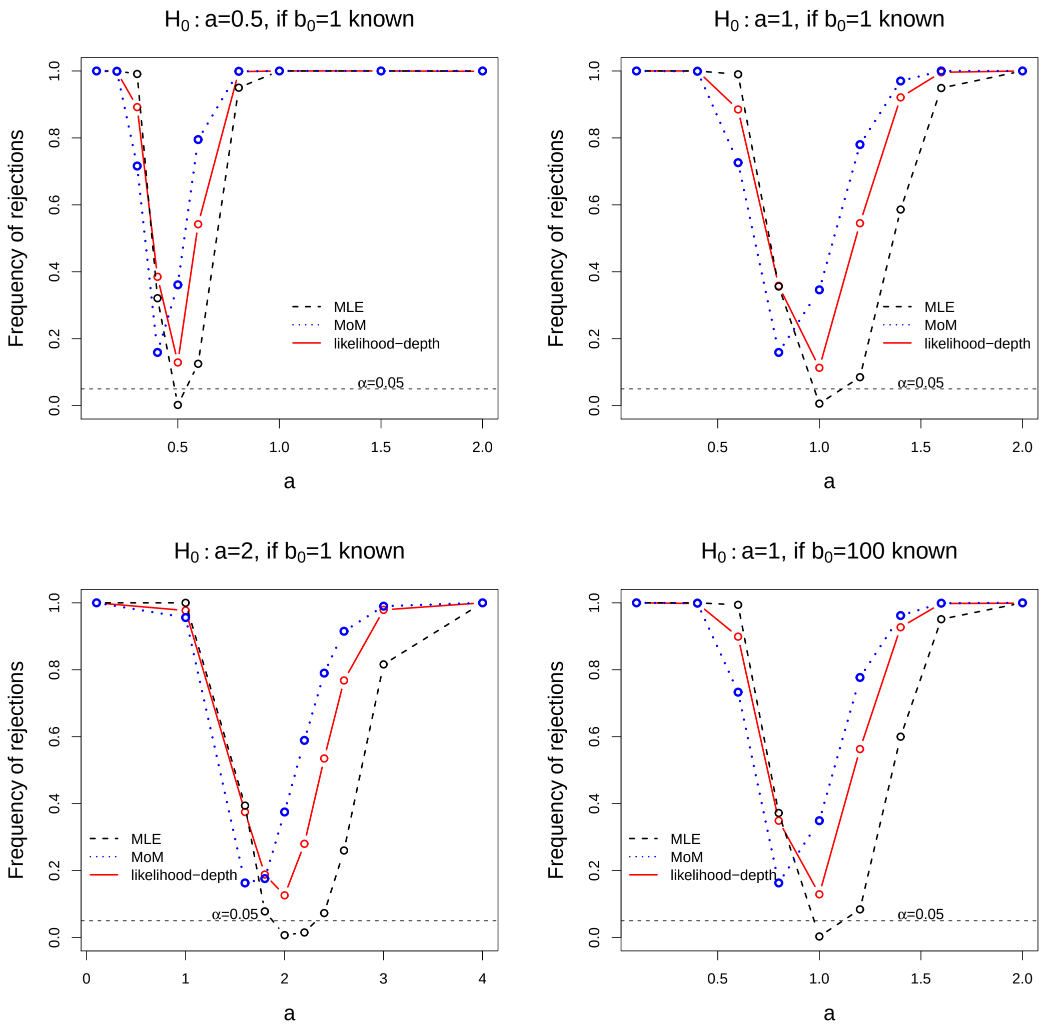

Figure 3.29.: Simulated power of the tests for $H_0 : a = a_0$ for censored data with known scale b_0, 20% of the data censored.

method based on the medians behaves worst. For a better view, we graph the results in Figure 3.30. We notice that the confidence intervals based on likelihood-depth have the smallest mean length in all cases.

In a second study we also consider ε-contaminated data, see Table 3.14. The contamination distribution is Wei(a_1, b) and $\varepsilon = 0.1$. In cases of contamination with $a_1 < 1$ and original $a > 1$, the confidence intervals based on likelihood-depth have the best covering rate nearest to 95%. The coverage rate of the MoM goes down to less than 25 % for censored data, so that this method seems not practical at all. The method of the MLE produces for uncontaminated covering rates that are very close to one, for some contaminations this happens too. Thus, it seems very conservative. But the coverage rate also goes down to less than five percent. We display the results in two graphics, that can be found in Figure 3.31. They reveal, that only the coverage rates of the confidence intervals

Table 3.13.: Confidence intervals for the shape parameter, censored data (20%), known scale b_0.

			MLE		MoM		lik-depth	
a)	1	1	0.997	0.568	0.657	0.578	0.866	0.503
b)	1	3	0.996	0.569	0.626	0.579	0.844	0.5
c)	1	0.2	0.996	0.565	0.655	0.575	0.868	0.504
d)	0.2	0.2	0.994	0.115	0.622	0.116	0.816	0.106
e)	0.2	3	0.998	0.115	0.629	0.116	0.837	0.104
f)	0.2	1	0.998	0.115	0.616	0.117	0.839	0.104
g)	3	3	0.994	1.675	0.635	1.746	0.868	1.561
h)	3	1	0.996	1.678	0.627	1.748	0.857	1.537
i)	3	0.2	0.997	1.674	0.632	1.733	0.879	1.519

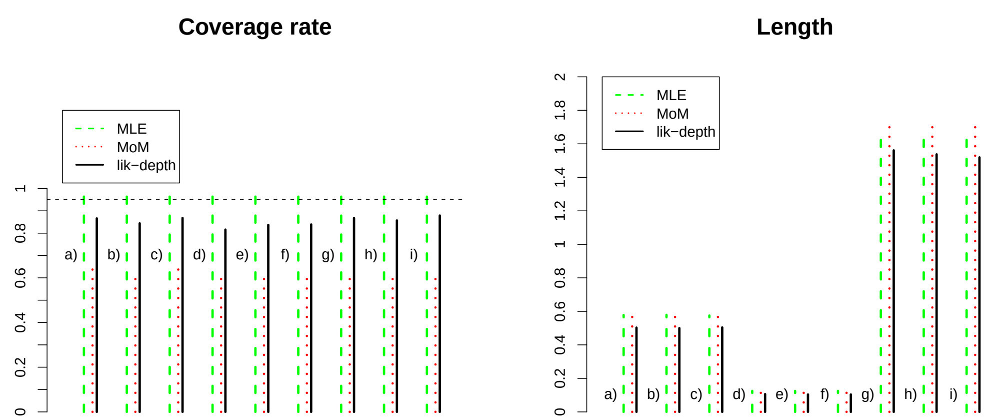

Figure 3.30.: Coverage rate and mean length of the confidence intervals for the shape parameter in censored data, see also Table 3.13.

based on likelihood-depth are stable.

Table 3.14.: Confidence intervals for the shape parameter, contaminated (10% with $\text{Wei}(a_1, b_1)$) and censored (20%) data, known scale b_0.

				MLE		MoM		lik-depth	
	a	b	a_1	cov.	length	cov.	length	cov.	length
a)	1	1	0.2	0.427	0.437	0.819	0.527	0.844	0.472
b)	1	1	2	0.997	0.603	0.429	0.63	0.774	0.57
c)	0.5	1	0.2	0.92	0.256	0.736	0.273	0.883	0.246
d)	0.5	1	2	0.985	0.314	0.245	0.345	0.649	0.364
e)	5	1	0.2	0.005	1.086	0.824	2.595	0.805	2.383
f)	5	1	2	0.807	2.434	0.789	2.697	0.834	2.421

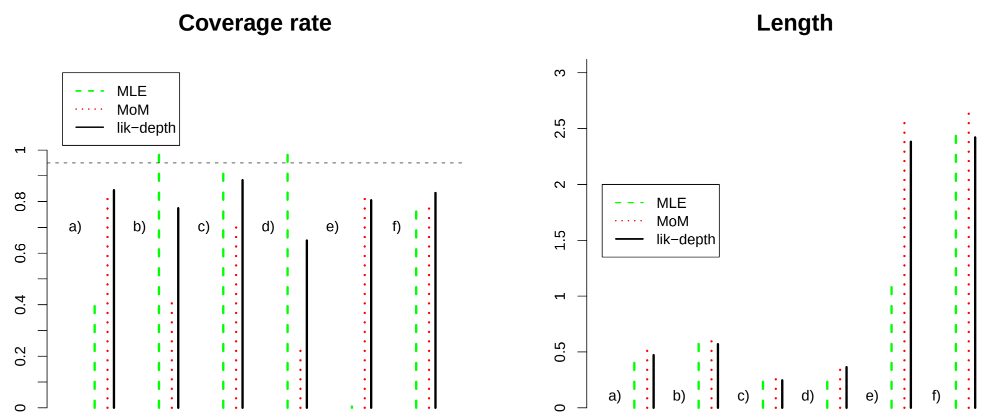

Figure 3.31.: Coverage rate and length of the confidence intervals for the shape parameter in censored and contaminated data, see also Table 3.14.

3.3.4. Type-I right-censored data with unknown scale parameter

If the scale parameter is unknown, the same problems as described in Section 3.3.2, see page 84, occur. We can not use the same theory as in the case, where the scale b_0 is known.

We plug $\tilde{b}_N = \text{med}(y_*)$ into the depth-function of a and calculate the latter as

$$\begin{aligned} d_T^{\tilde{b}_N}(a, z_*) &= \frac{1}{N} \min\Big(\sharp\{n; \delta_n = 1 \text{ and } t_{01}^{a,\tilde{b}_N} \le y_n \le t_{02}^{a,\tilde{b}_N}\}, \\ &\qquad \sharp\{n; \delta_n = 1 \text{ and } (y_n \ge t_{02}^{a,\tilde{b}_N} \text{ or } y_n \le t_{01}^{a,\tilde{b}_N})\} + (N-k)\Big), \end{aligned}$$

see Theorem 3.19 on page 55, where $t_{0i}^{a,\tilde{b}_N} = c_i^{\frac{1}{a}} \tilde{b}_N$, $i = 1, 2$. It was already discussed in Section 3.2.5 about estimation that

$$\frac{1}{N}\sharp\{n; \delta_n = 1 \text{ and } t_{01}^{a,\tilde{b}_N} \le y_n \le t_{02}^{a,\tilde{b}_N}\} \to -\exp\left(-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} b_0 (\ln 2)^{\frac{1}{a}})}{b_0}\right)^{a_0}\right) + 2^{-c_1^{\frac{a_0}{a}}}$$

as N tends to infinity. Because the limit is depending on the unknown parameter b_0, we use that it is approximated by $\frac{\tilde{b}_N}{(\ln 2)^{\frac{1}{a}}}$ and work with

$$p_{a,c_0}^{\tilde{b}_N} := \begin{cases} 2^{-c_1} - 2^{-c_2}, & c_2^{\frac{1}{a}} \tilde{b}_N < c_0 \\ 2^{-c_1} - 2^{-\left(\frac{c_0}{\tilde{b}_N}\right)^a}, & c_2^{\frac{1}{a}} \tilde{b}_N \ge c_0 \end{cases}.$$

Analog to Lemma 2.14, see page 17, the test statistic is defined as

$$\widetilde{T}(a, z_*) := \sqrt{N} \frac{d_S^{\tilde{b}_N}(a, z_*) - 2p_{a,c_0}^{\tilde{b}_N}(1 - p_{a,c_0}^{\tilde{b}_N})}{2\sqrt{p_{a,c_0}^{\tilde{b}_N}(1 - p_{a,c_0}^{\tilde{b}_N})(1 - 2p_{a,c_0}^{\tilde{b}_N})^2}}$$

and we work with this, as if we could use the same theory as in the case of b_0 known in the last subsection.

We begin with checking for which $a > 0$ it is $p_{a,c_0}^{\tilde{b}_N} > 0.5$, see also Lemma 3.25 on page 63. If $c_0 < t_{02}^{a,\tilde{b}_N} = c_2^{\frac{1}{a}} \tilde{b}_N$, it is $p_{a,c_0}^{\tilde{b}_N} = 2^{-c_1} - 2^{-c_2} \approx 0.624 > 0.5$. Now let be $t_{02}^{a,\tilde{b}_N} > c_0$. Then

$$\begin{aligned} & p_{a,c_0}^{\tilde{b}_N} > \frac{1}{2} \Leftrightarrow 2^{-c_1} - 2^{-\left(\frac{c_0}{\tilde{b}_N}\right)^a} > \frac{1}{2} \\ \Leftrightarrow\ & -\frac{\ln \overbrace{\left(2^{-c_1} - 2^{-1}\right)}^{=0.335<1}}{\ln(2)} < \left(\frac{c_0}{\tilde{b}_N}\right)^a \\ \Leftrightarrow\ & \underbrace{\ln\left(-\frac{\ln\left(2^{-c_1} - 2^{-1}\right)}{\ln(2)}\right)}_{:=k_1 \approx 0.455} \frac{1}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)} < a. \end{aligned}$$

Thus, if $c_0 < c_2^{\frac{1}{a_0}} \tilde{b}_N = t_{02}^{a_0,\tilde{b}_N}$ and $a_0 > \frac{k_1}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)}$ or if $c_0 > t_{02}^{a_0,\tilde{b}_N}$, it holds $s(a_0) > a_0$ (see Lemma 3.25) and $p_{a_0,c_0}^{\tilde{b}_N} > \frac{1}{2}$. We can use

$$\tilde{\varphi}_{a_0}^{\le}(z_*) := 1_{\{\sup_{a \le a_0} \tilde{T}(a, z_*) < \Phi^{-1}(\alpha)\}}(z_*)$$

as a test for $H_0 : a \leq a_0$ against $H_1 : a > a_0$. But if $c_0 < t_{02}^{a_0,\tilde{b}_N}$ and $a_0 < \frac{k_1}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)} \approx \frac{0.455}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)}$, we determine $\tilde{c}_\alpha^2(a_0)$ as an analogon to c_α^2 from the last subsection, to improve the power. Therefore, we determine an approximation of $P_{a_0,b_0}(T_{pos}^{a,b})$ for $b = (\ln 2)^{\frac{1}{a}} b_0$, with the same arguments as before and end up with

$$p_{a_0,a,c_0}^{\tilde{b}_N} := 2^{-c_1^{\frac{a_0}{a}}} - 2^{-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} \tilde{b}_N)}{\tilde{b}_N}\right)^{a_0}}.$$

Regard that, if $c_2^{\frac{1}{a_0}} \tilde{b}_N \geq c_0$ and $a \leq a_0$, then $c_2^{\frac{1}{a}} \tilde{b}_N \geq c_2^{\frac{1}{a_0}} \tilde{b}_N \geq c_0$. Hence, it holds $p_{a_0,a,c_0}^{\tilde{b}_N} = 2^{-c_1^{\frac{a_0}{a}}} - 2^{-\left(\frac{c_0}{\tilde{b}_N}\right)^{a_0}}$, if $c_2^{\frac{1}{a_0}} \tilde{b}_N \geq c_0$ and $a \leq a_0$. Analog to the calculations in Lemma 3.39 on page 93 we determine $\tilde{c}_\alpha^2(a_0)$ as that value a that solves

$$p_{a_0,a,c_0}^{\tilde{b}_N} = 1 - p_{a,a,c_0}^{\tilde{b}_N},$$

as it is $p_{a_0,(\cdot),c_0}^{\tilde{b}_N}$ strictly decreasing and $p_{(\cdot),a,c_0}^{\tilde{b}_N}$ strictly increasing. Solving $p_{a_0,a,c_0}^{\tilde{b}_N} = 1 - p_{a,a,c_0}^{\tilde{b}_N}$ means solving

$$2^{-c_1^{\frac{a_0}{a}}} - 2^{-\left(\frac{\min(c_0, c_2^{\frac{1}{a}} \tilde{b}_N)}{\tilde{b}_N}\right)^{a_0}} = 1 - 2^{-c_1} + 2^{-\left(\frac{\min(c_0, c_2^{\frac{1}{a}})}{\tilde{b}_N}\right)^{a}} \tag{3.13}$$

for a. As $p_{a_0,(\cdot),c_0}^{\tilde{b}_N}$ is strictly decreasing, $p_{a_0,a_0,c_0}^{\tilde{b}_N} < 0.5$, the solution $\tilde{c}_\alpha^2(a_0)$ is smaller than a_0. Thus (3.13) is equivalent to

$$2^{-c_1^{\frac{a_0}{a}}} - 2^{-\left(\frac{c_0}{\tilde{b}_N}\right)^{a_0}} = 1 - 2^{-c_1} + 2^{-\left(\frac{c_0}{\tilde{b}_N}\right)^{a}}.$$

We define a test for $H_0 : a \leq a_0$ against $H_1 : a > a_0$ for unknown scale parameter as

$$\tilde{\varphi}_{a_0}^{\leq}(z_*) := \begin{cases} 1_{\{\sup_{a \leq \tilde{c}_\alpha^2(a_0)} \tilde{T}(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*), & c_2^{\frac{1}{a_0}} \tilde{b}_N \geq c_0 \text{ and } a_0 < \frac{k_1}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)} \\ 1_{\{\sup_{a \leq a_0} \tilde{T}(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*), & \text{else} \end{cases},$$

where $k_1 = \ln\left(-\frac{\ln\left(2^{-c_1} - 2^{-1}\right)}{\ln(2)}\right) \approx 0.455$.

The power-function of this new test is simulated for sample size $N = 100$, 20% right-censored data and different shape and scale parameter. Figure 3.32 shows the power-function in comparison to the power-function of the test where b_0 is known.

We see that the power-functions do not really differ. It is even hard to distinguish the dashed, blue line, the power-function of the test for known scale, and the solid, black one that displays the simulated power-function of the test for unknown scale parameter.

Before we can also give a test for the hypothesis $H_0 : a \geq a_0$, the quantity $\tilde{c}_\alpha^1(a_0)$, as an analogon of $c_\alpha^1(a_0)$, has to be determined for the cases $c_2^{\frac{1}{a_0}} \tilde{b}_N < c_0$ as well as $c_2^{\frac{1}{a_0}} \tilde{b}_N \geq c_0$

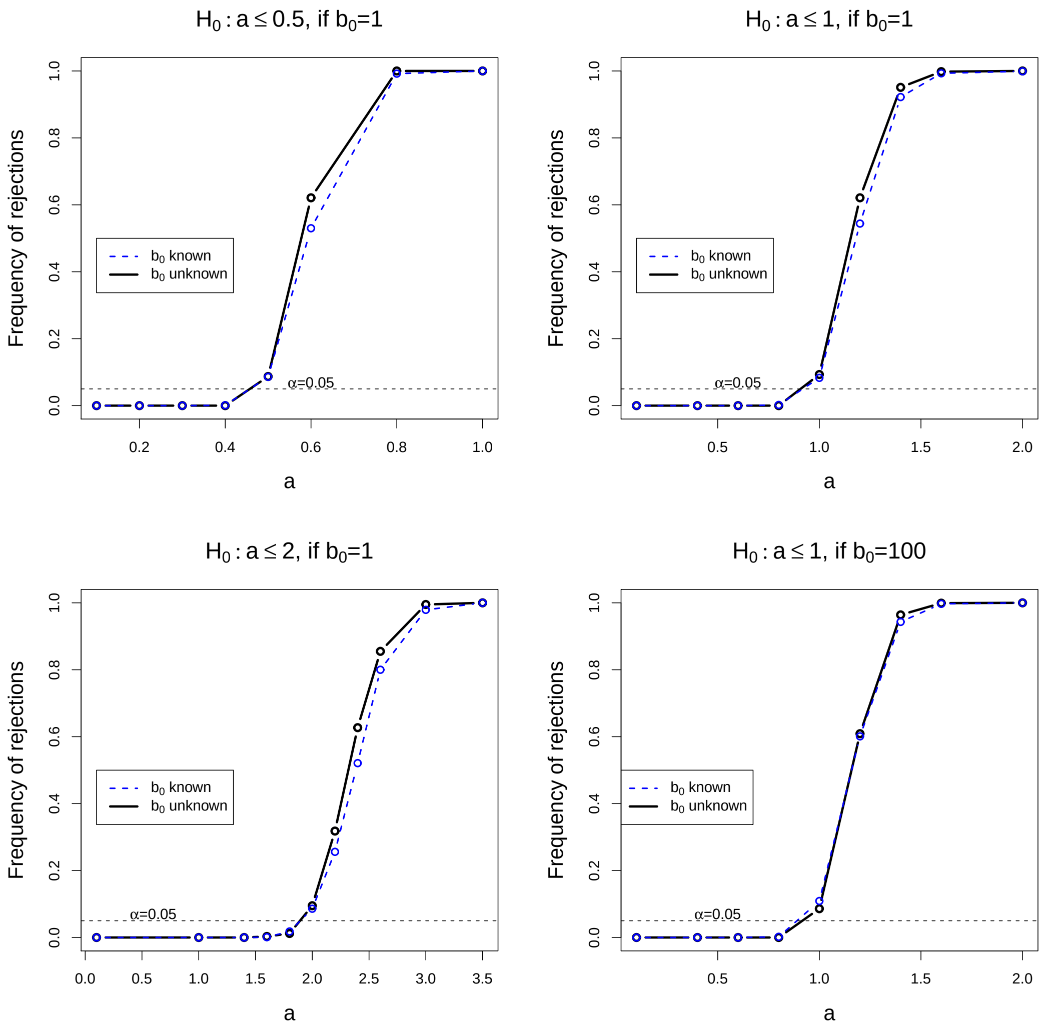

Figure 3.32.: Simulated power of $H_0 : a \leq a_0$ for 20% right-censored data, unknown scale parameter compared to known scale.

and $a_0 > \frac{k_1}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)} \approx \frac{0.455}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)}$, because in these situations it is $s(a_0) > a_0$ and $p^{\tilde{b}_N}_{a_0,c_0} > 0.5$ and we discussed in the second chapter that then the test for the hypothesis $H_0 : a \geq a_0$ has to be corrected. As before, it holds $p^{\tilde{b}_N}_{a_0,(\cdot),c_0}$ strictly decreasing and $p^{\tilde{b}_N}_{(\cdot),a,c_0}$ strictly increasing. For $c_2^{\frac{1}{a_0}}\tilde{b}_N < c_0$ we use the results from the uncensored case, where the scale parameter is unknown. Thus, we end up with

$$\tilde{c}^1_\alpha(a_0) = \tilde{k}_0 a_0,\ \tilde{k}_0 \approx 1.835.$$

If $c_2^{\frac{1}{a_0}}\tilde{b}_N \geq c_0$ and $a_0 > \frac{k_1}{\ln\left(\frac{c_0}{\tilde{b}_N}\right)}$, it is $\tilde{c}_\alpha^1(a_0)$ the solution of $1 - p_{a,a,c_0}^{\tilde{b}_N} = p_{a_0,a,c_0}^{\tilde{b}_N}$, i.e. of

$$1 - 2^{-c_1} + 2^{-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N)}{\tilde{b}_N}\right)^a} = 2^{-c_1^{\frac{a_0}{a}}} - 2^{-\left(\frac{\min(c_0, c_2^{\frac{1}{a}}\tilde{b}_N)}{\tilde{b}_N}\right)^{a_0}}$$

for $a > a_0$. Consequently, the test for $H_0 : a \geq a_0$ against $H_1 : a < a_0$ is given by

$$\tilde{\varphi}_{a_0}^{\geq}(z_*) = \begin{cases} 1_{\{\sup_{a \geq \tilde{k}_0 a_0} \tilde{T}(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*), & c_2^{\frac{1}{a_0}}\tilde{b}_N < c_0 \\ 1_{\{\sup_{a \geq a_0} \tilde{T}(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*), & c_2^{\frac{1}{a_0}}\tilde{b}_N \geq c_0 \text{ and } a_0 < \frac{k_1}{\ln \frac{c_0}{\tilde{b}_N}} \\ 1_{\{\sup_{a \geq \tilde{c}_\alpha^1(a_0)} \tilde{T}(a,z_*) < \Phi^{-1}(\alpha)\}}(z_*), & c_2^{\frac{1}{a_0}}\tilde{b}_N \geq c_0 \text{ and } a_0 > \frac{k_1}{\ln \frac{c_0}{\tilde{b}_N}} \end{cases},$$

where $k_1 = \ln\left(-\frac{\ln\left(2^{-c_1} - 2^{-1}\right)}{\ln(2)}\right) \approx 0.455$.

The power-function of $\tilde{\varphi}_{a_0}^{\geq}$ is simulated for sample size $N = 100$, 20% censored data and different shape and scale parameters. Figure 3.33 shows the results in comparsion to the power-function of the test with known scale parameter.

Here too, it is hard to detect differences between the two power-functions. The estimation of the scale parameter seems to have no influence on the power-function of the test for the shape parameter.

Using the tests for $H_0 : a \geq a_0$ and $H_0 : a \leq a_0$, a test for $H_0 : a = a_0$ against $H_1 : a \neq a_0$ is defined as

$$\tilde{\varphi}_{a_0}^{=}(z_*) := \max(1_{\{\tilde{T}(a_0,z_*) < \Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{\tilde{T}(a,z_*) < \Phi^{-1}(\frac{\alpha}{2})\}}(z_*)),$$

with

$$a = \begin{cases} \tilde{k}_0 a_0, & c_2^{\frac{1}{a_0}}\tilde{b}_N < c_0 \\ \tilde{c}_\alpha^1(a_0), & c_2^{\frac{1}{a_0}}\tilde{b}_N \geq c_0 \text{ and } a_0 > \frac{k_1}{\ln \frac{c_0}{\tilde{b}_N}} \\ \tilde{c}_\alpha^2(a_0), & c_2^{\frac{1}{a_0}}\tilde{b}_N \geq c_0 \text{ and } a_0 < \frac{k_1}{\ln \frac{c_0}{\tilde{b}_N}} \end{cases}$$

and $\tilde{k}_0 \approx 1.835$, $k_1 = \ln\left(-\frac{\ln\left(2^{-c_1} - 2^{-1}\right)}{\ln(2)}\right) \approx 0.455$.

Hence, confidence intervals for the shape parameter of the Weibull distribution in type-I right-censored data with unknown scale parameter are given by

$$\{a_0 > 0; \tilde{\varphi}_{a_0}^{=}(z_*) = 0\}.$$

Once more, we compare the simulated power-functions for the tests with known and unknown scale parameter, see Figure 3.34, and once more the simulations indicate that the estimation of the scale parameter has no influence on the power.

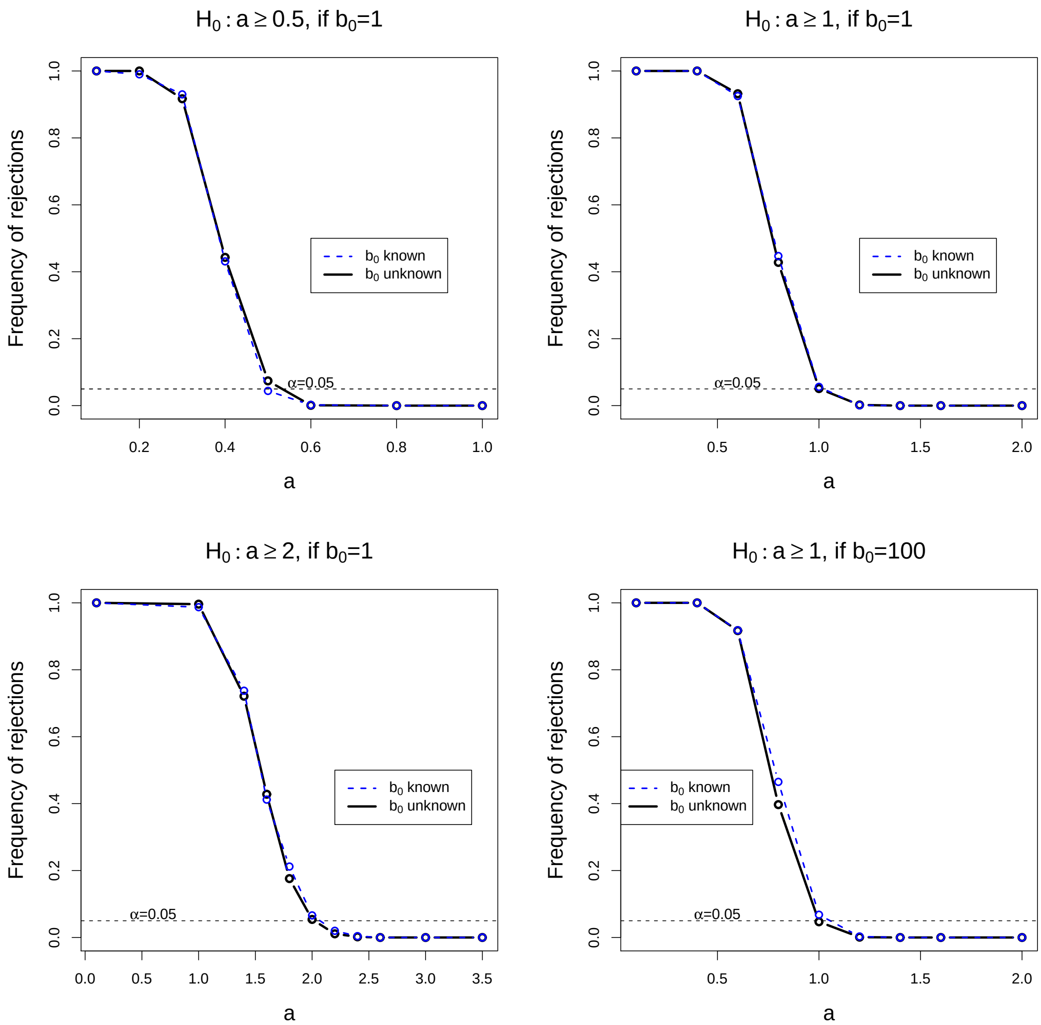

Figure 3.33.: Testing $H_0 : a \leq a_0$ for 20% right-censored data, compare unknown scale and known scale.

We compare 95%-confidence intervals (lik-depth) to the ones we get by the likelihood-ratio statistics (LRS) and the ones based on the method of medians (MoM) for the example of steel lifetimes.

Example 3.44. We introduce for the lifetime-data from Example 3.26, $y_* = (4030, 4680, 4860, 5750, 7170, 34100, 51000)$ two censorings, (a) $c_0 = 10000$ and (b) $c_0 = 7000$, then determine 95%-confidence intervals with the help of the new method based on likelihood-depth, the method based on MLE and the one based on MoM. The confidence intervals for the shape parameter we get by the methods from above are given in Table 3.15.

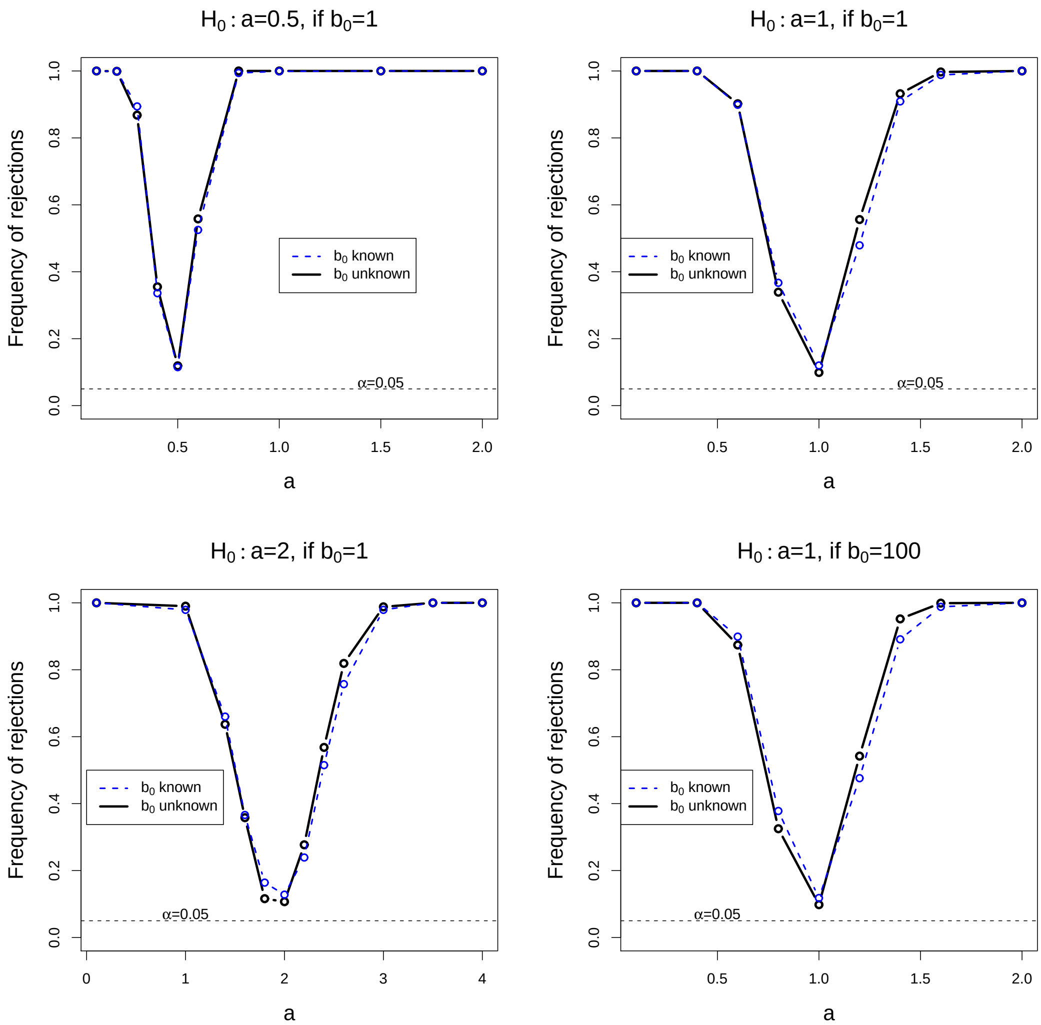

Figure 3.34.: Simulated power of the tests for $H_0 : a = a_0$ for 20% right-censored data, b_0 unknown, compared with the test when b_0 is supposed to be known.

Table 3.15.: 95%-confidence interval for the shape parameter for the steel data of Example 3.26.

c_0	MLE	MoM	lik-depth
∞	[0.63,2.33]	[1.01,6.00]	[0.46,4.38]
10000	[0.94,4.17]	[1.01,6.00]	[0.00,4.37]
7000	[1.20,7.14]	[2.04,12.06]	[0.00,4.37]

We see that only the confidence intervals based on likelihood-depth do not change for the heavier censoring, while the other two methods do.

3.4. Tests and confidence intervals for the scale parameter

In this section we derive tests for the scale parameter of the Weibull distribution. The results from Section 2.2 of the theory chapter and the considerations from Section 3.2 about estimators for the Weibull parameters are used.

3.4.1. Uncensored data with known shape parameter

Let be $T_* = (T_1, \ldots, T_N)$ as before i.i.d., $T_i \sim \text{Wei}(a_0, b)$. In the following we derive tests for the null hypothesis $H_0 : b \geq b_0$ against $H_1 : b < b_0$, $H_0 : b \leq b_0$ against the alternative $H_1 : b > b_0$, and match these two tests to a test for $H_0 : b = b_0$ against $H_1 : b \neq b_0$. We already showed in Section 3.2 about estimation that $s(b) = (\ln 2)^{\frac{1}{a_0}} b < b$ and $P_{a_0,b}(T^b_{pos}) < 0.5$. All through this subsection we consider the shape parameter to be known. For all tests we use the same test statistic, defined analog to Lemma 2.14, see page 17.

Definition 3.45. *For $t_* = (t_1, \ldots, t_N)$ consider*

$$T(b, t_*) := \sqrt{N} \frac{d_S(b, t_*) - 2p_{scale}(1 - p_{scale})}{2\sqrt{p_{scale}(1 - p_{scale})(1 - 2p_{scale})^2}},$$

where $p_{scale} = \exp(-1)$ as defined in the proof of Lemma 3.4 on page 39.

Using Corollary 2.17 on page 19 and Definition 2.19 on page 20, an asymptotic α-level test for $H_0 : b \geq b_0$ is easily defined.

Theorem 3.46. *We get an asymptotic α-level test for $H_0 : b \geq b_0$ by rejecting H_0, if $\sup_{b \geq b_0} T(b, t_*) < \Phi^{-1}(\alpha)$, i.e. we define*

$$\varphi^{\geq}_{b_0}(t_*) := 1_{\{\sup_{b \geq b_0} T(b,t_*) < \Phi^{-1}(\alpha)\}}(t_*).$$

Remark. *The test $\varphi^{\geq}_{b_0}$ is independent of the shape parameter. For this test we can also assume that the shape parameter is unknown.*

We compare the power of this new test in a simulation study with a test for the scale parameter given in the textbook of Rinne, [Rin 2009], based on the maximum likelihood estimator. The hypothesis $H_0 : b \geq b_0$ is rejected, if $\hat{b}_{MLE} < b_0 \exp\left(\frac{u_\alpha}{\hat{a}_{MLE}}\right)$, where $\hat{b}_{MLE}$ and $\hat{a}_{MLE}$ are the maximum likelihood estimates for the scale and shape parameter, respectively, see e.g. Section 3.2 in this work, and u_α are the percentage points of $U = \hat{a}_{MLE} \ln\left(\frac{\hat{b}_{MLE}}{b}\right)$, for a table see [Rin 2009]. For $N = 100$ and $\alpha = 0.05$ it is $u_\alpha = -0.174$. We simulate 1000 times 100 data with different scale parameter and test different hypotheses of type $H_0 : b \leq b_0$, see Figure 3.35.

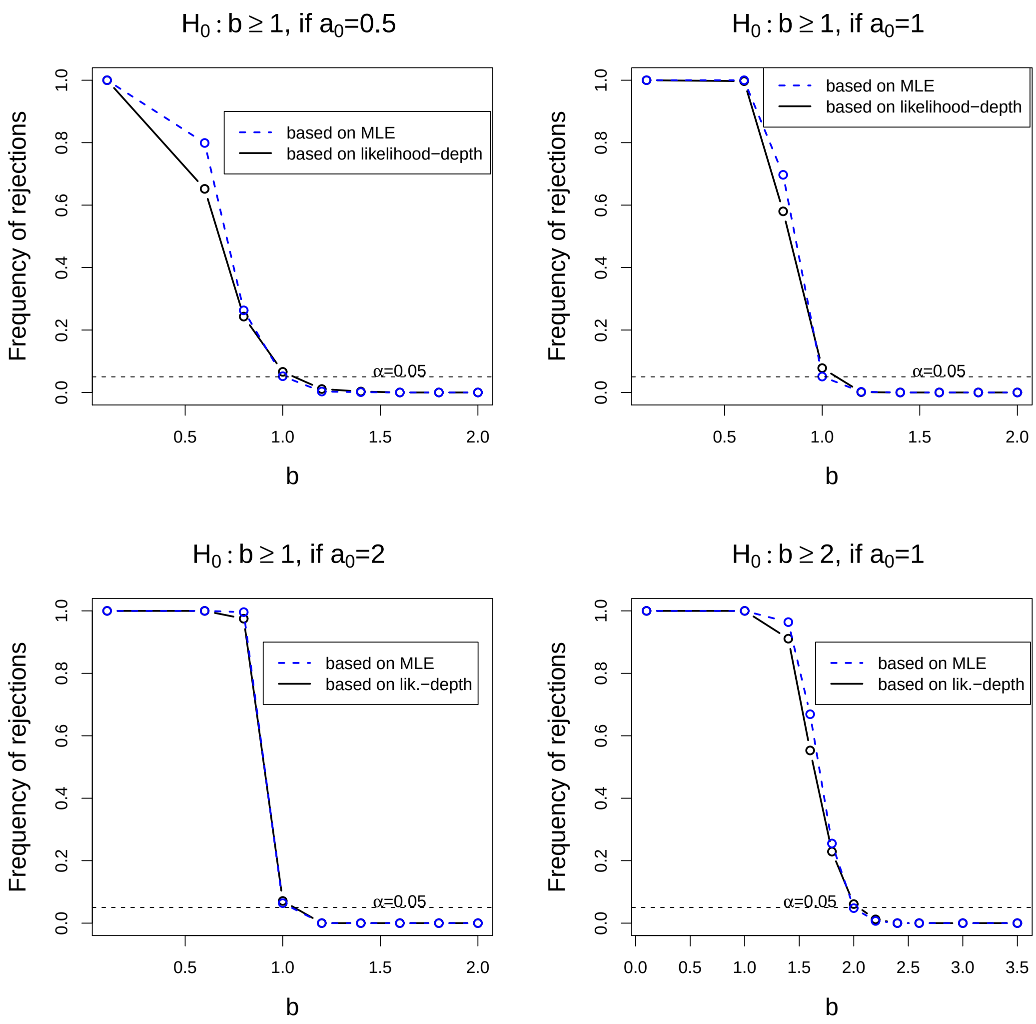

Figure 3.35.: Simulated power for $H_0 : b \geq b_0$.

We observe in Figure 3.35 that the shape parameter has a big influence on the power of the test. The test based on likelihood-depth and the test based on the MLE do not really differ in their power for uncontaminated data. The test based on the maximum likelihood estimator seems to give only slightly better results.

In a next step we consider data, for which some part is distributed with different scale and/or shape, i.e. ε-contaminated data. Here $\varepsilon = 0.1$ and the contamination distribution is Wei(a_1, b_1), see Figure 3.36. We suppose the shape parameter to be known as a_0 for the test based on MLE. For contamination with a small shape parameter the power of the test based on likelihood-depth seems to be better than the power of the test based on the MLE. For other contaminations we can not really find a difference in the behavior of the power-functions.

To receive also a good power of the test for $H_0 : b \leq b_0$ and prove consistency in both

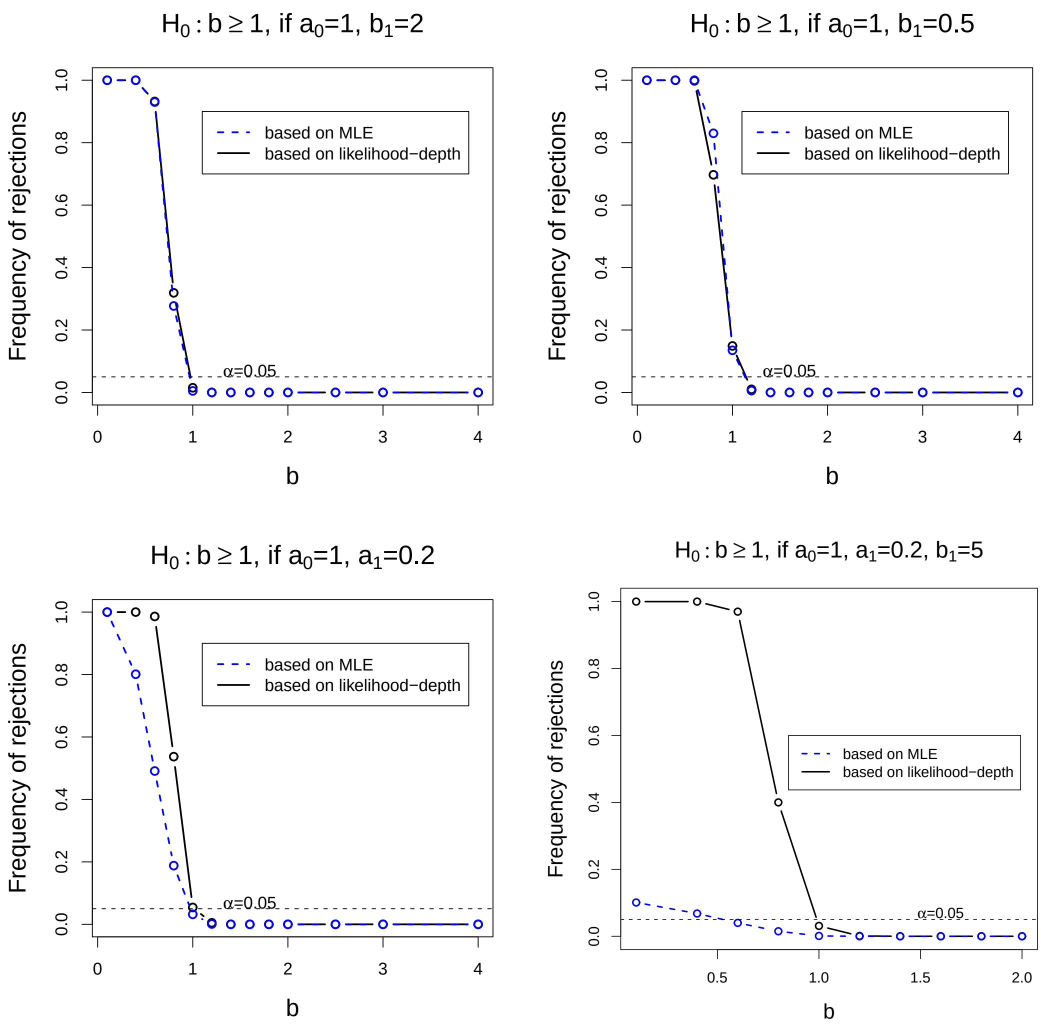

Figure 3.36.: Simulated power for $H_0 : b \geq b_0$ for contaminated data, i.e. 10% of the data is coming from $\text{Wei}(a_1, b_1)$ (if b_1 resp. a_1 not given $b_1 := b_0$ resp. $a_1 := a_0$).

cases $H_0 : b \geq b_0$ and $H_0 : b \leq b_0$, we have to determine

$$c^2_{\alpha,a_0}(b_0) := \min\{b; \lim_{N\to\infty} P_{a_0,b_0}\left(T(b,T_*) < \Phi^{-1}(\alpha)\right) \leq \alpha\},$$

as described in Definition 2.18, see page 19 and Definition 2.19 on page 20. Therefore we use Lemma 2.27 on page 25. Before we can do this we have to prove that the conditions of Lemma 2.27 are fulfilled.

Lemma 3.47. *It holds*

$$p^{a_0}_{b_0,b} := P_{a_0,b_0}(T^b_{pos}) = \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right).$$

Further, it is $p^{a_0}_{b_0,(\cdot)}$ *strictly decreasing,* $p^{a_0}_{(\cdot),b_0}$ *is strictly increasing, and* $\frac{1}{2} < 1 - p_{scale} = 1 - p^{a_0}_{b,b} = 1 - \exp(-1) \approx 0.623 < \frac{1}{2} + \frac{1}{\sqrt{8}}$.

Proof: At once, we get $p_{b_0,b}^{a_0} = P_{a_0,b_0}(T_{pos}^b) = P_{a_0,b_0}(T \geq b) = \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)$. Hence, for $b_1 < b_2$, i.e. $-\frac{b_1}{b_0} > -\frac{b_2}{b_0}$, it holds $p_{b_0,b_1}^{a_0} = \exp\left(-\left(\frac{b_1}{b_0}\right)^{a_0}\right) > \exp\left(-\left(\frac{b_2}{b_0}\right)^{a_0}\right) = p_{b_0,b_2}^{a_0}$. Thus, $p_{b_0,(\cdot)}^{a_0}$ is strictly decreasing for all $b_0 > 0$. The same arguments also yield that $p_{(\cdot),b_0}^{a_0}$ is strictly increasing. □

Application of Lemma 2.27 leads to

Lemma 3.48. *The correction for the power of the test for $H_0 : b \leq b_0$ is given by*

$$c_{\alpha,a_0}^2(b_0) = b_0(-\ln(1-\exp(-1)))^{\frac{1}{a_0}} \approx b_0(0.4587)^{\frac{1}{a_0}},$$

if $\alpha < 0.5$. Especially, it is $c_{\alpha,a_0}^2(\cdot)$ strictly increasing and as $-\ln(1-\exp(-1)) \approx 0.46 < 1$, it holds $c_\alpha^2(b_0) < b_0$ for all $b_0 > 0$.

Proof: Using Lemma 2.27 in connection with the last lemma we only have to solve $p_{b_0,b}^{a_0} = 1 - p_{scale}$ for b. Hence, solving

$$\exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right) = 1 - \exp(-1),$$

yields $b = (-\ln(1-\exp(-1)))^{\frac{1}{a_0}} b_0$. We end up with

$$c_{\alpha,a_0}^2(b_0) = (-\ln(1-\exp(-1)))^{\frac{1}{a_0}} b_0.$$

□

The improvement function c_{α,a_0}^2 depends on a_0, so here we really have to assume that a_0 is known. In the next section we also simulate the power of the test when a_0 is not known but has to be estimated.

Before defining the test for $H_0 : b \leq b_0$, we state the following theorem about the consistency of $\varphi_{b_0}^{\geq}$.

Theorem 3.49. *Let be $\alpha < 0.5$. Then $\varphi_{b_0}^{\geq}$ is a consistent test with asymptotic level α for $H_0 : b \geq b_0$.*

Proof: With Lemma 3.47, Lemma 3.48 and the fact that p_{scale} is continuous, the assumptions of Theorem 2.37 on page 31 are fulfilled, which yields the claim. □

Now we are able to give also a consistent test for $H_0 : b \leq b_0$.

Theorem 3.50. *Let be $\alpha < 0.5$. The test*

$$\varphi_{b_0}^{\leq}(t_*) := 1_{\{\sup_{b \leq b_0(-\ln(1-\exp(-1)))^{\frac{1}{a_0}}} T(b,t_*) < \Phi^{-1}(\alpha)\}}(t_*)$$

is consistent with asymptotic level α for the hypothesis $H_0 : b \leq b_0$ against the alternative $H_1 : b > b_0$.

Proof: Use analog arguments as in the proof of the last theorem and use Theorem 2.38 on page 31. □

We simulate the power of the new test and compare it to the test based on the MLE. Results are pictured in Figure 3.37 for uncontaminated data and in Figure 3.38 for contaminated data.

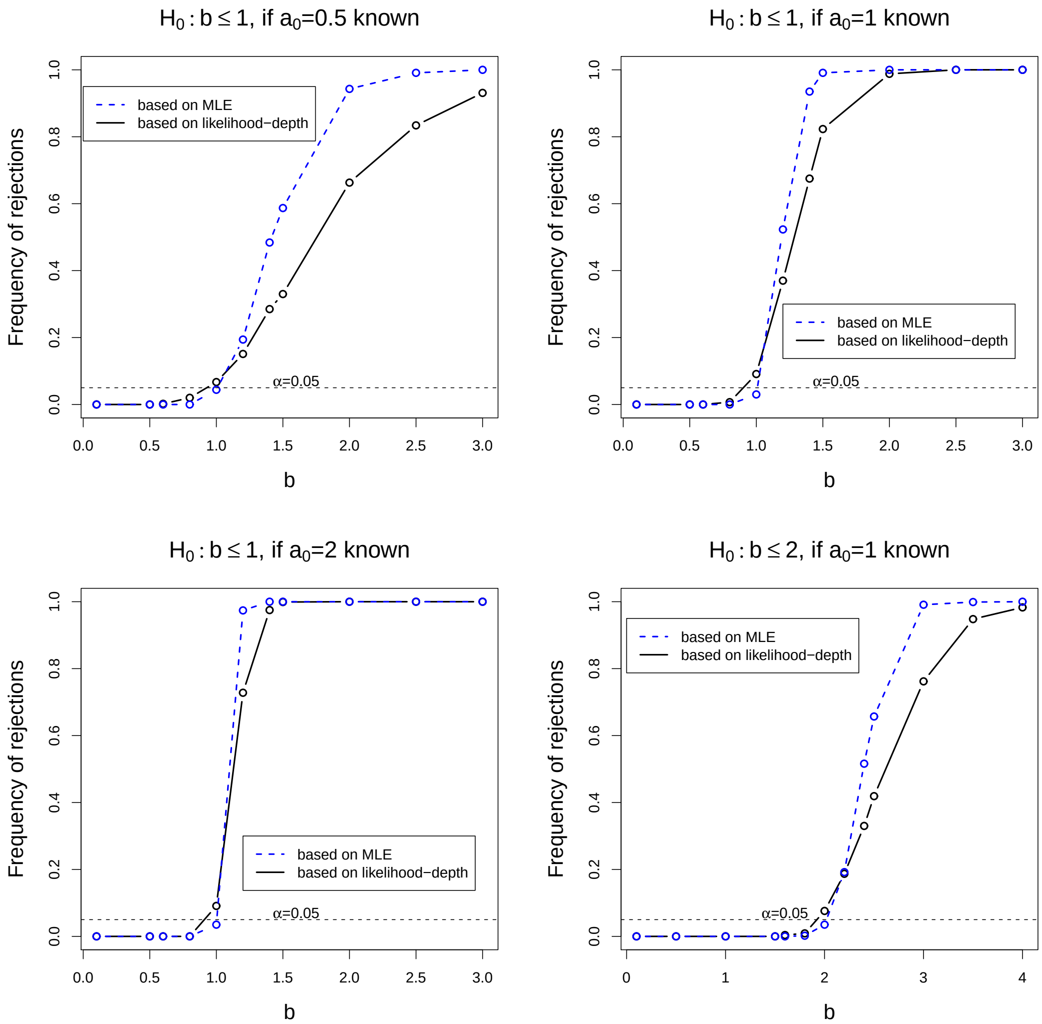

Figure 3.37.: Simulated power for $H_0 : b \leq b_0$, known shape parameter a_0.

We see that the power is depending on the shape parameter again. The power of the test based on MLE is better than the power of the new test for uncontaminated data. For contaminated data the new test is more robust, as expected. The test based on the maximum likelihood estimator is very sensitive to contamination with a small shape parameter in contrast to the new test, while it is not effected very much by contamination with a small scale.

With these two tests for $H_0 : b \geq b_0$ against $H_1 : b < b_0$ and $H_0 : b \leq b_0$ against $H_1 : b > b_0$

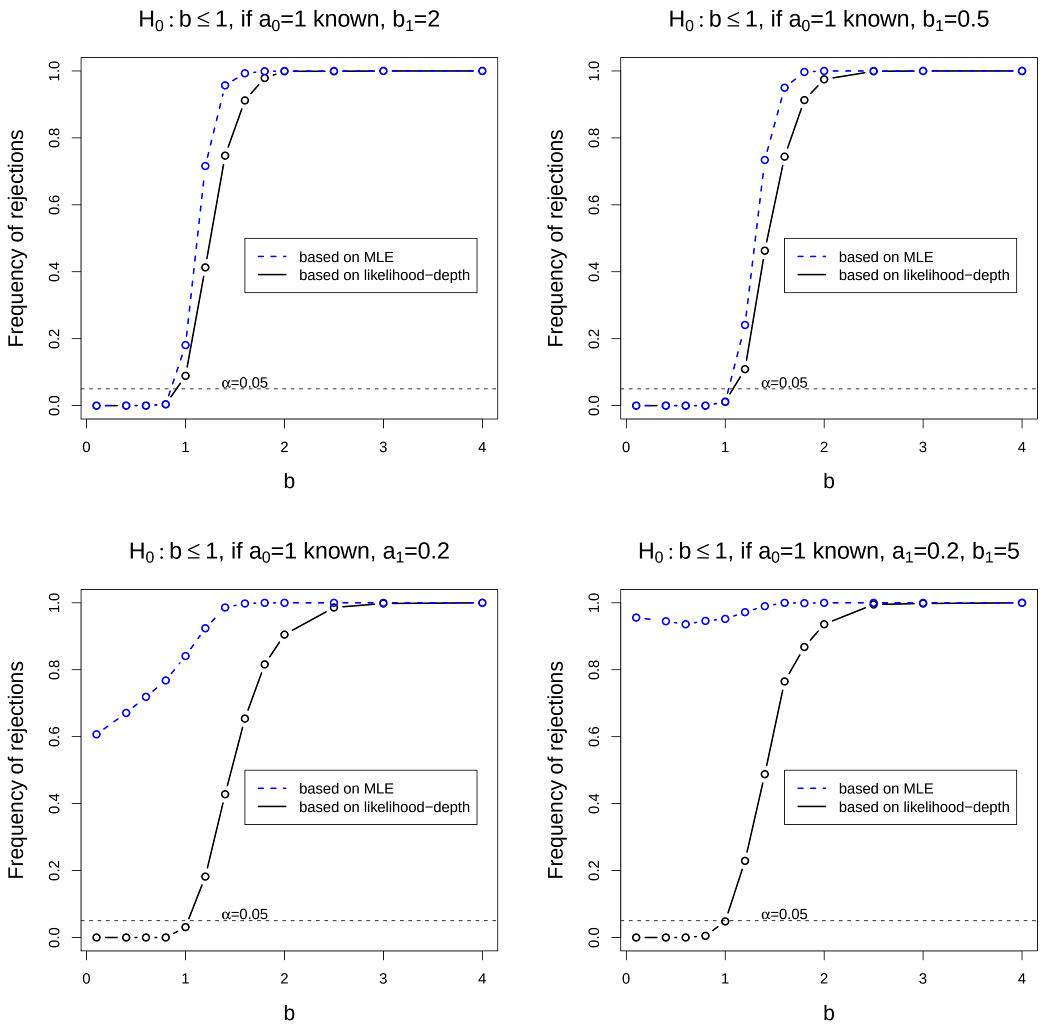

Figure 3.38.: Simulated power for $H_0 : b \le b_0$ with contaminated data, known shape parameter, contamination distribution Wei(a_1, b_1), (if b_1 resp. a_1 not given $b_1 := b_0$ resp. $a_1 := a_0$), ratio of contamination 10%.

we easily derive a consistent test for $H_0 : b = b_0$ against $H_1 : b \neq b_0$, see also Theorem 2.35 on page 30, and thereby we also get confidence intervals.

Theorem 3.51. *Let be $\alpha < 0.5$ and*

$$\varphi_{b_0}^{=}(t_*) := \max(1_{\{T(b,t_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(t_*), 1_{\{T(c^2_{\frac{\alpha}{2},a_0}(b),t_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(t_*)).$$

Then $\varphi_{b_0}^{=}$ is a consistent test with asymptotic level α for $H_0 : b = b_0$ against $H_0 : b \neq b_0$.

A confidence interval with asymptotic level $\gamma = 1 - \alpha$ for the scale parameter is given by

$$\{b_0 > 0; \varphi_{b_0}^{=}(t_*) = 0\}.$$

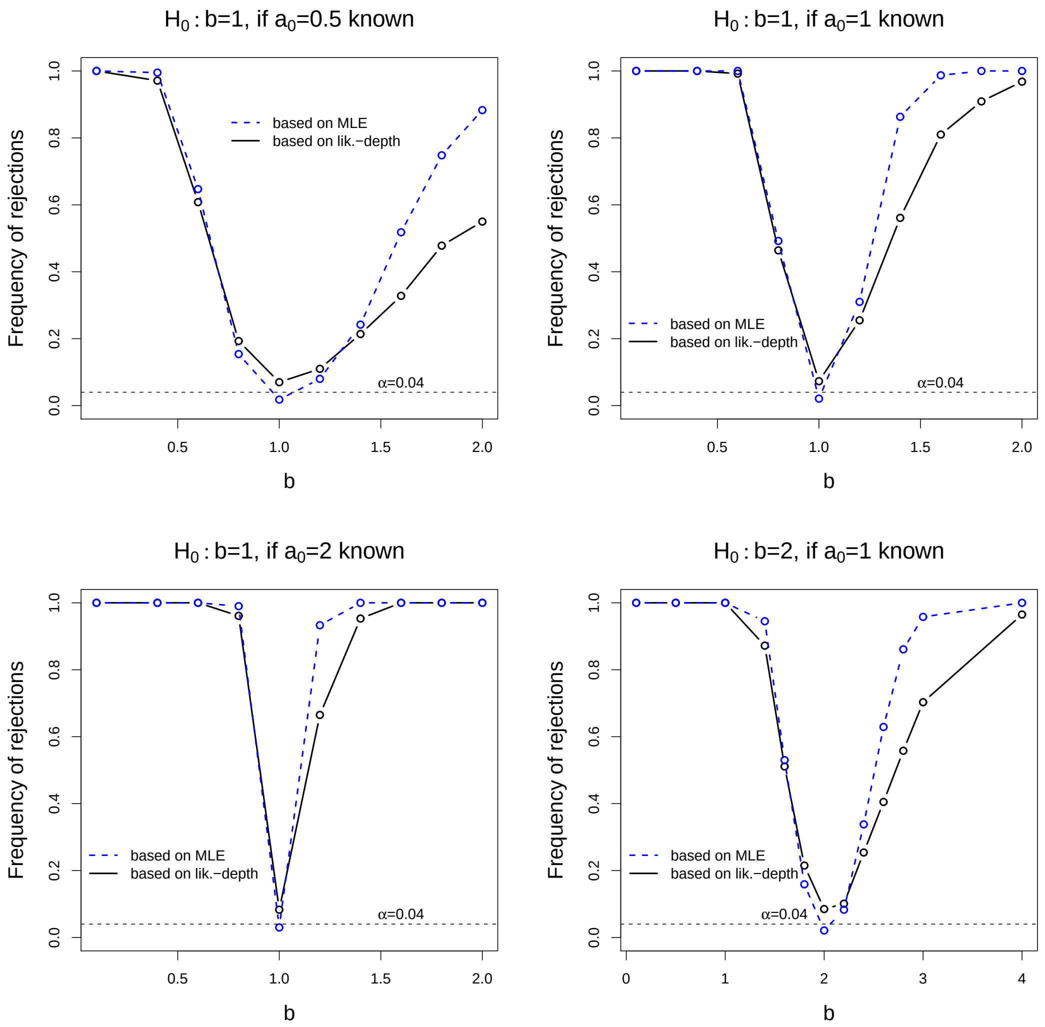

Figure 3.39.: Simulated power for $H_0 : b = b_0$, known shape parameter.

A simulation study shows that the new test for $H_0 : b = b_0$ based on likelihood-depth does not keep the level and that its power is worse than the power of the test based on the maximum likelihood estimator for $N = 100$ uncontaminated data, 1000 repetitions each, see Figure 3.39. We simulate tests with level $\alpha = 0.04$, as the tables for the values of u_α for the test based on the MLE are only available for $\alpha = 0.02$ (not $\alpha = 0.025$).

For contaminated data the simulation studies displayed in Figure 3.40 show that the new test is robust against ε-contamination in contrast to the test based on the MLE, where we choose $\varepsilon = 0.1$ in all studies. Also the confidence intervals for the scale parameter based on the method of likelihood-depth and confidence intervals based on the testing with the maximum likelihood estimator, as described before, are compared. We calculate 96%-confidence intervals for the same reasons that we considered tests with level $\alpha = 0.04$. The results for uncontaminated data are given in Table 3.16 and for contaminated data in Table 3.17. We simulate 1000 times 100 data each.

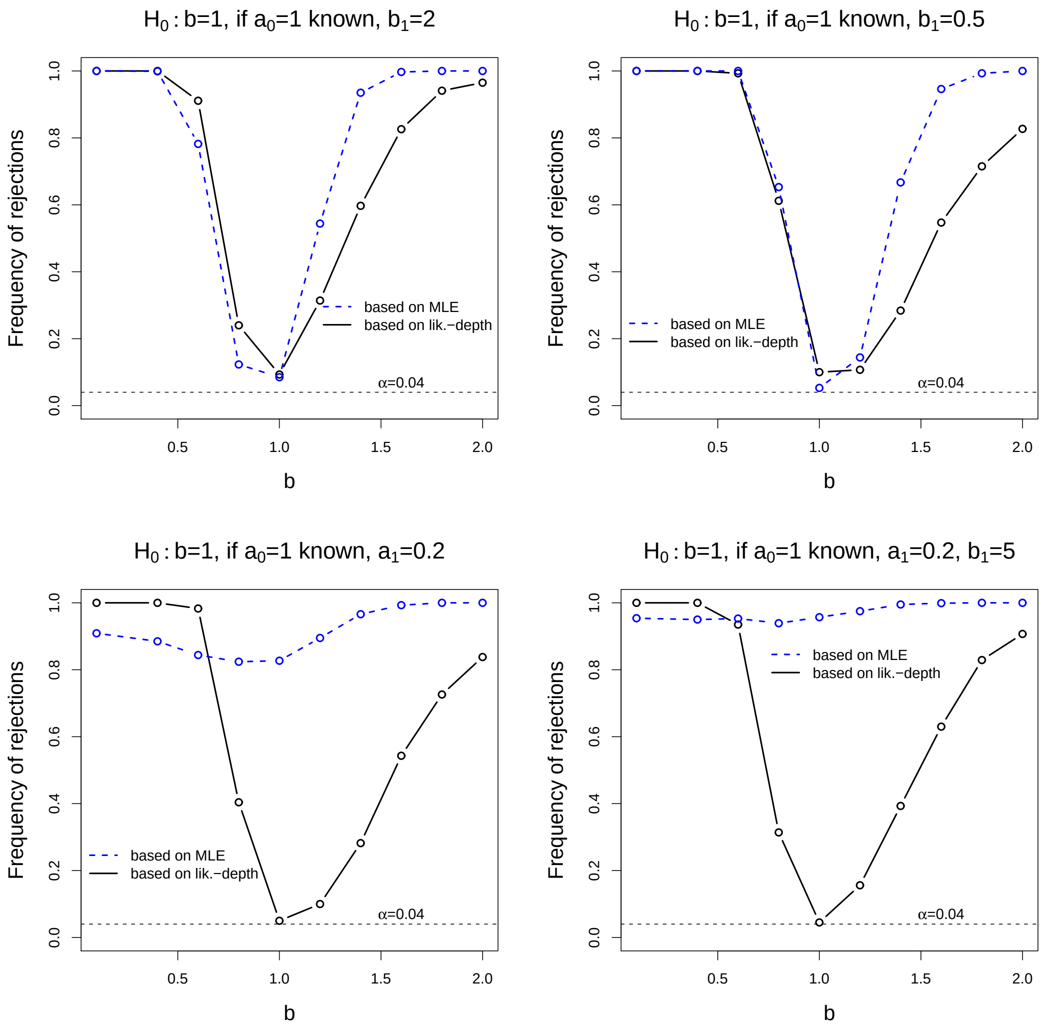

Figure 3.40.: Estimated power for $H_0 : b = b_0$ with contaminated data, known shape parameter, contamination distribution Wei(a_1, b_1) (if b_1 resp. a_1 not given $b_1 := b_0$ resp. $a_1 := a_0$).

Table 3.16.: 96%-confidence intervals for the scale parameter.

		MLE		likelihood-depth	
a	b	coverage	length	coverage	length
1	1	0.978	0.449	0.930	0.530
1	5	0.972	2.237	0.919	2.617
0.2	5	0.980	15.167	0.920	17.450
5	5	0.973	0.447	0.910	0.539
5	0.2	0.379	0.018	0.604	0.022
5	1	0.951	0.089	0.903	0.108
0.2	1	0.974	3.101	0.915	3.555

Table 3.17.: 96%-confidence intervals for the scale parameter for 0.1-contaminated data, with contamination distribution Wei(a_1, b_1).

				MLE		likelihood-depth	
a	b	a_1	b_1	coverage	length	coverage	length
1	1	0.5	1	0.796	0.496	0.941	0.592
1	1	5	1	0.976	0.445	0.805	0.360
1	1	1	10	0.008	0.847	0.857	0.683
5	5	1	10	0.005	1.120	0.952	0.650

We see that both methods give quite similar results. Most times the confidence intervals based on maximum likelihood estimation have a better coverage rate and smaller mean length. But in the case of $a = 5$ and $b = 0.2$, the coverage rate of the confidence intervals based on MLE is less than 40%. Here the confidence intervals based on likelihood-depth have a higher coverage rate at about 60%. If we consider contaminated data, the coverage rate of the confidence intervals based on the MLE goes down to less than one percent in some cases, while the new method is robust against contamination.

3.4.2. Uncensored data with unknown shape parameter

If the shape parameter is unknown, we can not use the theory of Section 2.2 anymore. As a_0 is unknown, we have to estimate it using Procedure 3.15 on page 46. We already mentioned that the test $\varphi_{b_0}^{\geq}$ for $H_0 : b \geq b_0$ against $H_1 : b < b_0$ is independent of a_0. Here we need not to do any extra work. For testing $H_0 : b \leq b_0$ we use $\varphi_{b_0}^{\leq}$. This test is depending on a_0, as the correction $c^2_{\alpha,a_0}(b_0)$ is, while the test statistic is independent of a_0. Considering a_0 unknown and estimated by $\hat{a}_{LDE}$, we get a new plug-in test by using $\hat{a}_{LDE} =: \hat{a}$, the estimator based on likelihood-depth instead of a_0. Then the correction becomes

$$\tilde{c}^2_{\alpha,\hat{a}}(b_0) = (-\ln(1 - \exp(-1)))^{\frac{1}{\hat{a}}} b_0.$$

We do some simulation studies to examine, if the estimation of the shape parameter has an influence on the power of the tests developed in the last subsection. The results can be found in Figure 3.41 and Figure 3.42. We compare the power of the test based on likelihood-depth with unknown shape to the test based on the MLE. For the definition of the latter test see Section 3.4.1, page 111. We set the sample size to $N = 100$ and repeat every simulation 1000 times.

The pictures show that the power does not really change, if the shape parameter is unknown and has to be estimated in case of uncontaminated data. We get the equal results as in the case, where it is supposed to be known, see Figure 3.37 and Figure 3.38. If we consider contaminated data, we see that the test based on the MLE becomes more

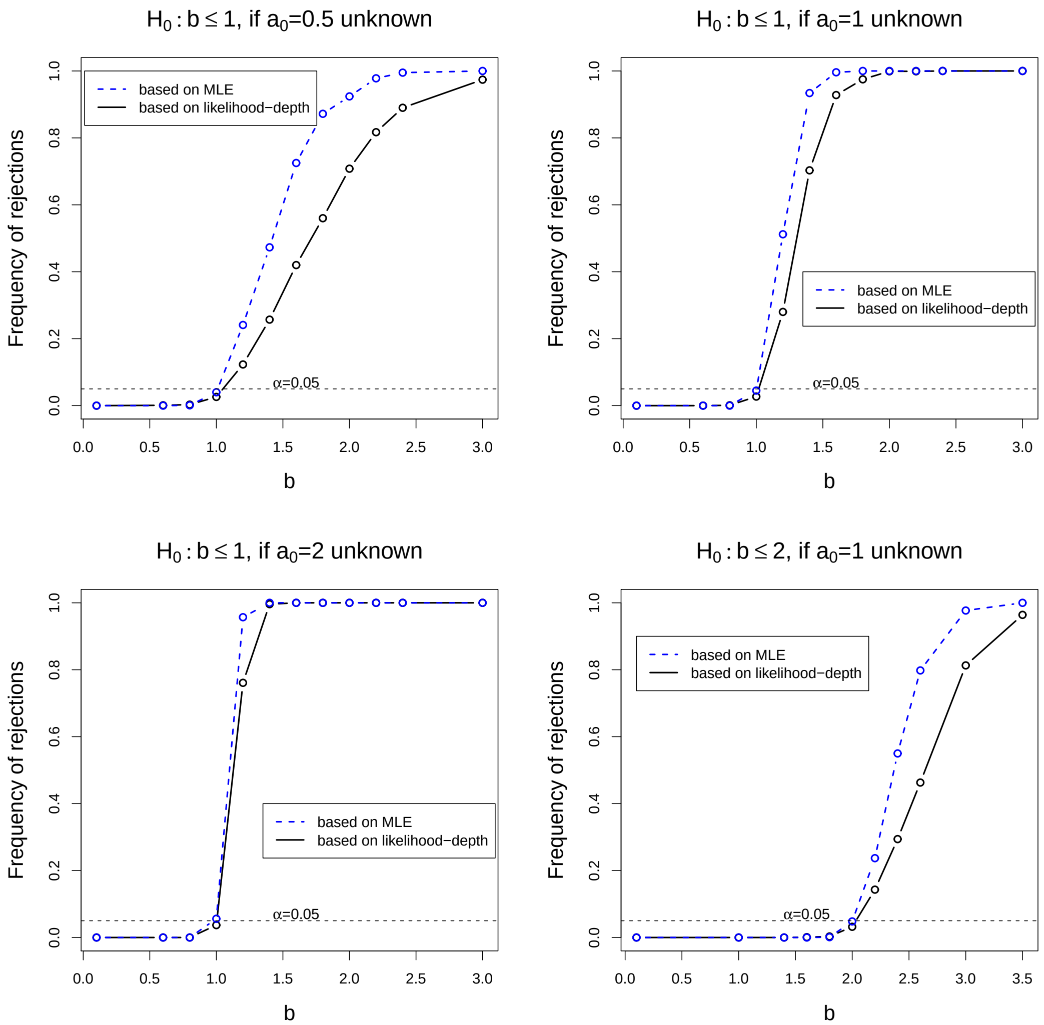

Figure 3.41.: Simulated power for $H_0 : b \leq b_0$, unknown shape parameter a_0.

robust, if the shape parameter is estimated by the MLE and not supposed to be known, see especially Figure 3.42 on the left and Figure 3.38, p. 116, in the second row on the left. The new test based on likelihood-depth is not influenced by the estimation at all.

Also we simulate the power of the test $\tilde{\varphi}^{=}_{b_0}$, when the shape parameter a_0 is unknown. Once more we compare this test to the test based on the MLE. The results for uncontaminated data are displayed in Figure 3.43 and for contaminated data in Figure 3.44. We see that for $a_0 = 2$ the power-functions of both tests are very close, while for $a_0 = 0.5$ the test based on the maximum likelihood estimator with estimated a_0 by the MLE is better. Again we see that for uncontaminated data the power is not really influenced by the estimation of the shape parameter. For contaminated data the test based on the MLE is more robust in some situations, when the shape is estimated. The test based on likelihood-depth is not disturbed by the estimation of the shape parameter.

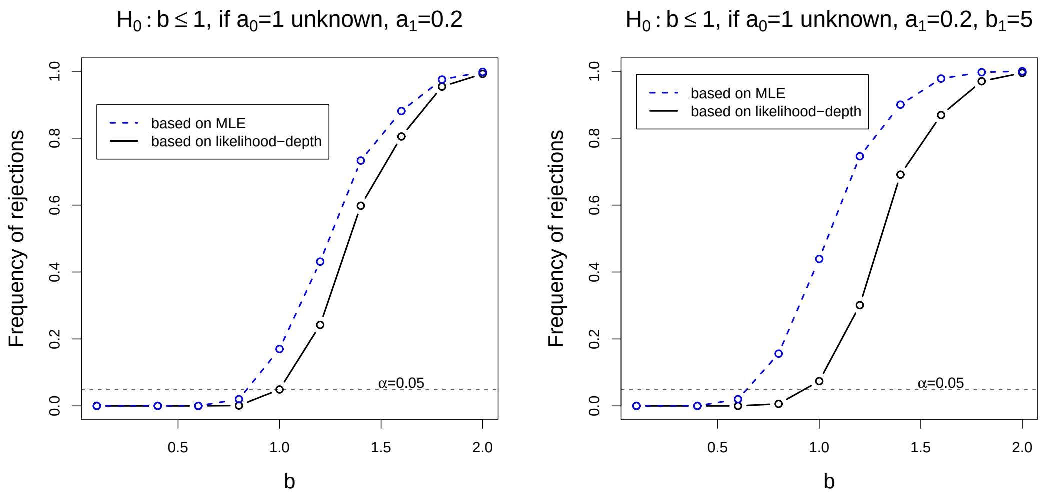

Figure 3.42.: Simulated power for $H_0 : b \leq b_0$ with contaminated data, unknown shape parameter a_0, contamination with Wei(a_1, b_1) samples, ratio of contamination 10% (if a_1 resp. b_1 not given, then $a_1 := a_0$ resp. $b_1 := b_0$).

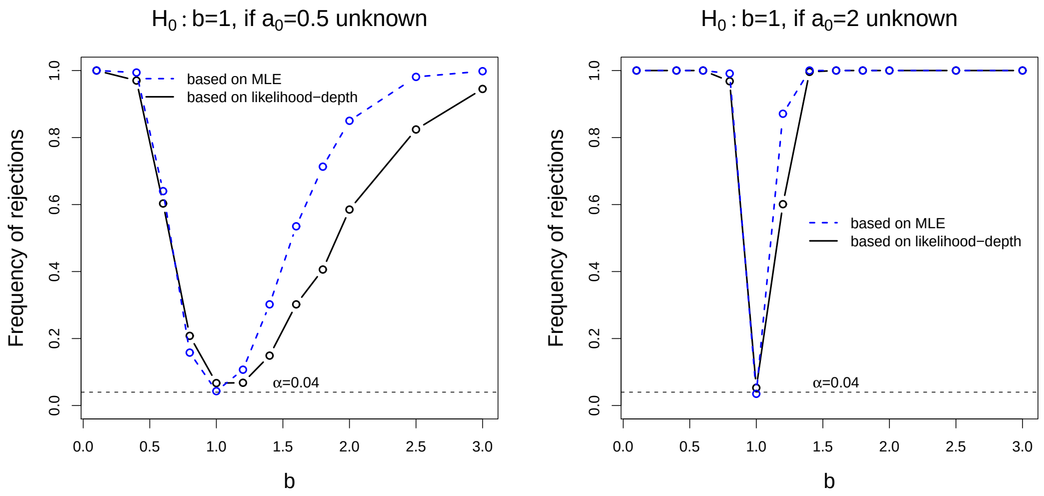

Figure 3.43.: Simulated power for $H_0 : b = b_0$, unknown shape parameter.

3.4.3. Type-I right-censored data with known shape parameter

Now consider type-I right-censored data and the shape parameter to be known. The number of uncensored data is denoted with k, according to (3.5) it holds $k > \frac{N}{2}$. In Theorem 3.17 on page 54 we proved that

$$T^b_{pos} = \begin{cases} [b\,, \infty), & b < c_0 \\ [c_0\,, \infty), & b \geq c_0 \end{cases} .$$

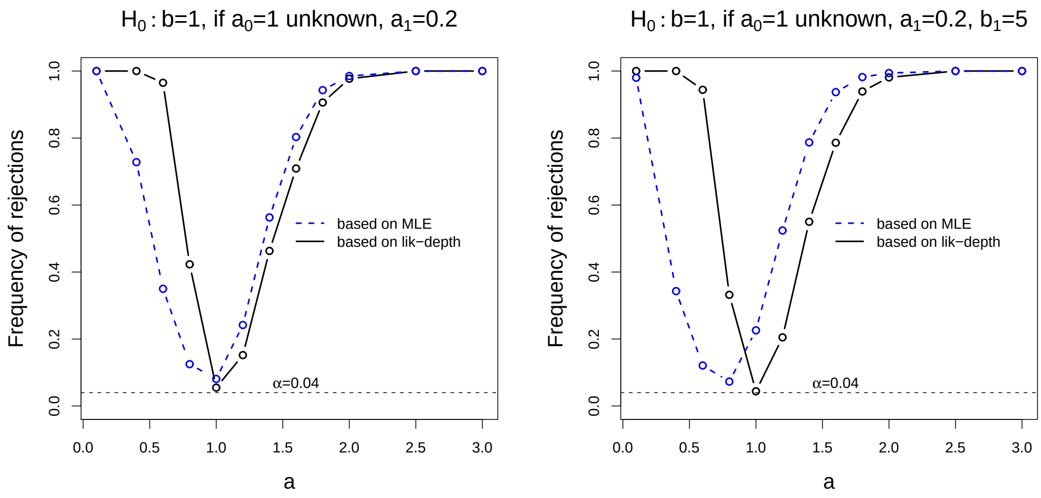

Figure 3.44.: Simulated power for $H_0 : b = b_0$ with contaminated data, unknown shape parameter a_0, contamination with Wei(a_1, b_1) samples, ratio of contamination 10% (if a_1 resp. b_1 not given, then $a_1 := a_0$ resp. $b_1 := b_0$).

Recall that $T^b_{pos} = \{t \in \mathbb{R}; \frac{\partial}{\partial b} \ln f_{a,b}(t) \geq 0\}$. Thus, it is $d_S(b, z_*) = N^b_{pos} \cdot N^b_{neg}$ with probability one, where $N^b_{pos} = \sharp\{z_n = (\delta_n, y_n); y_n > b\}$ and $N^b_{neg} = \sharp\{z_n; y_n < b\}$, if $c_0 < b$. If $c_0 \geq b$, it is $N_{pos} = N - k$ (number of censored data), $N_{neg} = k$ (number of uncensored data). Moreover, it holds

$$p_{b,c_0} := P_{a_0,b}(T^b_{pos}) = \begin{cases} \exp(-1) = p_{scale}, & b \leq c_0 \\ \exp\left(-\left(\frac{c_0}{b}\right)^{a_0}\right), & b > c_0 \end{cases}.$$

We define the test statistic analog to Lemma 2.14 on page 17:

Definition 3.52. $T(b, z_*) := \sqrt{N}\dfrac{d_S(b, z_*) - 2p_{b,c_0}(1 - p_{b,c_0})}{2\sqrt{p_{b,c_0}(1 - p_{b,c_0})(1 - 2p_{b,c_0})^2}}$.

If $b \leq c_0$, this is just the same as in the case of uncensored data, see page 111. If $b > c_0$, the simplicial likelihood-depth is constant, but p_{b,c_0} is growing with b up to $\exp(0) = 1$. Thus, the test statistic is growing to infinity. Hence, testing hypotheses for $b_0 > c_0$ does not make sense. Moreover, when testing $H_0 : b \geq b_0$ for $b_0 \leq c_0$, we only consider the supremum of the test statistics over $b \in \{b; b_0 \leq b \leq c_0\}$. As for $b \leq c_0$ the test statistic is the same as in the uncensored case and also

$$P_{a_0,b_0}(T^b_{pos}) = P_{a_0,b_0}(T \geq b) = \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right),$$

for $b \leq b_0 \leq c_0$, is the same as in the uncensored case, we can use the results from there, see page 114 to 116. This leads to the following

Theorem 3.53. *Let be $b_0 \leq c_0$ and $\alpha < 0.5$.*

(a) A test for $H_0 : b \geq b_0$ against $H_1 : b < b_0$ is given by

$$\varphi^{\geq}_{b_0}(z_*) := 1_{\{\sup_{b_0 \leq b \leq c_0} T(b,z_*) < \Phi^{-1}(\alpha)\}}(z_*).$$

This is a consistent test with asymptotic level α.

(b) To test the hypothesis $H_0 : b \leq b_0$ against $H_1 : b > b_0$ we use

$$\varphi^{\leq}_{b_0}(z_*) := 1_{\{\sup_{b_0 \leq c^2_{\alpha,a_0}(b_0)} T(b,z_*) < \Phi^{-1}(\alpha)\}}(z_*),$$

where $c^2_{\alpha,a_0}(b_0) = b_0(-\ln(1-\exp(-1)))^{\frac{1}{a_0}}$, see Lemma 3.48 on page 114. This test is consistent with asymptotic level α.

(c) For testing $H_0 : b = b_0$ against $H_1 : b \neq b_0$ we use

$$\varphi^{=}_{b_0}(z_*) := \max(1_{\{T(b_0,z_*) < \Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{T(c^2_{\frac{\alpha}{2},a_0}(b_0),z_*) < \Phi^{-1}(\frac{\alpha}{2})\}}(z_*)),$$

where $c^2_{\alpha,a_0}(b_0) = b_0(-\ln(1-\exp(-1)))^{\frac{1}{a_0}}$. This test is consistent with asymptotic level α.

Proof: As discussed right before the theorem, we use the same arguments as in the case of uncensored data, see Theorem 3.49 on page 114, Theorem 3.50 on page 114 and Theorem 3.51 on page 116. The claims follow immediately. □

The Figures 3.45, 3.46 and 3.47 show some simulation results for the estimation of the power-function for the different hypotheses. As a comparison we consider the Wald test, see e.g. the textbook of Lawless [Law 2003], Section 5.1.1. There it is shown that the test statistic

$$\frac{\hat{u}_{MLE} - u_0}{se(\hat{u}_{MLE})},$$

is under $H_0 : u = u_0$ approximately standard normal distributed. Here $u = \exp(b)$ and $se(\hat{u}_{MLE}) = (\hat{V}_{11})^{\frac{1}{2}}$, the square root of the upper left element of the inverse of the observed information matrix $\hat{V} = I^{-1}$,

$$\hat{V} = (I(\hat{u}_{MLE}, \hat{v}_{MLE}))^{-1} = \frac{1}{\hat{v}^2_{MLE}} \begin{pmatrix} r & \sum_{i=1}^N \hat{z}_i e^{\hat{z}_i} \\ \sum_{i=1}^N \hat{z}_i e^{\hat{z}_i} & r + \sum_{i=1}^N \hat{z}_i^2 e^{\hat{z}_i} \end{pmatrix}^{-1},$$

$\hat{z}_i = \frac{\ln(y_i) - \hat{u}_{MLE}}{\hat{v}_{MLE}}$, and $\hat{u}_{MLE} = \exp(\hat{b}_{MLE})$, $\hat{b}_{MLE}$ is the maximum likelihood estimator for the scale parameter b, $\hat{v}_{MLE} = \frac{1}{\hat{a}_{MLE}}$, $\hat{a}_{MLE}$ is the maximum likelihood estimator for a.

In every simulation the largest 20% of the data are censored, the sample size is $N = 100$ and every simualtion is repeated 1000 times. We see that the level is nearly kept and that the shape parameter has a great influence on the power of the test. A small shape parameter causes a worse power than a shape greater than one.

For $H_0 : b \geq b_0$ the power-functions of the new test and the Wald test do not really differ. For $H_0 : b \leq b_0$ and $H_0 : b = b_0$ the power of the Wald test is better.

We use the test for $H_0 : b = b_0$ to give confidence intervals for the scale parameter of the Weibull distribution.

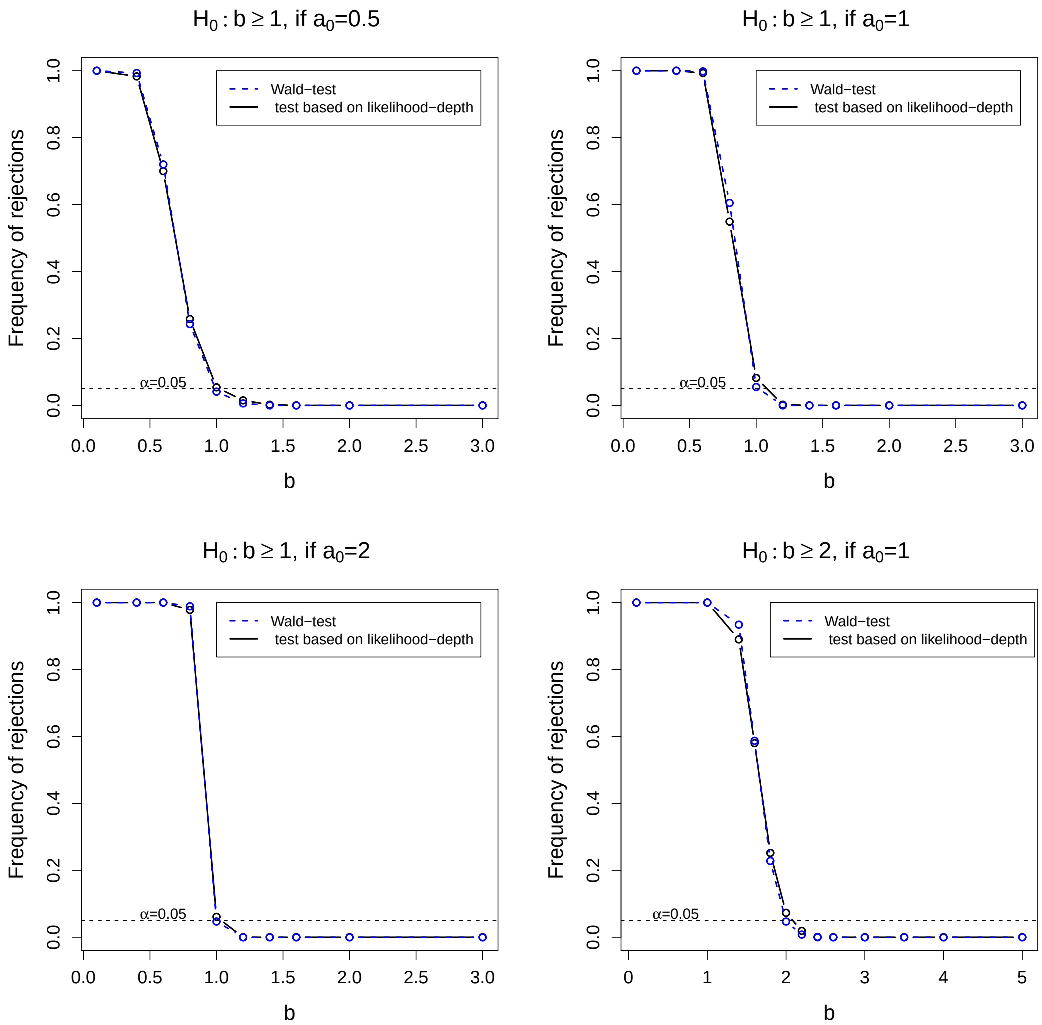

Figure 3.45.: Simulated power-function of the tests for $H_0 : b \geq b_0$ with 20% right-censored data.

Theorem 3.54. *A confidence interval for the scale parameter with asymptotic level* $\gamma = 1 - \alpha$ *is given by*

$$\{b_0 > 0; \varphi^{=}_{b_0}(z_*) = 0\},$$

where $\varphi^{=}_{b_0}$ *denotes the test for* $H_0 : b = b_0$*.*

We compare the confidence intervals based on likelihood-depth for censored data with confidence intervals given by Wald-type confidence procedures, see for example Lawless [Law 2003], Section 5.1.1. A confidence interval for the transformed scale parameter $u = \exp(b)$ based on the Wald-type methods with level γ is

$$[\hat{u}_{MLE} - \Phi^{-1}(\gamma)se(\hat{u}_{MLE}), \hat{u}_{MLE} + \Phi^{-1}(\gamma)se(\hat{u}_{MLE})],$$

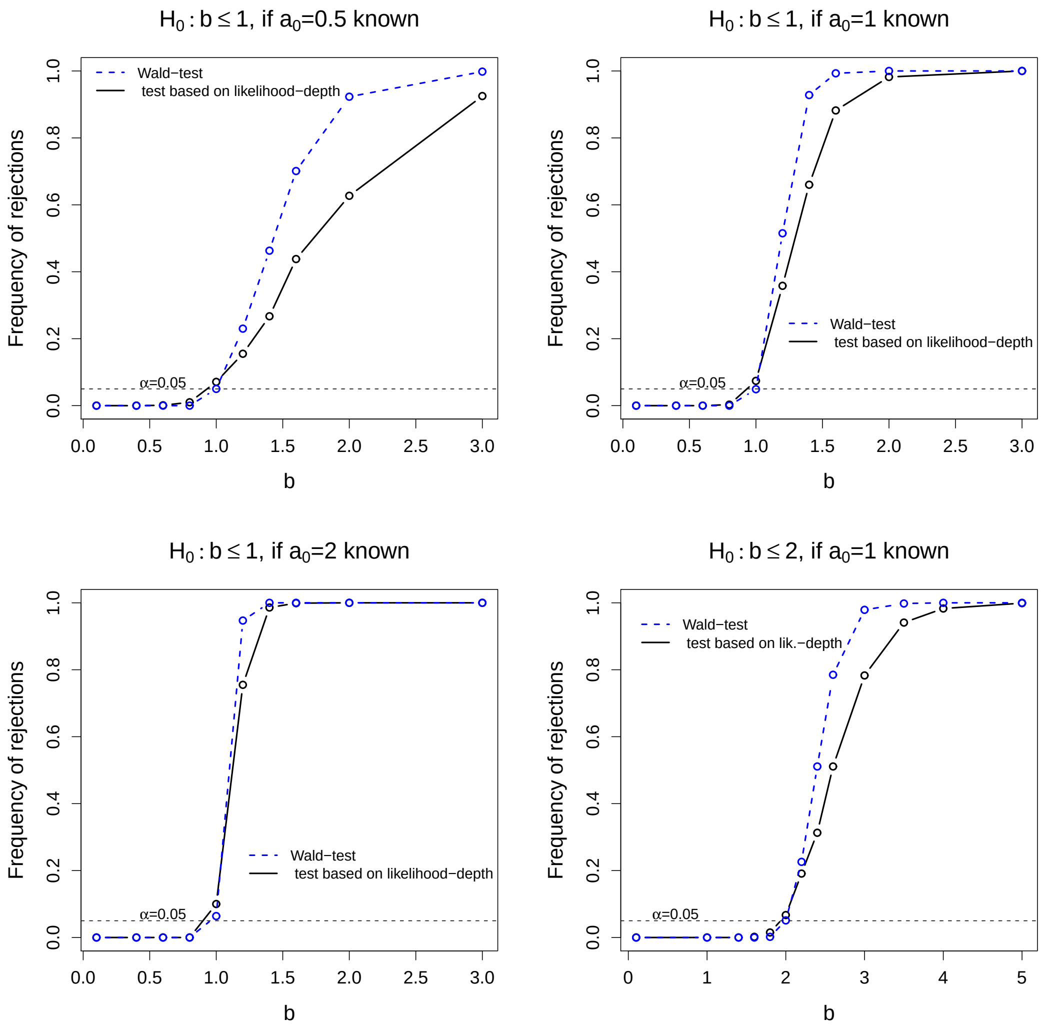

Figure 3.46.: Simulated power-function of the tests for $H_0 : b \leq b_0$ with 20% right-censored data.

where $\hat{u}_{MLE}$ denotes the MLE for u, $se(\hat{u}_{MLE}) = \hat{V}_{1,1}^{\frac{1}{2}}$. We simulate each 100 data with various scale and shape parameter and censored the largest 20% of it. Then we determine the confidence intervals with the methods described above. We repeat every simulation 1000 times. Table 3.18 shows the results.

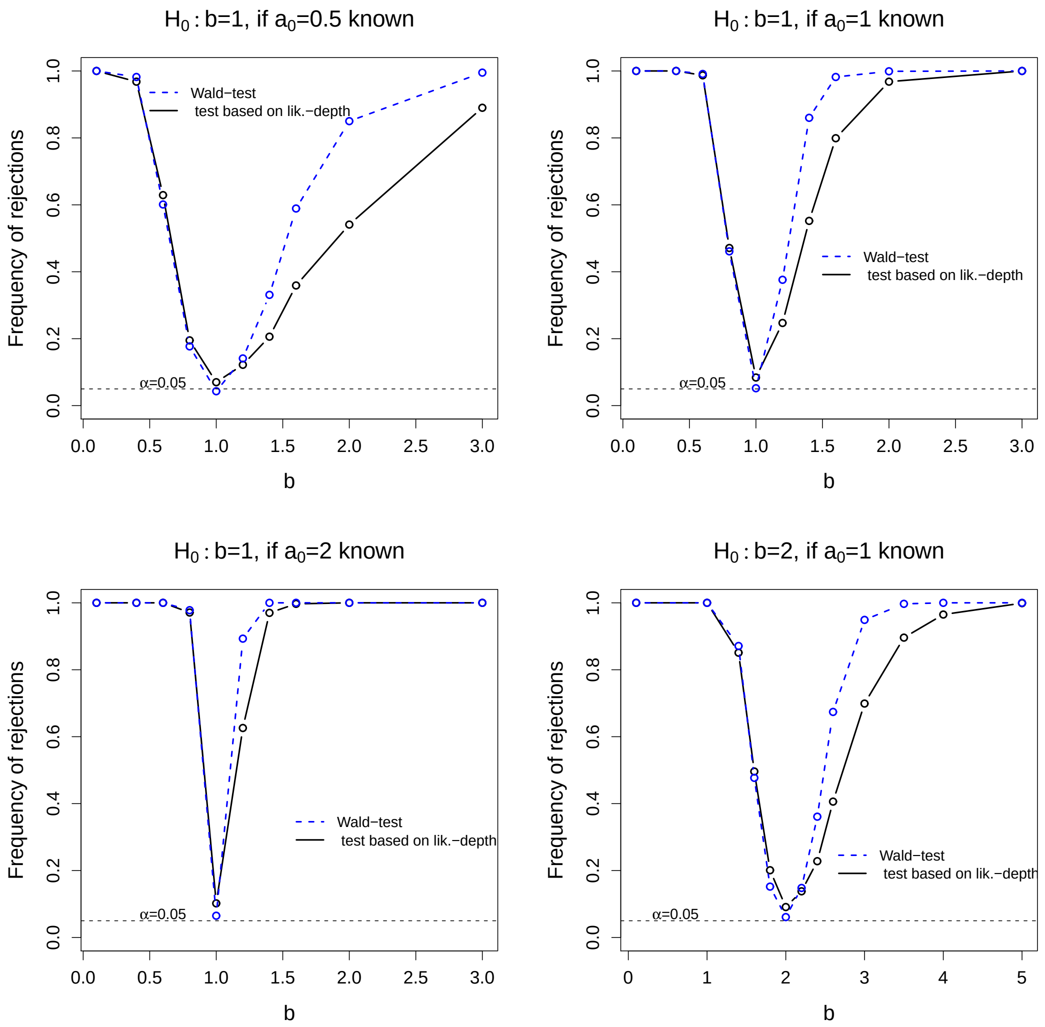

Figure 3.47.: Simulated power-function of the tests for $H_0 : b = b_0$ with 20% right-censored data.

Table 3.18.: 95%-confidence intervals for the scale parameter for 20% right-censored data.

		Wald-type		likelihood-depth	
a	b	coverage	length	coverage	length
1	1	0.951	0.452	0.912	0.522
0.2	1	0.943	3.348	0.907	3.535
0.2	1000	0.958	3274.662	0.932	2874.491
1	10	0.947	4.517	0.918	2.553
1	0.2	0.956	0.09	0.917	0.104
2	0.5	0.956	0.111	0.928	0.131

Table 3.18 shows that the confidence intervals based on the Wald test have higher covering rates than the ones based on the new test. We also consider some simulation of confidence intervals in ε-contaminated data, with contamination distribution Wei(a_1, b_1), $\varepsilon = 0.1$ and 20% right-censored data. The results for sample size 100 and 1000 repetitions each are displayed in Table 3.19.

Table 3.19.: 95%-confidence intervals for the scale parameter for 20% right-censored and ε-contaminated data, $\varepsilon = 0.1$.

				Wald-type		likelihood-depth	
a	b	a_1	b_1	coverage	length	coverage	length
1	1	1	0.1	0.8	0.4	0.844	0.614
1	1	1	10	0.586	0.543	0.844	0.687
0.2	1	0.2	10	0.926	3.996	0.923	4.161
0.2	1	0.2	0.1	0.941	2.653	0.904	2.843
1	1	0.2	0.1	0.903	0.429	0.915	0.617
1	1	0.2	5	0.911	0.471	0.956	0.639
1	5	0.2	0.5	0.909	2.134	0.918	3.086

In Table 3.19 we see that the covering rates of the confidence intervals based on the new test are more stable than the ones based on the Wald test, which seem still quite robust.

3.4.4. Type-I right-censored data with unknown shape parameter

Up to know, for the type-I right-censored data, we only considered tests in the situation of known shape parameter a_0. If it is unknown, we use an estimation based on likelihood-depth for it, see Procedure 3.23 on page 60, and plug it into the the correction, c^2_{α,a_0}, of the tests $\varphi^{\leq}_{b_0}$ and $\varphi^{=}_{b_0}$ instead of a_0. As in the uncensored case, the test statistic is independent of a_0. For this "new" tests we can not prove the consistency as before. But as in the Section 3.4.2 we will do some simulation studies to analyze the behavior of the power-function when the shape parameter is not known. We consider the test for $H_0 : b \leq b_0$ and $H_0 : b = b_0$, where we estimate the shape parameter with the help of Procedure 3.23. The results are displayed in Figure 3.48 and Figure 3.49. The sample size is $N = 100$, 20% of the data are right-censored and every simulation is repeated 100 times. Again the power-function is compared to the power-function of the Wald test, here we used for the estimation of the shape parameter the MLE.

If we compare the estimated power-functions of the test with known shape parameter in Figure 3.46 resp. Figure 3.47 to the power-functions of the test with estimated shape parameter in Figure 3.48 resp. Figure 3.49, the only difference we notice is that the level is kept by the latter ones in more cases. Thus, also for censored data, the estimation of the shape parameter seems not to (negatively) influence the power of the test.

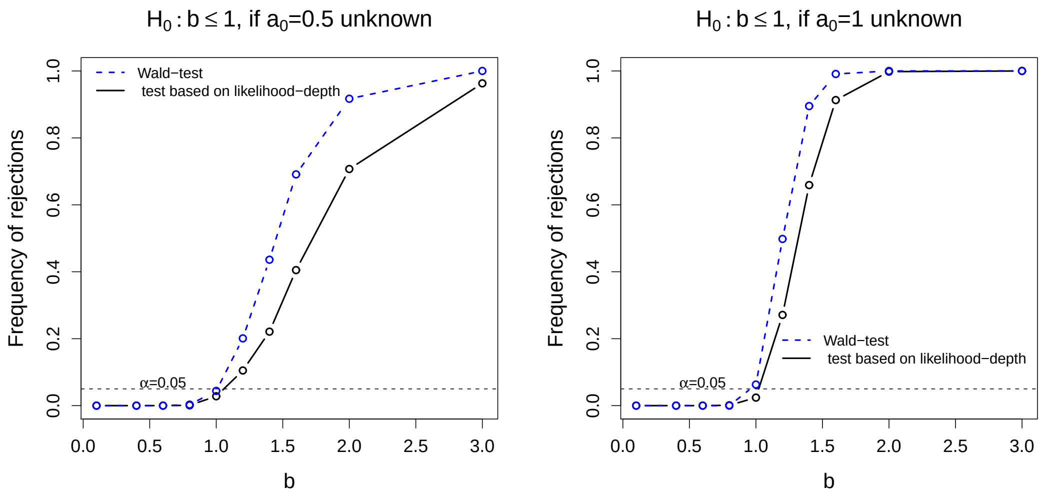

Figure 3.48.: Test for $H_0 : b \leq b_0$ for 20% right-censored data and unknown shape.

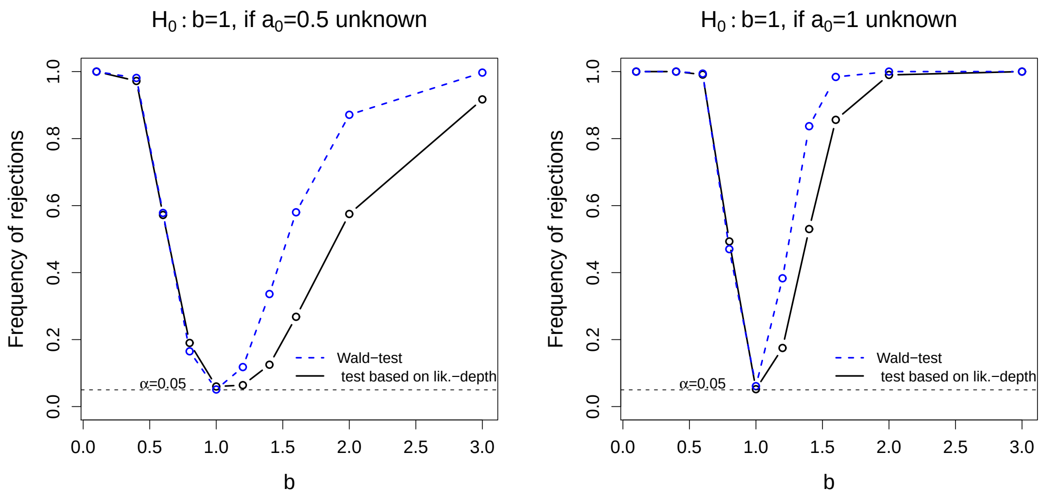

Figure 3.49.: Test for $H_0 : b = b_0$ for 20% right-censored data and unknown shape.

3.5. Open problems

In this chapter we developed estimation procedures and tests for the two parameters of the Weibull distribution in uncensored and censored data and we proved consistency of the estimators under the restriction, that less than half the data is censored. Also we proved that the tests for the shape parameter, respectively for the scale parameter, are consistent for uncensored data, if the scale, respectively the shape parameter is known. For censored data we also proved consistency of the tests for the scale parameter.

But in some situations we have not proven consistency in case of the shape parameter yet. This is the aim of ongoing analysis. Here we have to show the monotonicity of c_α^1

and c_α^2. Further studies could also be concerned with the analysis of the behavior of the power functions, when the parameter not tested is also unknown and has to be estimated first.

4. Copulas

4.1. Preliminaries

The copula model has a variety of applications, because it models dependence structures. For example in finance, especially in the analysis of credit risks the insolvency of several debtors at the same time, or for insurances, the risk of appearance of different claims at the same time, have to be modeled to insure solvency of the bank and insurance respectively all the time. Copulas are also used in the simulation of technical production processes to model e.g. the occurrence of coupled failures. We consider the two-dimensional case only. The following Definition 4.1 and the Theorems 4.2 and 4.3 are taken from the textbook [Nel 2006].

Definition 4.1. *[Nel 2006, Definition 2.2.2.] A function* $C : [0,1] \times [0,1] \to [0,1]$ *is called a (2-dimensional)* copula*, if it has the following properties:*

- *For all* $u, v \in [0,1] : C(u,0) = 0 = C(0,v)$ *and* $C(u,1) = u$, $C(1,v) = v$.
- *For all* $u_1, u_2, v_1, v_2 \in [0,1]$ *with* $u_1 \leq u_2$ *and* $v_1 \leq v_2$*:*

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Copulas are special distribution functions. There is an upper and a lower bound for every copula:

Theorem 4.2. *[Nel 2006, Theorem 2.2.3.] Let* C *be a* 2*-dimensional copula,* $u, v \in [0,1]$*, then we have:*

$$\max(u + v - 1, 0) =: W(u,v) \leq C(u,v) \leq M(u,v) := \min(u,v).$$

$W(u,v)$ *and* $M(u,v)$ *are both copulas (for dimension 2) and called the* Fréchet-Hoeffding lower bound *and* Fréchet-Hoeffding upper bound *respectively.*

A third important copula is the *product copula* $\Pi(u,v) = uv$, which models independence of U and V. The graphics in Figure 4.1 show the contour plot of the Fréchet-Hoeffding bounds and the independence copula.

One of the major properties of copulas is given in the theorem of Sklar. It is the foundation for many of the applications of the copula theory. The theorem yields that copulas couple the multivariate distribution function and the univariate margins.

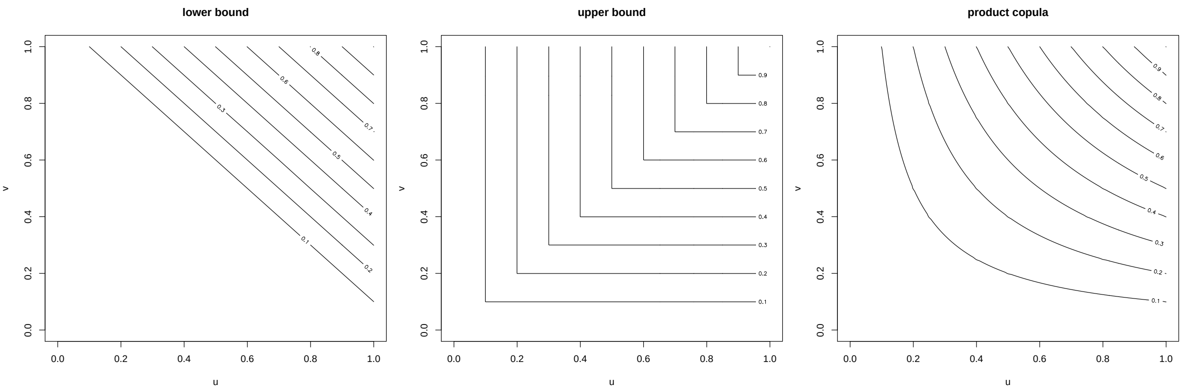

Figure 4.1.: Contour-plots of the Fréchet- Hoeffding bounds and the independence copula

Theorem 4.3 (Sklar)**.** *[Nel 2006, Theorem 2.3.3.] Let H be a 2-dimensional distribution function with margins F and G. Then there exists a copula C such that for all $x, y \in \bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$:*

$$H(x, y) = C(F(x), G(y)).$$

If F and G are continuous, C is unique. Then C can be determined with the help of the quasi-inverses of F and G, $F^{(-1)}$ and $G^{(-1)}$ respectively, for all $u, v \in [0, 1]$ as

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v)).$$

If F and G are not continuous, C is only unique on $range(F) \times range(G)$.

With a copula C and one-dimensional distribution functions F and G, we can define a two-dimensional distribution function with margins F and G for all $x, y \in \bar{\mathbb{R}}$ as

$$H(x, y) := C(F(x), G(y)).$$

With this theorem we are able to split a two-dimensional distribution function into the margins and the copula, which models the dependence structure between the two variables. On the other hand we can combine arbitrary margins and dependence structures and receive a two-dimensional distribution function.

In Section 4.2 and Section 4.3 we concern ourselves with a copula called the Gaussian copula :

Definition 4.4. *[Aas 2004] The* Gaussian copula *is defined as*

$$C_\rho(u, v) := \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left(\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right) ds\, dt,$$

$u, v \in (0, 1)$ and $-1 < \rho < 1$.

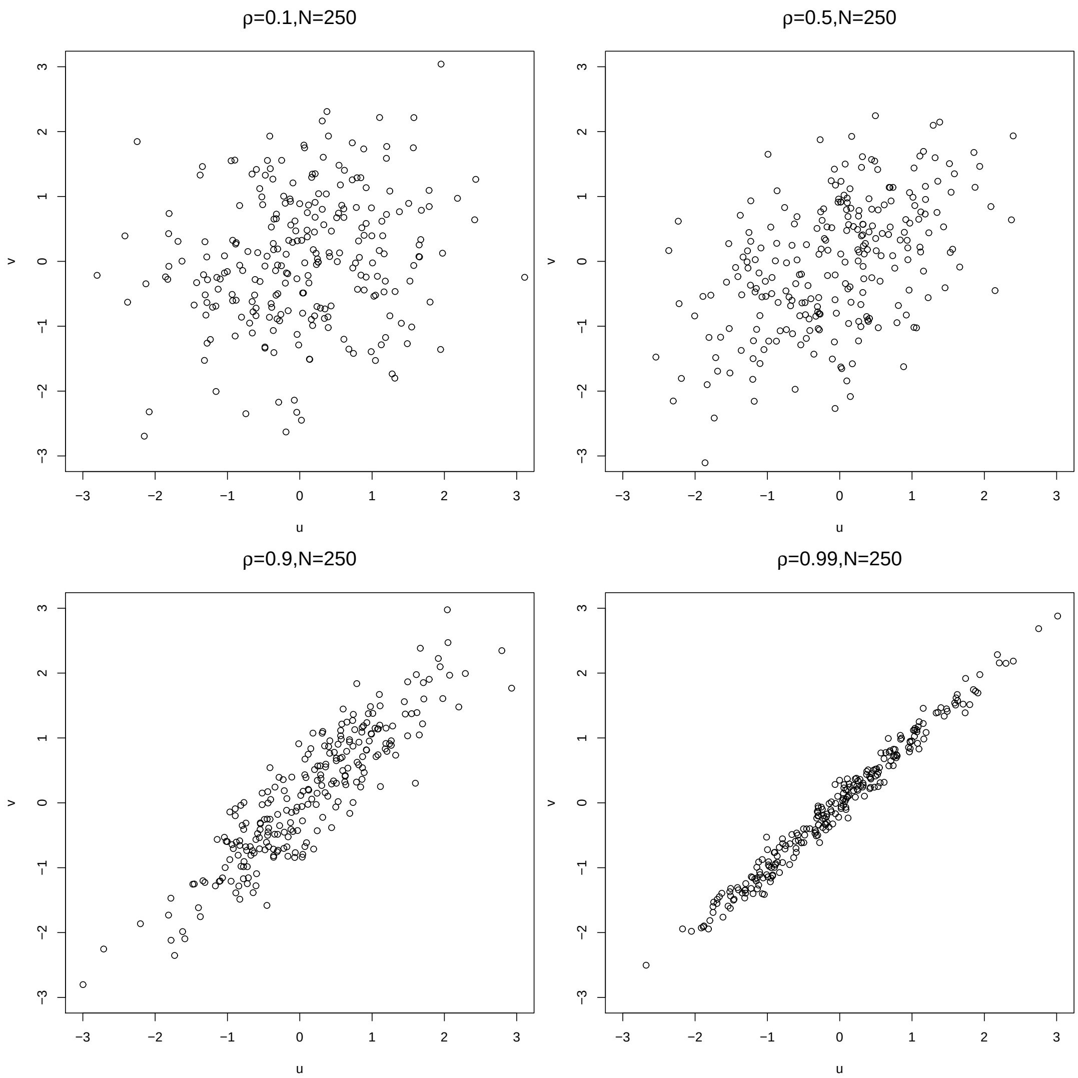

Figure 4.2.: Simulated data points for the Gauss distribution with varying ρ.

The parameter ρ is the correlation of U and V. We only consider the case $0 \leq \rho$, due to the symmetry. If (X, Y) has a two-dimensional normal distribution, X and Y having standard normal distribution, then the copula is the Gaussian copula. These variables X, Y shall be considered later. In Figure 4.2 data points with this distribution for various ρ are depicted. We used the R-package "copula", see [Yan 2007], for simulation.

Another special class of copulas are the Archimedean copulas, which are generated by convex functions:

Definition 4.5. *[Nel 2006] Let φ be a convex, continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(1) = 0$. Let $\varphi^{[-1]}$ denote the pseudo-inverse of φ, i.e.*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t \leq \infty \end{cases} .$$

Then the function

$$C : [0,1] \times [0,1] \to [0,1], \qquad (u,v) \mapsto C(u,v) := \varphi^{[-1]}(\varphi(u) + \varphi(v))$$

is a copula with generator φ. *These copulas are called* Archimedean copulas.

In the following we mainly look at an one-parametric family of Archimedean copulas, the *Gumbel* copulas:

Definition 4.6. *[Nel 2006, Table 4.1] Let* $\theta \geq 1$ *and* $\varphi_\theta : [0,1] \to [0,\infty]$, $t \mapsto \varphi_\theta(t) := (-\ln t)^\theta$. *Then the generated copula*

$$C_\theta : [0,1] \times [0,1] \to [0,1], \qquad (u,v) \mapsto C_\theta(u,v) := \exp\left(-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\right)$$

is called Gumbel copula.

Theorem 4.7. *For* $\theta = 1$ *the Gumbel copula is equal to the product copula* $C(u,v) = uv$, *which simulates independence. For* $\theta \to \infty$ *the limit of the Gumbel copula is the Fréchet-Hoeffding upper bound* $M(u,v)$, *which simulates total positive dependency.*

We proof this with the help of the following Lemma:

Lemma 4.8 (Theorem 4.4.8.). *[Nel 2006] Let* $\{C_\theta; \theta \in \Theta\}$ *be a family of Archimedean copulas with differentiable generators* φ_θ. *Then* $\lim C_\theta(u,v) = M(u,v)$, *if and only if*

$$\lim \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = 0 \text{ for all } t \in (0,1),$$

where "$\lim$*" denotes the appropriate one-sided limit as* θ *approaches an end point of the parameter interval* Θ.

Now we proof the above stated claims about the convergence of the copula.

Proof: (of Theorem 4.7) Clearly for $\theta = 1$ it follows

$$C_\theta(u,v) = \exp\left(-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\right) = \exp\left(-((-\ln u) + (-\ln v))\right) = u \cdot v.$$

From

$$\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{(-\ln t)^\theta}{-\frac{\theta}{t}(-\ln t)^{\theta-1}} = -\frac{(-\ln t)t}{\theta} = \frac{t \ln t}{\theta}$$

we get $\lim\limits_{\theta\to\infty} \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = 0$, so that the second theorem implies $\lim\limits_{\theta\to\infty} C_\theta(u,v) = M(u,v)$. □

As an example for data $z_i = (x_i, y_i)$, $i = 1, \ldots, N$, with joint distribution function given by the Gumbel copula (margins are the uniform distribution-function on $[0,1]$), the graphics in Figure 4.3 show simulated data ($N = 100$ data points) for this copula with varying parameter θ. Again we used the R-package "copula", see [Yan 2007], for the simulation.

For $\theta = 1.1$ the points are nearly independent distributed and for growing θ the convergence to $M(u,v)$ is reflected in the cumulation of the points along the line $x = y$. A special effect of the Gumbel copula for $\theta > 1$ is the concentration of points in $(1,1)$ and $(0,0)$.

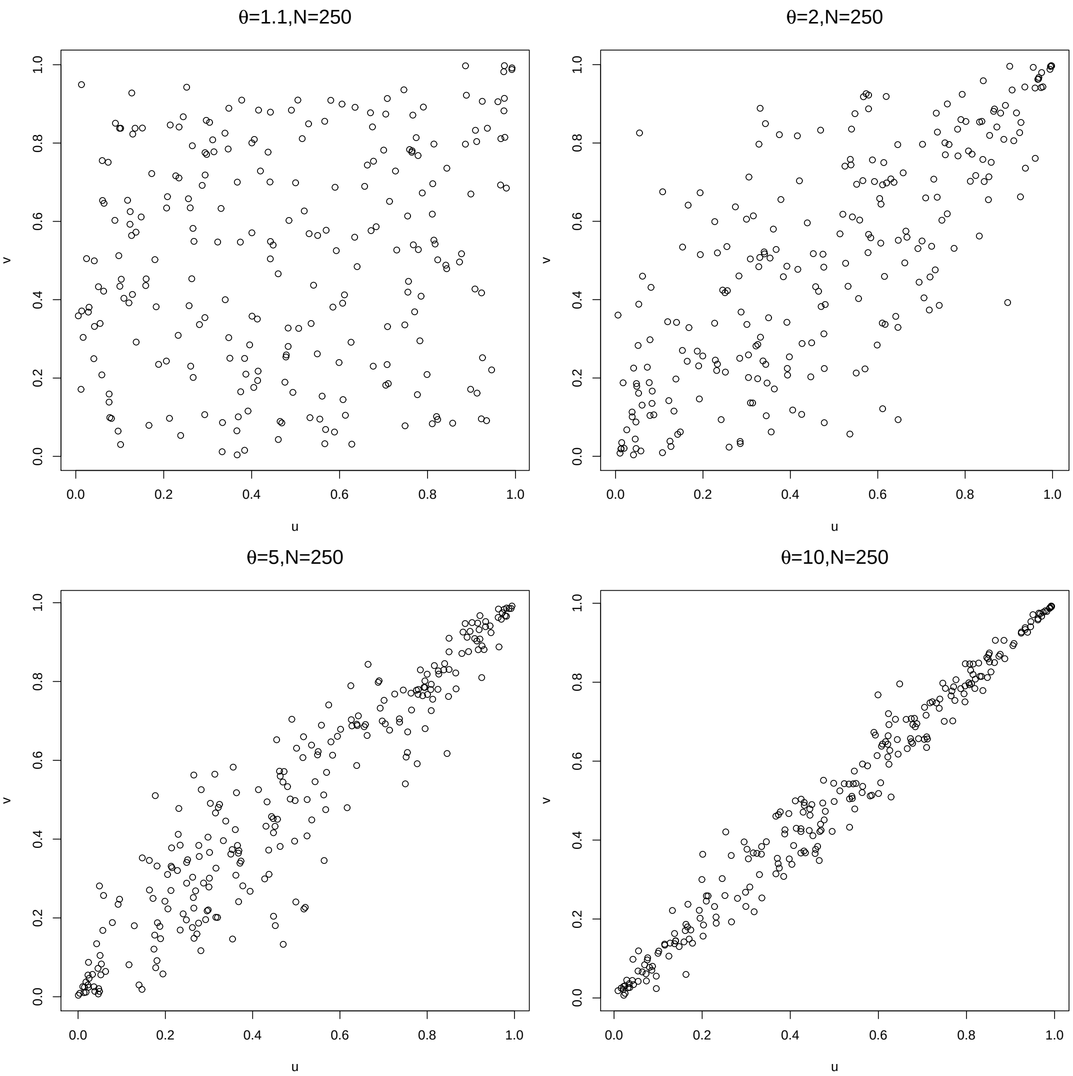

Figure 4.3.: Simulated data points for the Gumbel copula with varying θ.

4.2. Estimator for the correlation coefficient

First we take a look at the already mentioned Gaussian copula or more precisely at 2-dimensional i.i.d. random variables $Z_i = (X_i, Y_i), i = 1, \dots, N$, with standard normal variables X_i and Y_i and dependence structure given by the Gaussian copula, see Definition 4.4. We want to find an estimator for the parameter ρ based on likelihood-depth. As noted before, ρ is the correlation between X and Y.

To calculate the depth we make use of the function $h'(\rho, (x, y)) = \ln \frac{\partial}{\partial \rho} f_\rho(x, y)$, where f_ρ denotes the density of the two-dimensional normal distribution with mean $\mu = (0, 0)$ and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

Lemma 4.9. *It holds*

$$h'(\rho,(x,y)) = \frac{-\rho y^2 + (x+\rho^2 x)y + \rho - \rho^3 - \rho x^2}{(1-\rho^2)^2}.$$

Proof: Let be f_ρ the density function of the 2-dimensional normal distribution with mean and covariance matrix as given above. We derive

$$\ln(f_\rho(x,y)) = -\ln(2\pi) - \ln(\sqrt{1-\rho^2}) - \frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}$$

and

$$\begin{aligned} h'(\rho,(x,y)) &= \frac{\partial}{\partial\rho}\ln(f_\rho(x,y)) \\ &= \frac{\rho}{1-\rho^2} - \frac{(-2xy)2(1-\rho^2) - (x^2 - 2\rho xy + y^2)2(-2\rho)}{4(1-\rho^2)^2} \\ &= \frac{4\rho - 4\rho^3 + 4xy + 4\rho^2 xy - 4\rho x^2 - 4\rho y^2}{4(1-\rho^2)^2} \\ &= \frac{-\rho y^2 + (x+\rho^2 x)y + \rho - \rho^3 - \rho x^2}{(1-\rho^2)^2}. \end{aligned}$$

□

Figure 4.4 shows the mean of the parameter with maximum depth and the standard deviation for different underlying distributions with correlation coefficient ρ_0. For every ρ_0 we simulated 1000 times $N = 100$ data and calculated the parameter with maximum depth.

The parameter $\tilde{\rho}$ with maximum likelihood-depth is always greater than the real parameter ρ_0, which means the maximum depth estimator is biased. We will see later on that the only exception is $\rho_0 = 0$. As in Chapter 2, we use the following shortcuts

$$T^\rho_{pos} := \{z = (x,y); h'(\rho,z) \geq 0\},\ T^\rho_{neg} := \{z = (x,y); h'(\rho,z) \leq 0\}$$

and

$$p_{\rho,\rho'} := P_\rho(T^{\rho'}_{pos}) = \int\int 1_{T^{\rho'}_{pos}}(x,y) f_\rho(x,y)\,dx\,dy.$$

To determine $p_\rho := p_{\rho,\rho}$ for a fixed ρ, we need the boundaries of T^ρ_{pos}, which are given by the zeros of $h'(\rho,\cdot)$.

For $\rho = 0$ the probability that a data lies inside the region T^ρ_{pos} is $\frac{1}{2}$, because

$$h'(\rho,(x,y)) = h'(0,(x,y)) = \frac{-0\cdot y^2 + (x + 0^2\cdot x)y + 0 - 0^3 - 0\cdot x^2}{(1-0^2)^2} = xy.$$

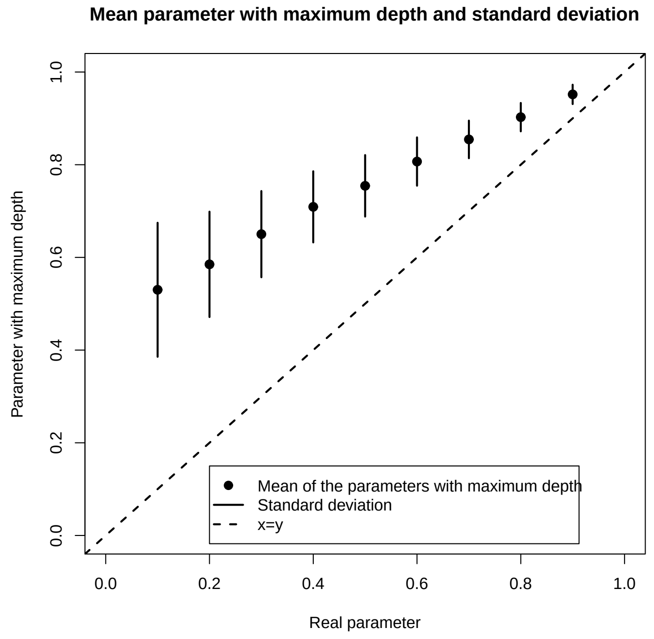

Figure 4.4.: Mean and standard deviation of the parameters with maximum depth for the 2-dimensional normal distribution.

This means that $h'(0,(x,y)) < 0$ if and only if x and y have different sign. Thus, the parameter with maximum depth is not an asymptotic biased estimator for $\rho = 0$. From now on let be $\rho > 0$. Then we find algebraic terms for the zeros of $h'(\rho,\cdot)$, i.e. we determine $v_{1/2}(x,\rho)$ such that

$$h'(\rho,(x,v_{1/2}(x,\rho))) = 0$$

for all $x \in \mathbb{R}$. Recall that $h'(\rho,\cdot) : \mathbb{R}^2 \to \mathbb{R}$.

Lemma 4.10. *The zeros of* $h'(\rho,(x,y)) = \frac{\partial}{\partial\rho} \ln f_\rho(x,y)$ *are*

$$v_1(x,\rho) = \frac{\rho^2 x + x + \sqrt{\rho^4 x^2 - 2\rho^2 x^2 + x^2 - 4\rho^4 + 4\rho^2}}{2\rho}$$

and

$$v_2(x,\rho) = \frac{\rho^2 x + x - \sqrt{\rho^4 x^2 - 2\rho^2 x^2 + x^2 - 4\rho^4 + 4\rho^2}}{2\rho}.$$

Furthermore we have $v_2(x,\rho) < x < v_1(x,\rho)$.

Proof: To determine the zeros, we have to solve a quadratic equation:

$$\frac{\partial}{\partial\rho}\ln f_\rho(x,y) = \frac{-\rho y^2+(x+\rho^2 x)y+\rho-\rho^3-\rho x^2}{(1-\rho^2)^2} = 0$$
$$\Leftrightarrow y^2 - \tfrac{x+\rho^2 x}{\rho}y - 1 + \rho^2 + x^2 = 0$$
$$\Leftrightarrow y_{1,2} = \tfrac{x+\rho^2 x}{2\rho} \pm \sqrt{\tfrac{x^2-2\rho^2x^2+\rho^4x^2+4\rho^2-4\rho^4}{4\rho^2}}.$$

That $v_2(x,\rho) < x < v_1(x,\rho)$ can be seen as follows, where especially the estimate against x is important. In a first step let be $x \geq 0$, then we obtain

$$\begin{aligned} v_1(x,\rho) &= \frac{\rho^2 x + x + \sqrt{\rho^4x^2-2\rho^2x^2+x^2-4\rho^4+4\rho^2}}{2\rho} \\ &= x\frac{1+\rho^2}{2\rho} + \underbrace{\frac{\sqrt{\rho^4x^2-2\rho^2x^2+x^2-4\rho^4+4\rho^2}}{2\rho}}_{>0} \\ &> x\underbrace{\frac{1+\rho^2}{2\rho}}_{>1} > x. \end{aligned}$$

For $x < 0$ it is

$$\begin{aligned} v_1(x,\rho) &= x\frac{1+\rho^2}{2\rho} + \frac{\sqrt{\rho^4x^2-2\rho^2x^2+x^2-4\rho^4+4\rho^2}}{2\rho} \\ &> x\frac{1+\rho^2}{2\rho} + \frac{-x(1+\rho^2)}{2\rho} = 0 > x. \end{aligned}$$

Now we show the second inequality $x < v_2(x,\rho)$:

$$\begin{aligned} v_2(x,\rho) &= \frac{\rho^2 x + x - \sqrt{\rho^4x^2-2\rho^2x^2+x^2-4\rho^4+4\rho^2}}{2\rho} \\ &= x\frac{1+\rho^2}{2\rho} - \frac{\sqrt{x^2(1-\rho^2)^2 \underbrace{-4\rho^2(\rho^2-1)}_{>0}}}{2\rho} \\ &< x\frac{1+\rho^2}{2\rho} - \frac{\sqrt{x^2(1-\rho^2)^2}}{2\rho} =: (\star), \end{aligned}$$

so $x < 0$ yields $(\star) = x\underbrace{\frac{1+\rho^2}{2\rho}}_{>1} - \underbrace{\frac{-x(1-\rho^2)}{2\rho}}_{>0} < x$. If $x \geq 0$ it is

$$(\star) = x\frac{1+\rho^2}{2\rho} - x\frac{1-\rho^2}{2\rho} = x\rho < x.$$

The proof is complete. □

The next lemma gives T_{pos}^{ρ} in terms of the zeros.

Lemma 4.11. *For $\rho > 0$ it is $T_{pos}^{\rho} = \{z = (x, y); v_2(x, \rho) \leq y \leq v_1(x, \rho)\}$.*

Proof: It is easily shown, that $h'(\rho, z)$ is positive for $z = (x, x)$ and fixed ρ:

$$\begin{aligned} h'(\rho,(x,x)) &= \frac{-\rho x^2 + (x+\rho^2 x)x + \rho - \rho^3 - \rho x^2}{(1-\rho^2)^2} \\ &= \frac{-2\rho x^2 + x^2 + \rho^2 x^2 + \rho(1-\rho^2)}{(1-\rho^2)^2} \\ &= \frac{(1-\rho)^2 x^2}{(1-\rho)^2(1+\rho)^2}\frac{\rho}{1-\rho^2} \\ &= \frac{x^2}{(1+\rho)^2}\frac{\rho}{1-\rho^2} > 0, \end{aligned}$$

since $0 < \rho < 1$. Because of this and due to the fact that $h'(\rho, \cdot)$ is continuous in (x, y) for fixed ρ, we have $T_{\rho}^{pos} = \{z_n = (x, y); v_2(x, \rho) \leq y \leq v_1(x, \rho)\}$. □

This consideration leads to the following

Lemma 4.12. *We have*

$$p_\rho = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \left(\Phi\left(\frac{v_1(x,\rho) - \rho x}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{v_2(x,\rho) - \rho x}{\sqrt{1-\rho^2}}\right)\right) dx,$$

where Φ denotes the one-dimensional standard normal distribution function.

Proof:

$$\begin{aligned} p_\rho = P_\rho(T_{pos}^{\rho}) &= \int_{-\infty}^{\infty}\int_{v_2(x,\rho)}^{v_1(x,\rho)} f_\rho(x,y)dydx \\ &= \int_{-\infty}^{\infty}\int_{v_2(x,\rho)}^{v_1(x,\rho)} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}} dydx \\ &= \int_{-\infty}^{\infty}\int_{v_2(x,\rho)}^{v_1(x,\rho)} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2(1-\rho^2+\rho^2) - 2\rho xy + y^2}{2(1-\rho^2)}} dydx \\ &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{v_2(x,\rho)}^{v_1(x,\rho)} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dydx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \left(\Phi\left(\frac{v_1(x,\rho) - \rho x}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{v_2(x,\rho) - \rho x}{\sqrt{1-\rho^2}}\right)\right) dx. \quad \square \end{aligned}$$

The last integral could not be determined analytically, so it was evaluated numerically for values in $[0, 1]$. Some of the results are shown in Table 4.1, to give an insight of its behavior.

Table 4.1.: p_ρ for the correlation coefficient.

p_ρ	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.	0.5000	0.5150	0.5256	0.5346	0.5425	0.5496	0.5561	0.5621	0.5677	0.5729
0.1	0.5777	0.5823	0.5866	0.5907	0.5945	0.5982	0.6017	0.6050	0.6082	0.6112
0.2	0.6141	0.6168	0.6195	0.6220	0.6244	0.6267	0.6290	0.6311	0.6332	0.6352
0.3	0.6371	0.6389	0.6407	0.6424	0.6441	0.6457	0.6472	0.6487	0.6501	0.6515
0.4	0.6528	0.6541	0.6553	0.6565	0.6576	0.6587	0.6598	0.6609	0.6619	0.6628
0.5	0.6638	0.6647	0.6655	0.6664	0.6672	0.6679	0.6687	0.6694	0.6701	0.6708
0.6	0.6714	0.6721	0.6727	0.6733	0.6738	0.6744	0.6749	0.6754	0.6758	0.6763
0.7	0.6767	0.6772	0.6775	0.6779	0.6783	0.6787	0.6790	0.6793	0.6796	0.6799
0.8	0.6802	0.6804	0.6806	0.6809	0.6811	0.6813	0.6815	0.6816	0.6818	0.6819
0.9	0.6821	0.6822	0.6823	0.6824	0.6825	0.6825	0.6826	0.6826	0.6827	0.6827

Here we see again that the maximum likelihood-depth estimator is biased. The probability that a data, coming from the distribution with ρ_0, lies inside $T_{pos}^{\rho_0}$ is clearly different from 0.5. So we have to determine the shift function $s(\rho)$, resp. s^{-1}, by solving the equation $P_\rho(T_{pos}^{s(\rho)}) = 0.5$, as described in Section 2.1.

The function s^{-1} can be approximated by evaluating the shift between $s(\rho)$ and ρ for some points $s(\rho)$ and compensate it with a polynomial of order three. For $s(\rho) \in \{0.47, 0.48, \ldots, 0.99\}$ ρ is numerically determined, such that $|P_\rho(T_{pos}^{s(\rho)}) - 0.5| \leq 10^{-4}$. We start at $s(\rho) = 0.47$, because for $s(\rho) = 0.46$ we got a solution $\rho < 0$. The results are displayed in Table 4.2 and Figure 4.5 shows the approximated $s^{-1}(\cdot)$, where the points $(s(\rho), \rho)$ are emphasized.

In $\rho = 0$ there does not exist a deviation, because $p_0 = 0.5$. But if we fix ρ and search for solutions $s(\rho)$ such that $p_{\rho,s(\rho)} = 0.5$, the question is, why there appears such a big jump from the solution $s(0) = 0$ to the solution $s(0.01) = 0.47$. The reason for this jump near $\rho = 0.01$ shows Figure 4.6. It displays that for $\rho = 0$ there exists two parameter ρ', namely $\rho' = 0$ and $\rho' = 0.461$, with $p_{0,\rho'} = \frac{1}{2}$. $p_{0,\rho}$ is larger than $\frac{1}{2}$ for $\rho \in (0, 0.461)$ and then decreases for $\rho > 0.461$. Numerical calculations showed that only for $\rho = 0$ there exist two solutions ρ' with $p_{0,\rho'} = \frac{1}{2}$, but we have no proof for this. For small $\rho > 0$, like $\rho = 0.01$ in Figure 4.6, the function $p_{\rho,\cdot}$ changes only a little bit.

For $\rho > 0$ it is $P_\rho(T_{pos}^0) > 0.5$. This can be seen as follows:

Lemma 4.13. *It is*

$$\lim_{\rho \to 0} v_1(x, \rho) = \begin{cases} \infty, & x > 0 \\ 1, & x = 0 \\ 0, & x < 0 \end{cases}$$

and

$$\lim_{\rho \to 0} v_2(x, \rho) = \begin{cases} 0, & x > 0 \\ -1, & x = 0 \\ -\infty, & x < 0 \end{cases} .$$

Table 4.2.: The shift $s(\rho)$.

$s(\rho)$	ρ	$s(\rho)$	ρ	$s(\rho)$	ρ
0.47	0.0101	0.65	0.2800	0.83	0.6317
0.48	0.0223	0.66	0.2980	0.84	0.6525
0.49	0.0348	0.67	0.3160	0.85	0.6740
0.50	0.0477	0.68	0.3340	0.86	0.6952
0.51	0.0612	0.69	0.3525	0.87	0.7165
0.52	0.0746	0.70	0.3712	0.88	0.7380
0.53	0.0886	0.71	0.3905	0.89	0.7597
0.54	0.1030	0.72	0.4090	0.90	0.7813
0.55	0.1175	0.73	0.4290	0.91	0.8030
0.56	0.1322	0.74	0.4485	0.92	0.8247
0.57	0.1475	0.75	0.4680	0.93	0.8465
0.58	0.1630	0.76	0.4878	0.94	0.8684
0.59	0.1790	0.77	0.5080	0.95	0.8902
0.60	0.1951	0.78	0.5280	0.96	0.9121
0.61	0.2115	0.79	0.5490	0.97	0.9341
0.62	0.2280	0.80	0.5690	0.98	0.9560
0.63	0.2455	0.81	0.5900	0.99	0.9780
0.64	0.2625	0.82	0.6107		

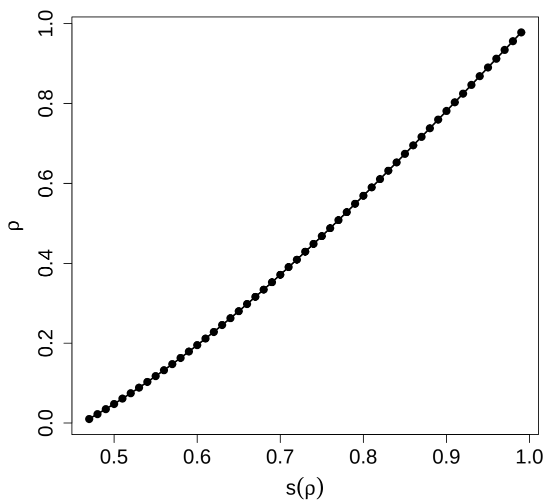

Figure 4.5.: Points $(s(\rho), \rho)$ and graph of the approximated $s^{-1}(\cdot)$.

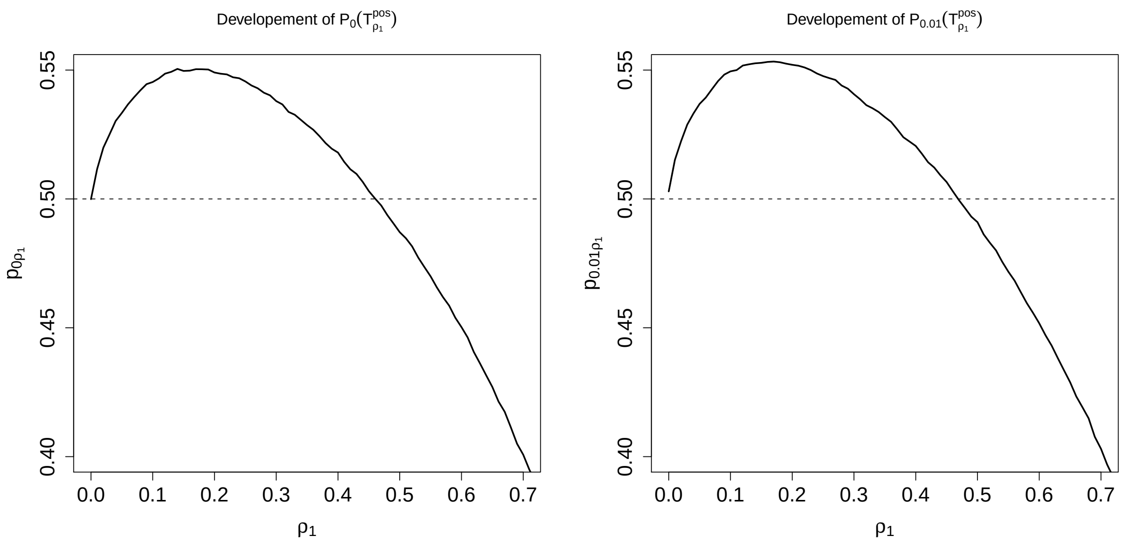

Figure 4.6.: Evolution of $p_{\rho,\rho_1} = P_\rho(T^{\rho_1}_{pos})$ for $\rho = 0$ and $\rho = 0.01$.

Proof: We show the claim for $v_1(x,\rho)$. The proof for $v_2(x,\rho)$ works analogously. Let be $x \neq 0$. Then we have

$$v_1(x,\rho) = \frac{1}{2\rho}\left(x(\rho^2+1)^2 + \sqrt{x^2(1-\rho^2) - 4\rho^2(\rho^2-1)}\right)$$

and for $\rho \to 0$

$$\left(\underbrace{x(\rho^2+1)}_{\to x} + \sqrt{\underbrace{x^2(1-\rho^2)^2}_{\to x^2} - \underbrace{4\rho^2(\rho^2-1)}_{\to 0}}\right) \to \begin{cases} 2x, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

This yields

$$\lim_{\rho\to 0} v_1(x,\rho) = \lim_{\rho \to 0} \frac{1}{2\rho}\left(x(\rho^2+1) + \sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}\right) = \infty,$$

if $x > 0$. Further, with the rule of l'Hospital it is $\lim\limits_{\rho\to 0} v_1(x,\rho) = 0$ for $x < 0$:

$$\begin{aligned}
\lim_{\rho\to 0} v_1(x,\rho) &= \lim_{\rho\to 0} \frac{x(\rho^2+1) + \sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}}{2\rho} \\
&= \lim_{\rho\to 0} \frac{2\rho x + (x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1))^{-\frac{1}{2}}(-2x^2(1-\rho^2)2\rho - 16\rho^3 + 8\rho)}{2} \\
&= \lim_{\rho\to 0} \underbrace{\rho x}_{\to 0} + \frac{\overbrace{\rho(-x^2+\rho^2x^2 - 8\rho^2 + 4)}^{\to 0}}{\underbrace{\sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}}_{\to |x|}} = 0
\end{aligned}$$

For $x = 0$ we see

$$\lim_{\rho\to 0} v_1(0,\rho) = \lim_{\rho\to 0}\left(0 + \frac{\sqrt{-4\rho^4+4\rho^2}}{2\rho}\right) = \lim_{\rho\to 0}\frac{2\rho\sqrt{1-\rho^2}}{2\rho} = \lim_{\rho\to 0}\sqrt{1-\rho^2} = 1$$

and the claim is proved. □

Further, we obtain

$$\begin{aligned}
P_\rho(T^0_{pos}) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{x^2}{2}}\left(\Phi\left(-\frac{\rho x}{\sqrt{1-\rho^2}}\right) - 0\right)dx \\
&\quad + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\left(1-\Phi\left(-\frac{\rho x}{\sqrt{1-\rho^2}}\right)\right)dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\,dx + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{x^2}{2}}\Phi\left(-\frac{\rho x}{\sqrt{1-\rho^2}}\right)dx \\
&\quad - \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\Phi\left(-\frac{\rho x}{\sqrt{1-\rho^2}}\right)dx \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\left(\Phi\left(\frac{\rho x}{\sqrt{1-\rho^2}}\right) - \Phi\left(-\frac{\rho x}{\sqrt{1-\rho^2}}\right)\right)dx \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\left(\Phi\left(\frac{\rho x}{\sqrt{1-\rho^2}}\right) - 1 + \Phi\left(\frac{\rho x}{\sqrt{1-\rho^2}}\right)\right)dx \\
&= \sqrt{\frac{2}{\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\underbrace{\Phi\left(\frac{\rho x}{\sqrt{1-\rho^2}}\right)}_{>\frac{1}{2},\text{ if }\rho>0(x>0)}dx \\
&> \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{x^2}{2}}\,dx = \frac{1}{2}.
\end{aligned}$$

We have $p_{\rho,0} > \frac{1}{2}$ so that the solution ρ' with $p_{\rho,\rho'} = \frac{1}{2}$ is unique, implying that $s(\rho) = \rho'$ is well defined. We estimate the correlation as zero, if we receive maximal likelihood-depth for a parameter $\tilde{\rho} < 0.461$ and correct it with the inverse of the shift-function otherwise.

A least square fit of a polynomial of degree three to the points $(s(\rho), \rho)$, see Table 4.2, leads to

$$s^{-1}(s(\rho)) = -1.24101\, s(\rho)^3 + 3.68702\, s(\rho)^2 - 1.4546\, s(\rho) + 0.00857.$$

We decided for a polynomial of degree three, because the maximum absolute error for compensation with a polynomial of degree two was 0.006, for compensation with degree three 0.00041 and for degree four only little smaller (0.0004). A new estimator for the correlation in a dataset can be defined by

$$\begin{aligned}
\hat{\rho} &= -1.24101(\arg\max_\rho d_T(\rho, z_*))^3 + 3.68702(\arg\max_\rho d_T(\rho, z_*))^2 \\
&\quad -1.4546(\arg\max_\rho d_T(\rho, z_*)), +0.00857
\end{aligned} \tag{4.1}$$

if $\arg\max_\rho d_T(\rho, z_*) \geq 0.461$ and $\hat{\rho} = 0$ else. Recall that $x_m = \arg\max f(x)$ iff $f(x_m) = \max_{x\in D} f(x)$, D the domain of f.

In the following we show that this new estimator is strongly consistent. Therefore we use Proposition 2.8 on page 13 and, to show that the assumptions of this theorem are fulfilled, Theorem A.15 on page 182.

We start by proving that $\mathscr{C} := \{T^\rho_{pos}; 0 < \rho < 1\}$ is a VC-class.

Theorem 4.14. *The conjunction of the sets T^ρ_{pos}, $0 < \rho < 1$, is a VC-class and has a VC-index $V(\mathscr{C}) < 7$.*

Proof: We already elaborated in Lemma 4.11 that $T^\rho_{pos} = \{(x,y) \in \mathbb{R}^2; v_2(x,\rho) \leq y \leq v_1(x,\rho)\}$, where

$$v_{1,2}(x,\rho) = \frac{1}{2\rho}\left(x(\rho^2+1) \pm \sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}\right).$$

With $(x,y) \in T^\rho_{pos}$ we obtain $(y,x) \in T^\rho_{pos}$ and $(-x,-y) \in T^\rho_{pos}$, and we also have $(-y,-x) \in T^\rho_{pos}$. Thus, if we check, if $(x,y) \in T^\rho_{pos}$, we can transform (x,y) to $(\tilde{x},\tilde{y})$, such that $\tilde{x} \geq 0$ and $\tilde{y} \leq \tilde{x}$. Then $(x,y) \in T^\rho_{pos}$, iff $\tilde{y} \geq v_2(\tilde{x},\rho)$, as $\tilde{y} \leq v_1(\tilde{x},\rho)$ is always true because $\tilde{y} \leq \tilde{x} \leq v_1(\tilde{x},\rho)$. Because of this, it is sufficient to consider points (x,y) with $x \geq 0$ and $y \leq x$.

The next step is to show that for every $z = (x,y)$ there are only finitely many intervals $[\rho_{i_1}, \rho_{i_2}]$, $0 < \rho_{i_1} < \rho_{i_2} < 1$, such that $z \in T^\rho_{pos}$ for $\rho \in [\rho_{i_1}, \rho_{i_2}]$. That is true, if $v_2(x,\cdot)$ takes every value only a finite time, i.e. $v_2(x,\cdot)$ has only for a finite number of values the slope zero. Therefore we regard the derivative of $v_2(x,\cdot)$. Since it is $0 < \rho < 1$, it holds

$$\begin{aligned}
\frac{\partial}{\partial\rho} v_2(x,\rho) &= \frac{\partial}{\partial\rho}\frac{1}{2\rho}(x(\rho^2+1) - \sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}) \\
&= -\frac{1}{2\rho^2}(x(\rho^2+1) - \sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}) \\
&\quad + \frac{1}{2\rho}\left(u2\rho - \frac{-2x^2\rho + 2x^2\rho^3 - 8\rho^3 + 4\rho}{2\rho\sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}}\right) \\
&= \frac{x(\rho^2-1)\sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)} + x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}{2\rho^2\sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}} \\
&\quad + \frac{2x^2\rho^2 - 2x^2\rho^4 + 8\rho^4 - 4\rho^2}{2\rho^2\sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)}}.
\end{aligned}$$

$\frac{\partial}{\partial\rho} v_2(x,\rho) = 0$ is true, iff

$$x(\rho^2-1)\sqrt{x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1)} + x^2 + 2x^2\rho^4 + 4\rho^4 = 0,$$

which is equivalent to

$$x^2(1-\rho^2)^2 - 4\rho^2(\rho^2-1) = \frac{(-x^2-2x^2\rho^4-4\rho^4)^2}{x^2(\rho^2-1)^2}$$
$$\Leftrightarrow -4\rho^2x^4 + 2\rho^4x^4 - 4\rho^6x^4 - 3\rho^8x^4 - 12\rho^4x^2 + 12\rho^6x^2 - 8\rho^8x^2 + 4\rho^2x^2 + 16\rho^8 = 0.$$

This is a polynomial with degree 8 for ρ, so it has at most 8 zeros, especially the number of zeros is finite. That means for every $z = (x, y)$, with $x \geq 0, y \leq x$ there are at most $l = 9$ intervals $[\rho_{i_1}, \rho_{i_2}]$ such that $z \in T^\rho_{pos}$ for $\rho \in [\rho_{i_1}, \rho_{i_2}]$.

Now we show that $V(\mathscr{C}) < 7$. Let be $\{z_1, \ldots, z_7\}$ with $z_k = (x_k, y_k)$, where it is enough to consider $x_k \geq 0, y_k \leq x_k$, $k = 1, \ldots, 7$, as discussed above. We already stated that for every z there are at most $l = 9$ intervals $[\rho_{i_1}, \rho_{i_2}]$ such that $z \in T^\rho_{pos}$ for $\rho \in [\rho_{i_1}, \rho_{i_2}]$, $1 \leq i \leq 9$. Every interval has 2 endpoints so there are at most $2 \cdot 9$ endpoints for every z. The first point z_1 divides the interval $[0, 1]$ into maximal $2 \cdot 9 + 1$ subsets. For every point that is added, there will be at most $2 \cdot 9$ new subsets. All in all we get at most $7 \cdot 2 \cdot 9 + 1 = 127$ subsets. To shatter the points there are $2^7 = 128$ subsets needed. Therefore not all possible subsets of $\{z_1, \ldots, z_7\}$ are picked out. This is $V(\mathscr{C}) < 7$. □

We prove that $\mathscr{C} = \{T^\rho_{pos}; 0 < \rho < 1\}$ is a VC-class, which yields together with Theorem A.15 on page 182, that $\lambda^+_N(\cdot, Z_{*,N} = (Z_1, \ldots, Z_N)) = \frac{1}{N}\sharp\{Z_n = (X_n, Y_n); (X_n, Y_n) \in T^{(\cdot)}_{pos}\}$ converges uniformly to $\lambda^+_{\rho_0}(\cdot) = P_{\rho_0}(v_2(X, \cdot) \leq Y \leq v_1(X, \cdot))$ and $\lambda^-_N(\cdot, Z_{*,N})$ converges uniformly to $\lambda^-_{\rho_0}(\cdot)$.

Proposition 4.15. *The corrected maximum likelihood-depth estimator $\hat{\rho}$ given by (4.1) is a strongly consistent estimator for ρ.*

Proof: We use Theorem 2.8 on page 13, Theorem 2.12 on page 16 and the above proven statements to show the claim. We can assume that $\lambda^+_{\rho_0}(\rho) = P_{\rho_0}(T^\rho_{pos}) > \frac{1}{2}$ for $\rho < s(\rho_0)$. To explain this see Figure 4.7, where the evolution of $P_{\rho_0}(T^\rho_{pos})$ is displayed for different ρ_0. We see also that we can assume $\lambda^-_{\rho_0}(\rho) = 1 - \lambda^+_{\rho_0} < \frac{1}{2}$ for $\rho < s(\rho_0)$ and also $\lambda^+_{\rho_0}(\rho) < \frac{1}{2}$ for $\rho > s(\rho_0)$, i.e. the conditions of Theorem 2.8 are true. So $\tilde{\rho}_N = \arg\max_{0<\rho<1} d_T(\rho, Z_{*,N})$ is a consistent estimator for $s(\rho_0)$. As s^{-1} is continuous, we obtain $\hat{\rho} = s^{-1}(\tilde{\rho}_N)$ being also consistent. □

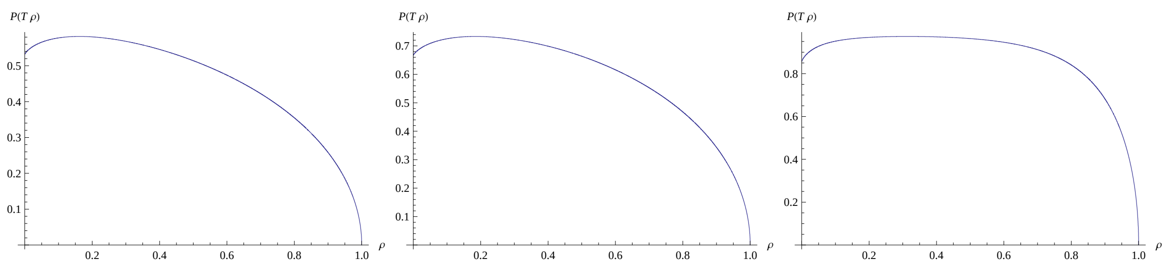

Figure 4.7.: $\lambda^+_{\rho_0}(\rho)$ for $\rho_0 = 0.1, 0.5$ and 0.9

The new estimator is compared to the correlation coefficient of Bravais-Pearson. This

Table 4.3.: New estimator LDE compared to Pearson's correlation coefficient in samples with different ρ, $N = 100$, 1000 repetitions each.

ρ	Pearson	LDE
0.01	0.0647	0.0063
0.1	0.0946	0.1213
0.2	0.2032	0.2039
0.3	0.2935	0.2920
0.4	0.4017	0.3943
0.5	0.4984	0.4891
0.6	0.5993	0.5926
0.7	0.6987	0.6945
0.8	0.7999	0.7963
0.9	0.8999	0.8989

correlation coefficient is defined for data $(x_1, y_1), \ldots, (x_N, y_N)$ as

$$\frac{\sum_{j=1}^{N}(x_j - \overline{x_*})(y_j - \overline{y_*})}{\sqrt{\sum_{j=1}^{N}(x_j - \overline{x_*})^2 \sum_{j=1}^{N}(y_j - \overline{y_*})^2}},$$

where $\overline{w_*}$ denotes the mean of $w_* = (w_1, \ldots, w_N)$.

For every $\rho \in \{0.1, 0.2, \ldots, 0.9\}$ 1000 datasets consisting of 100 data each were simulated. Table 4.3 shows the means of the new estimator (LDE) in comparison to the means of the correlation coefficient and Table 4.4 tables the mean squared errors (MSE) of the estimators. If there is more than one parameter with maximum depth, the mean of the arguments $\tilde{\rho} := \text{mean}(\tilde{\rho}_1, \tilde{\rho}_2, \ldots)$, $\tilde{\rho}_1, \tilde{\rho}_2 \ldots \in \arg\max_\rho d_T(\rho, z_*)$, is calculated and the estimator is given by $\hat{\rho} = s^{-1}(\tilde{\rho})$.

This shows that Pearson's correlation coefficient is most times slightly better than the new estimator. If we look at datasets with ε-contamination, the new estimator achieves some times better results than the correlation coefficient (especially for high correlation). We compare again the means of both estimators, this time for ε-contaminated data from $(1-\varepsilon)P_{\rho_0} + \varepsilon P_{\rho_1}$, with $\varepsilon > 0$ and P_ρ the two-dimensional normal distribution with mean $\mu = (0, 0)$ and covariance $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The ratio of contaminated data was $\varepsilon = 0.1$, see Table 4.5.

Tables 4.5 and 4.6 show that in case of data with contamination with low correlation the new estimator is better than the correlation coefficient of Bravais-Pearson for high correlated data, even for a small ratio of contamination. The greater the correlation is, the better are the results the new estimator achieves in comparison to Pearson's coefficient. If we construct contaminated data with high correlation, Pearson's correlation coefficient receives better results, especially for low correlation.

Table 4.4.: MSE of the LDE and Pearson's correlation coefficient in samples with different ρ, $N = 100$, 1000 repetitions each.

	Mean squared error	
ρ	Pearson	LDE
0.1	0.01031	0.01543
0.2	0.00843	0.02083
0.3	0.00843	0.02076
0.4	0.00675	0.01739
0.5	0.00532	0.01278
0.6	0.00450	0.00839
0.7	0.00257	0.00502
0.8	0.00134	0.00229
0.9	0.00036	0.00079

Table 4.5.: Estimation for ε-contaminated data with contamination correlation ρ_1, $\varepsilon = 0.1$.

ρ_0	ρ_1	Pearson	LDE	ρ_1	Pearson	LDE
0.10	0.01	0.0936	0.1276	0.99	0.1913	0.2877
0.20	0.01	0.1828	0.1891	0.99	0.2780	0.3729
0.30	0.01	0.2659	0.2741	0.99	0.3683	0.4494
0.40	0.01	0.3636	0.3653	0.99	0.4577	0.5221
0.50	0.01	0.4544	0.4606	0.99	0.5455	0.6011
0.60	0.01	0.5432	0.5603	0.99	0.7300	0.7657
0.70	0.01	0.6316	0.6608	0.99	0.8175	0.8392
0.80	0.01	0.7197	0.7645	0.99	0.8440	0.8180
0.90	0.01	0.8111	0.8796	0.99	0.9085	0.9167

Table 4.6.: MSE for ε-contaminated data with correlation ρ_1, $\varepsilon = 0.1$.

ρ_0	ρ_1	Pearson	LDE	ρ_1	Pearson	LDE
0.1	0.01	0.00975	0.01852	0.99	0.01847	0.06137
0.2	0.01	0.00946	0.02015	0.99	0.01578	0.05403
0.3	0.01	0.00986	0.02168	0.99	0.01296	0.04016
0.4	0.01	0.00891	0.01942	0.99	0.01037	0.02928
0.5	0.01	0.00901	0.01502	0.99	0.00742	0.02121
0.6	0.01	0.00832	0.01105	0.99	0.00510	0.01322
0.7	0.01	0.00922	0.00681	0.99	0.00318	0.00792
0.8	0.01	0.00972	0.00440	0.99	0.00146	0.00331
0.9	0.01	0.01026	0.00132	0.99	0.00041	0.00129

Table 4.7.: Estimation of ρ for data coming from $(1-\varepsilon)P_\rho + \varepsilon\delta_{x_0,x_0}$

ρ	x_0	$\varepsilon \cdot 100$	$\hat{\rho}$	Pearson
0.1	10^1	10	0.26	0.91
0.1	10^2	10	0.27	1
0.1	10^3	10	0.27	1
0.1	10^4	10	0.26	1
0.1	10^6	10	0.26	1
0.1	10^3	20	0.43	1
0.5	10^3	10	0.59	1

To study the behavior of the new estimator in comparison to the correlation coefficient of Pearson for growing sample size, we simulated data with correlation $\rho = 0.5$ and calculated the mean squared error for different sizes. For each N 1000 samples were constructed. The resulting roots of the means squared errors are displayed in Figure 4.8. We observe, that the difference in the MSE between the estimators shrinks with growing sample size. But Pearson's correlation coefficient has smaller MSEs for all sample sizes in uncontaminated data. We will show in a next simulation study that Pearson's correlation coefficient breaks down for contamination with real outliers, in contrast to the LDE.

As a last consideration we simulated outliers, i.e. we contaminated with a completely different distribution and simulated data from $(1-\varepsilon)P_\rho + \varepsilon\delta_{x_0,x_0}$, where δ_{x_0,x_0} is the Dirac measure on (x_0, x_0) and P_ρ, as before, the two-dimensional normal distribution with correlation ρ. Table 4.7 shows that the new estimator $\hat{\rho}$ is much more robust. In particular, there is no breaking down, i.e. it reaches not the upper bound 1 in contrast to the correlation coefficient of Bravais-Pearson. This is clearly an advantage of the LDE.

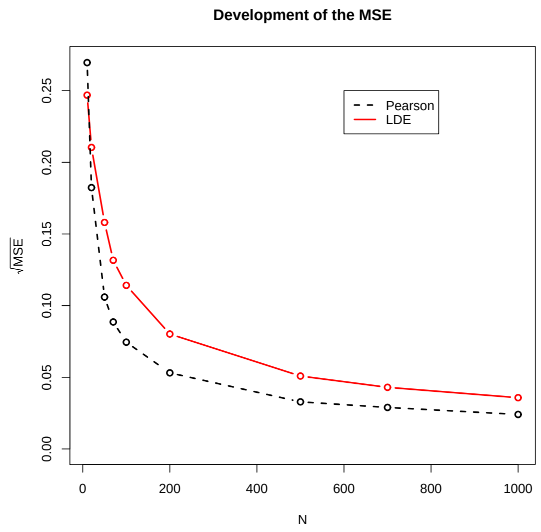

Figure 4.8.: MSE for growing sample size for the estimators of the correlation coefficient.

4.3. Tests and confidence intervals for the correlation coefficient

We still consider $Z_i = (X_i, Y_i)$, $i = 1, \ldots, N$, i.i.d. $X_i, Y_i \sim \mathcal{N}(0, 1)$ with dependence structure given by the Gaussian copula. Using the methods of Section 2.2 we construct tests for the hypotheses $H_0 : \rho \le \rho_0$, $H_0 : \rho \ge \rho_0$ and $H_0 : \rho = \rho_0$. We mentioned in Section 2.2, the power of the test for $H_0 : \rho \ge \rho_0$ is bad just using Corollary 2.17 on page 19. As we showed in the last section $s(\rho) > \rho$, and an improvement of the test has to be made in this case using Definition 2.18 on page 19 and Definition 2.19 on page 20.

We start with constructing a test for the null hypothesis $H_0 : \rho \le \rho_0$ against $H_1 : \rho > \rho_0$. The last section shows how the likelihood-depth can be calculated for the two-dimensional normal distribution, so we can use this to determine the simplicial depth of ρ. We also determined p_ρ in the section about estimation, see Table 4.1 on page 140. The test statistic is defined according to Lemma 2.14 as

$$T(\rho, z_*) := \sqrt{N} \frac{d_S(\rho, z_*) - 2p_\rho(1 - p_\rho)}{2\sqrt{p_\rho(1 - p_\rho)(1 - 2p_\rho)^2}}.$$

A direct conclusion from Corollary 2.17 on page 19 leads to

Corollary 4.16. *The test*

$$\varphi_{\rho_0}^{\le}(z_*) = 1_{\{\sup_{\rho \le \rho_0} T(\rho, z_*) < \Phi^{-1}(\alpha)\}}(z_*)$$

is an asymptotic α-level test for the hypothesis $H_0 : \rho \le \rho_0$ against $H_1 : \rho > \rho_0$, where $T(\rho, z_*)$ is defined as above and $\Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution.

This test shall be compared to an already existing test, the Fisher(-Samiuddin)-test, in a simulation study. The test statistic of the Fisher-test, see [Sac 2004] and [Sam 1970], is defined as

$$\hat{t}(Z) = \frac{r(Z) - \rho_0\sqrt{N-2}}{\sqrt{(1-r^2(Z))(1-\rho_0^2)}}$$

with $r(z) = \frac{\sum_{n=1}^{N}(x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^{N}(x_n - \bar{x})^2 \sum_{n=1}^{N}(y_n - \bar{y})^2}}$ and has an asymptotic t_{N-2} distribution.

We compare the power-function of the tests by simulating each 1000 times 100 data with distribution ρ, $\rho \in \{0.01, 0.02, \dots, 0.99\}$ and count, how often $H_0 : \rho \leq \rho_0$ is rejected. The results can be found in Figure 4.9.

Before taking a closer look at the results of this study, the robustness of the new test is compared to the robustness of the Fisher-(Samiuddin-)test. Therefore ε-contaminated datasets are simulated. Contaminated data means that $(1-\varepsilon)\cdot 100\%$ (here 95% and 90%) of the data are constructed with correlation ρ and $\varepsilon \cdot 100\%$ (here 5% and 10%) are constructed with $\rho_1 = 0.01$. The graphics in Figure 4.10 show the frequency of rejections of $H_0 : \rho \leq \rho_0$ for both tests with $\varepsilon = 0.05$ and Figure 4.11 shows the case with $\varepsilon = 0.1$. Figure 4.12 displays the power for contamination with $\rho_1 = 0.99 \geq \rho_0$.

For uncontaminated data the Fisher-Samiuddin-test has most times more power. For ε-contaminated data the new test succeeds in the cases of contamination data with low correlation in contrast to the Fisher-test. For the case of mixing with high correlated data, both tests do not keep the level.

As for the estimators, we also consider a contamination with a completely different distribution, i.e. data coming from $(1-\varepsilon)P_\rho + \varepsilon\delta_{(x_0,x_0)}$, with $\delta_{(x_0,x_0)}$ being the Dirac measure in (x_0, x_0) and P_ρ being the two-dimensional normal distribution with correlation ρ as before.

The new test is now much more robust than the Fisher-(Samiuddin-)test, see Figure 4.13. In particular, the level of the new test is not breaking down for x_0 tending to infinity which is the case for the Fisher-test.

Now we consider the hypothesis $H_0 : \rho \geq \rho_0$. Figure 4.14 shows the estimated power-function of the test, if we would not improve the power, for the example $\rho_0 = 0.8$.

So we will use Definition 2.18 on page 19 and Definition 2.19 on page 20 and try to improve the power of the test with the quantity

$$c_\alpha^1(\rho) = \max\{\rho; \lim_{N\to\infty} P_{\rho_0}(T(\rho, Z_*) < \Phi^{-1}(\alpha)) \leq \alpha\}.$$

The results of estimation of c_α^1 for $\rho = 0.05, \dots, 0.9$ are printed in Table 4.8. They were received by simulating datasets z_* with correlation ρ, determining the value of the teststatistic $T(\rho_1, z_*)$ for $\rho_1 = \rho, \dots, 0.999$ and counting how often $T(\rho_1, z_*) < \Phi^{-1}(\alpha)$. The values in Table 4.8 were achieved for $\alpha = 0.05$ and are the maximum values such

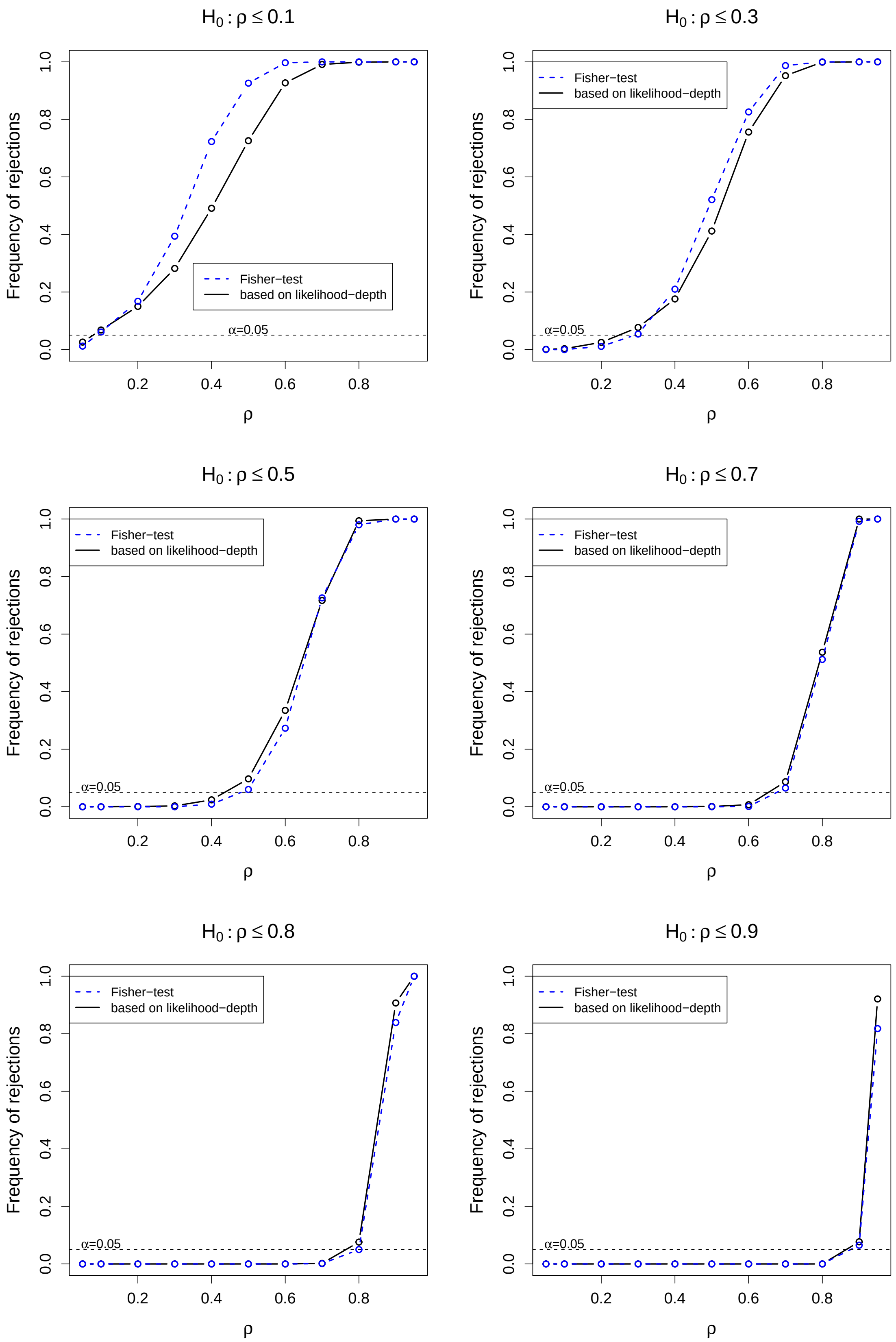

Figure 4.9.: Comparison of the power of the tests for $H_0 : \rho \leq \rho_0$.

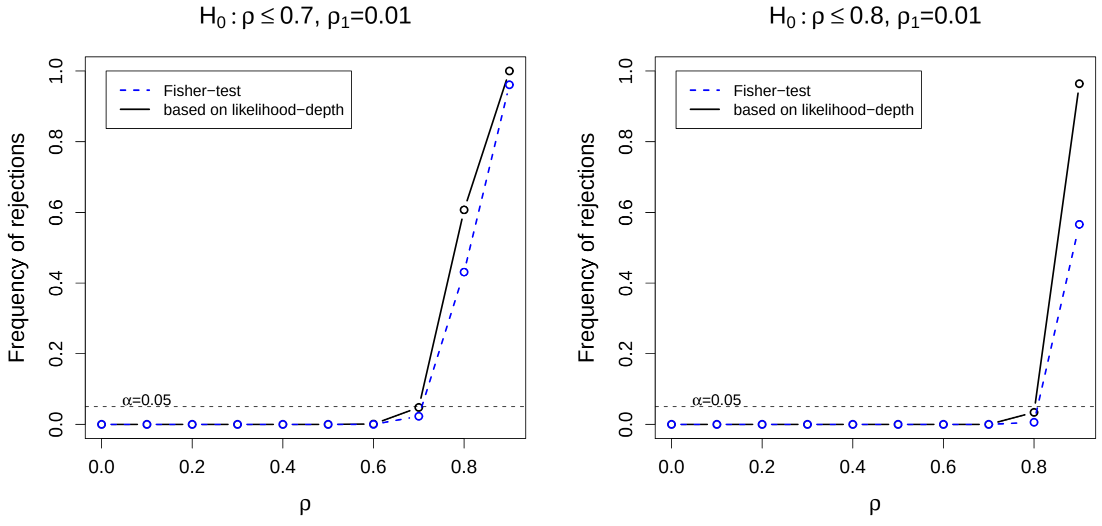

Figure 4.10.: Power of the tests for $H_0 : \rho \leq \rho_0$ with 5% contamination ($\rho_1 = 0.01$).

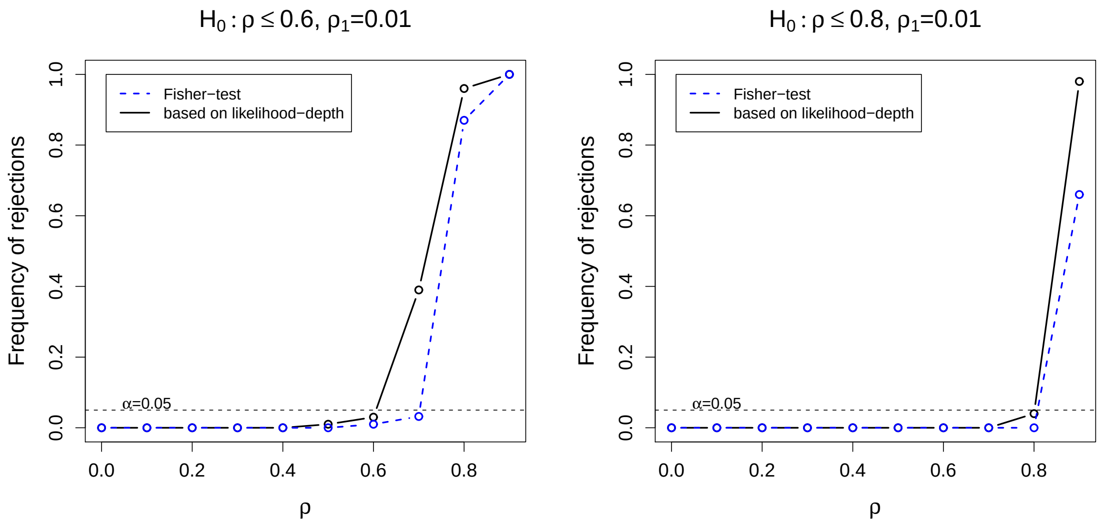

Figure 4.11.: Power of the tests for $H_0 : \rho \leq \rho_0$ with 10% contamination ($\rho_1 = 0.01$).

that $T(\rho_1, z_*) < \Phi^{-1}(\alpha)$ in maximum $\alpha \cdot 100\%$ of the cases. We made 1000 repetitions for every ρ and considered datasets with 1000 data.

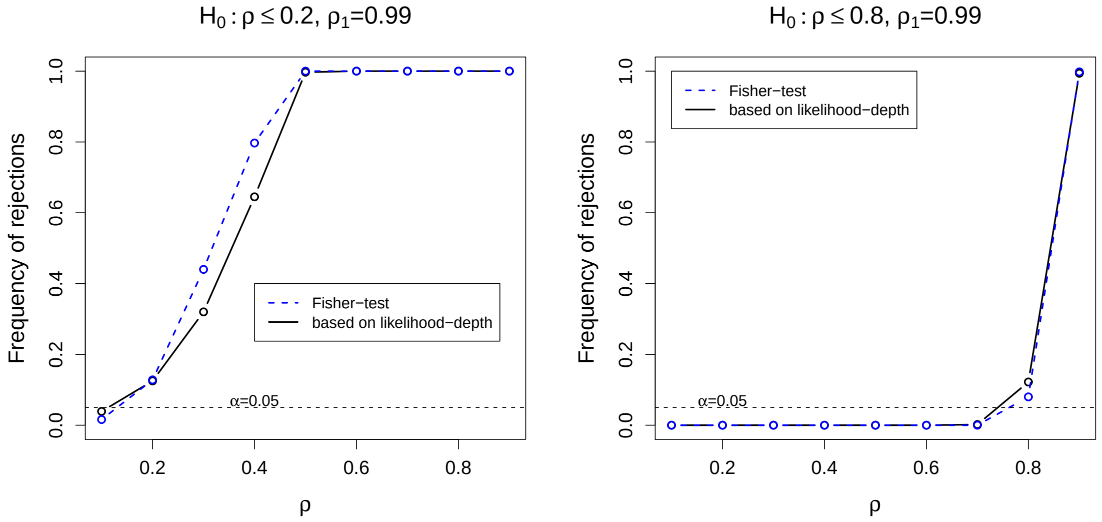

Figure 4.12.: Power of the tests for $H_0 : \rho \leq \rho_0$ with 5% contamination ($\rho_2 = 0.99$).

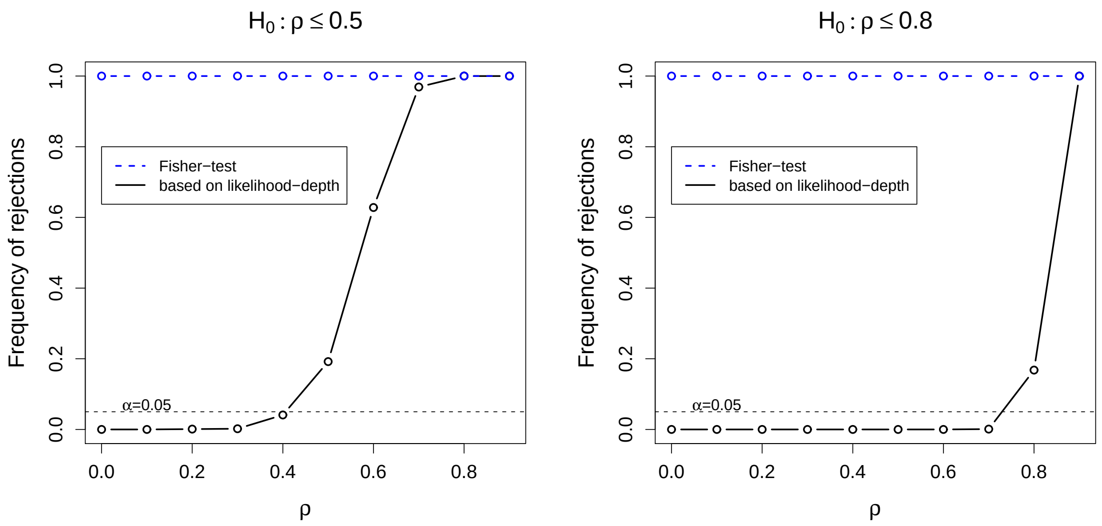

Figure 4.13.: Power-functions of $H_0 : \rho \leq \rho_0$ for data with 10% outliers in $(x_0, x_0) = (10^4, 10^4)$, $\rho_0 = 0.5, 0.8$.

Table 4.8.: Values of $c^1_{\alpha=0.05}(\rho)$.

ρ	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$\hat{c}^1_{\alpha=0.05}(\rho)$	0.819	0.831	0.852	0.862	0.872	0.880	0.888	0.897	0.902
ρ	0.5	0.55	0.6	0.65	0.7	0.70	0.8	0.85	0.9
$\hat{c}^1_{\alpha=0.05}(\rho)$	0.917	0.923	0.935	0.939	0.948	0.957	0.959	0.976	0.983

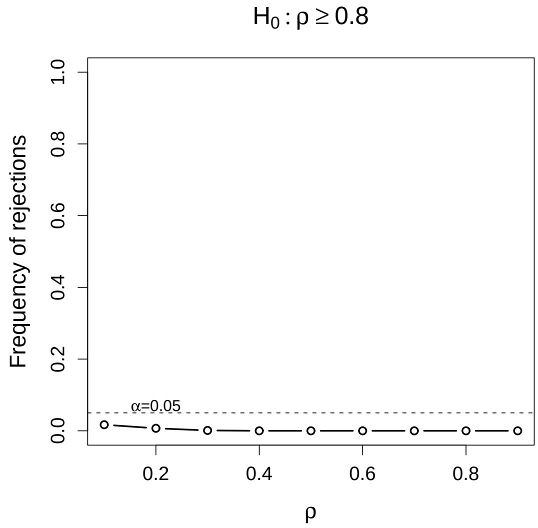

Figure 4.14.: Power of the new uncorrected test based on likelihood-depth for $H_0 : \rho \geq 0.8$.

We see that $\hat{c}_\alpha^1$ is increasing, so the assumptions to c_α^1, approximated by $\hat{c}_\alpha^1$, of Theorem 2.20 on page 20 seem to be satisfied. Under this assumption, we state the following

Corollary 4.17. *Assume c_α^1 to be increasing. The test*

$$\varphi_{\rho_0}^{\geq}(z_*) = 1_{\{\sup_{\rho \geq \hat{c}_\alpha^1(\rho_0)} T(\rho, z_*) < \Phi^{-1}(\alpha)\}}(z_*)$$

is an asymptotic α-level test for the hypothesis $H_0 : \rho \geq \rho_0$ against $H_1 : \rho < \rho_0$, when $T(\rho, z_*)$ is defined as in the beginning of this section, $\hat{c}_\alpha^1$ as above and $\Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution.

As $p_{\rho_0,(\cdot)}$ is not strictly decreasing, see also Figure 4.7 on page 145, unfortunately we can not prove consistency of the tests using the theorems of Section 2.2.

We could increase the accuracy and calculate $\hat{c}_\alpha^1(\rho)$ for more ρ or compensate with a polynomial to be able to get values for $\hat{c}_\alpha^1$ not estimated. But here we just check, if the power of the test for some ρ_0 improved or not. The results can be found in the graphics of Figure 4.15. We see is that the power has been really improved in comparison to the power of the test without correction, although it is not as good as the power of the Fisher-Samiuddin-test.

Now again we compare both tests for ε-contaminated data. In a first simulation every time 5% of the data is simulated with $\rho_1 = 0.1$, i.e. $\varepsilon = 0.05$. Here the Fisher(-Samiuddin)-test does not keep the level, see Figure 4.16. In a second study 10% of the data are simulated with $\rho_1 = 0.01$, i.e. $\varepsilon = 0.1$. For the results see Figure 4.17. Again the Fisher-(Samiuddin-)test does not keep the level, see especially the cases where $\rho_0 = 0.8$

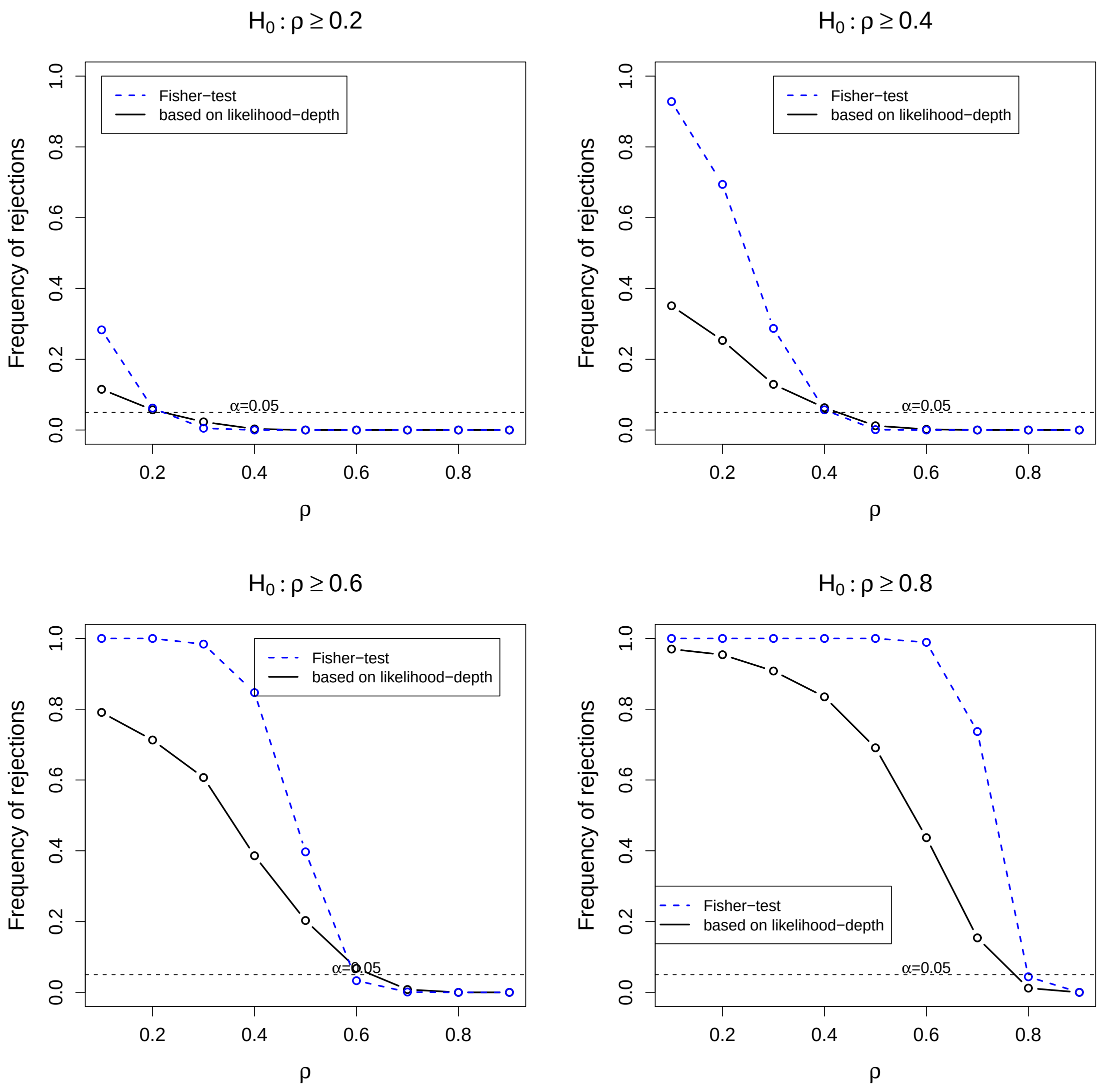

Figure 4.15.: Power of the tests (corrected test based on likelihood-depth and Fisher-test) for $H_0 : \rho \geq \rho_0$.

and $\rho_0 = 0.9$. For contamination with high correlation, $\rho_1 = 0.99$, the Fisher-test has more power, see Figure 4.18. But for contamination with nearly independent data the new test succeeds.

We simulate also data with outliers, i.e. data coming from $(1-\varepsilon)P_\rho + \varepsilon\delta_{(x_0,x_0)}$, with P_ρ and $\delta_{(x_0,x_0)}$ defined as before. The comparisons of the power of both tests for $x_0 = 10^4$, $\varepsilon = 0.1$ and $\rho_0 = 0.5$ resp. $\rho = 0.8$ are displayed in Figure 4.19. The Fisher test now never rejects $H_0 : \rho \geq \rho_0$ in contrast to the new test, which does not. Unfortunately, the new test has a bad power.

We developed robust tests for the hypotheses $H_0 : \rho \leq \rho_0$ and $H_0 : \rho \geq \rho_0$ based on likelihood depth. The next corollary gives the resulting test for $H_0 : \rho = \rho_0$.

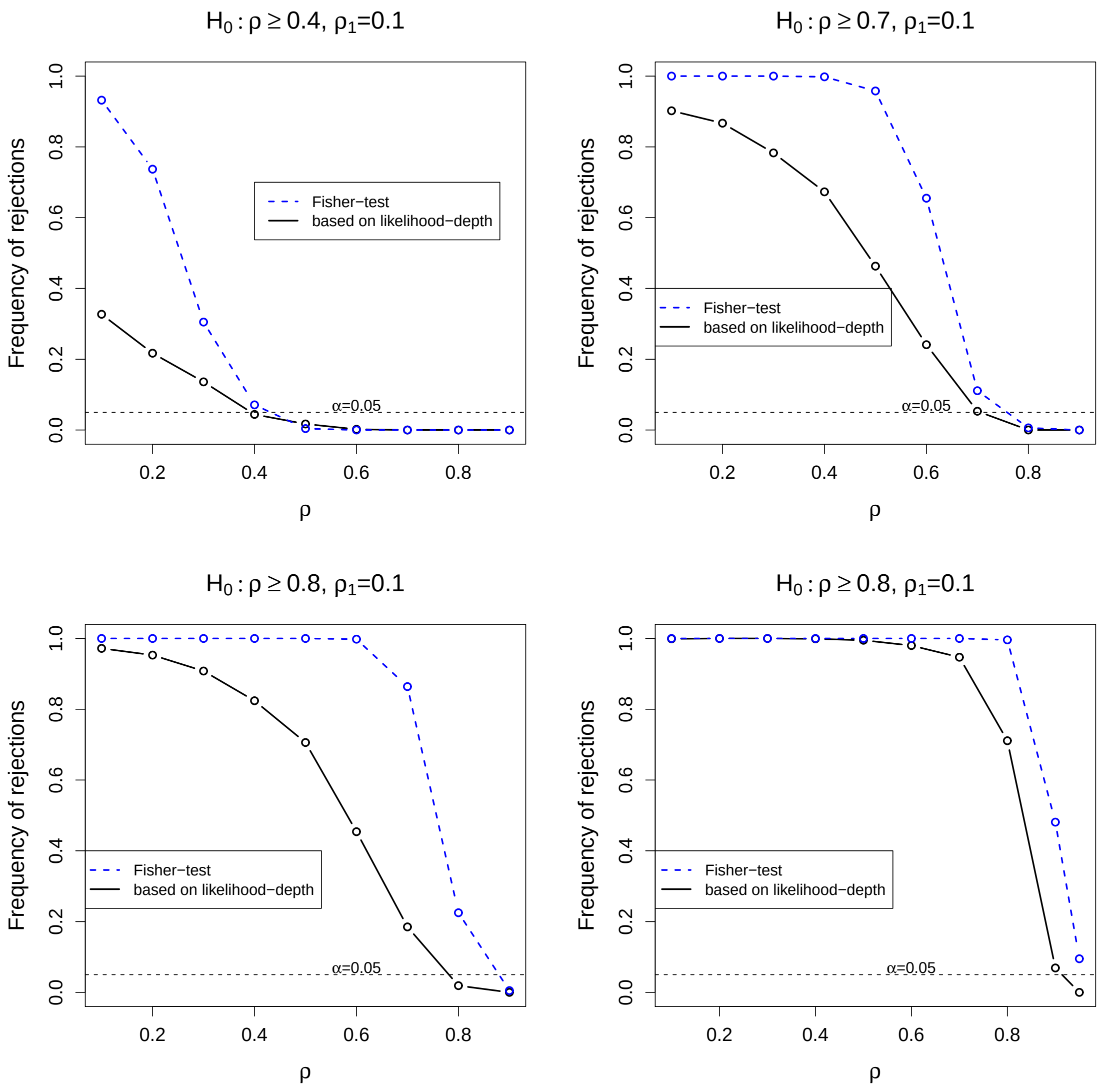

Figure 4.16.: Comparison of the tests of $H_0 : \rho \geq \rho_0$ for ε-contaminated data ($\rho_1 = 0.1$, $\varepsilon = 0.05$).

Corollary 4.18. *Assume c^1_α to be increasing. The test*

$$\varphi^{=}_{\rho_0}(z_*) = \max(1_{\{T(\rho_0,z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{T(\hat{c}^1_{\frac{\alpha}{2}}(\rho_0),z_*)<\Phi^{-1}(\frac{\alpha}{2})\}}(z_*)),$$

with $\hat{c}^1_\alpha$ as in Table 4.8, resp. Table 4.9, is an asymptotic α-level test for the hypothesis $H_0 : \rho = \rho_0$ against $H_1 : \rho \neq \rho_0$.

To test with level $\alpha = 0.05$ we have to determine $\hat{c}^1_{0.025}(\rho)$. We do this the same way we estimated $c^1_{0.05}(\rho)$ and receive the values given in Table 4.9.

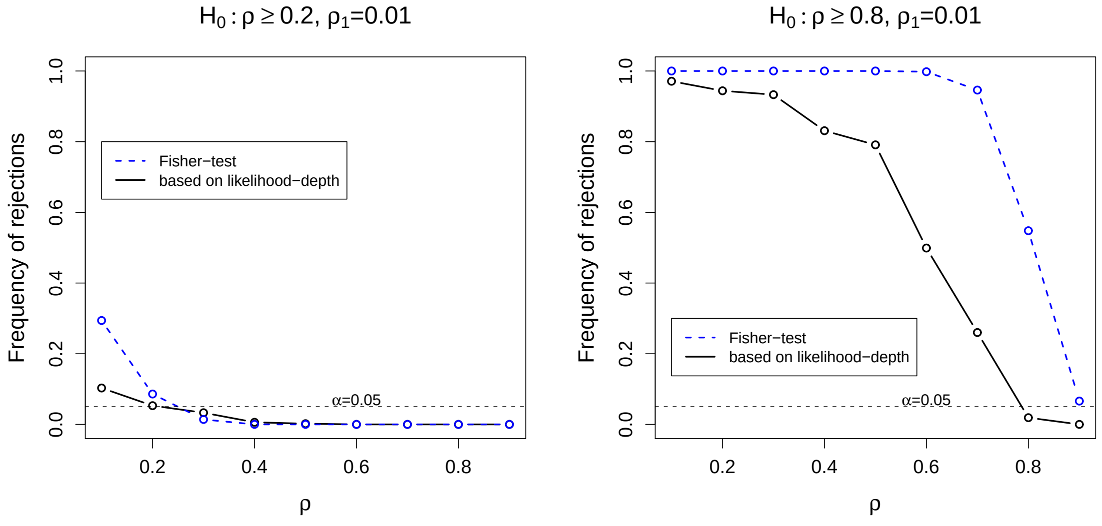

Figure 4.17.: Comparison of the tests of $H_0 : \rho \geq \rho_0$ for ε-contaminated data ($\rho_1 = 0.01$, $\varepsilon = 0.1$).

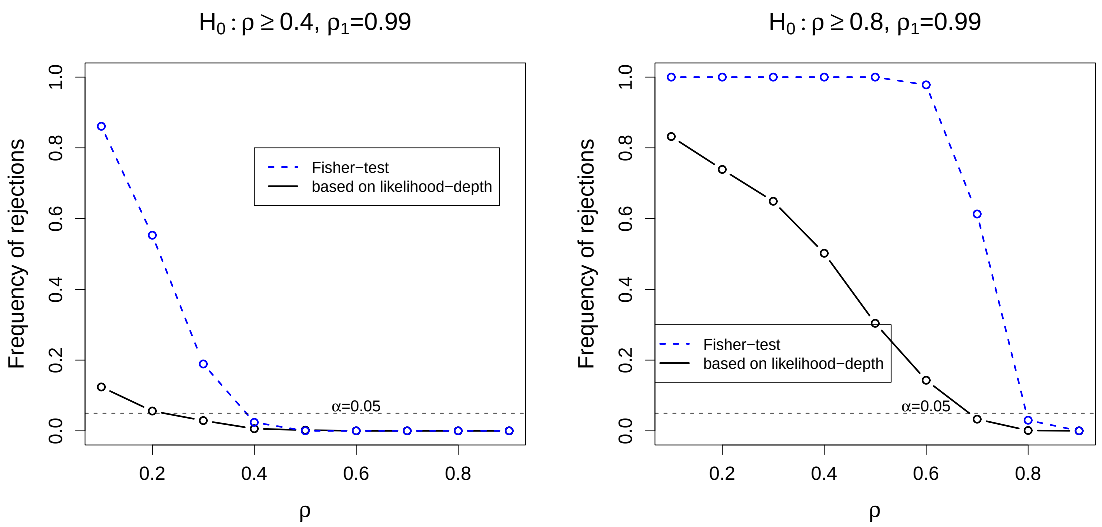

Figure 4.18.: Comparison of the tests of $H_0 : \rho \geq \rho_0$ for ε-contaminated data ($\rho_1 = 0.99$, $\varepsilon = 0.05$).

Table 4.9.: Values of $c^1_{\alpha=0.025}(\rho)$.

0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.82	0.838	0.852	0.867	0.892	0.914	0.929	0.948	0.965	0.982

The estimated power function of the new test based on likelihood-depth compared to the test by Fisher and Samiuddin is displayed in Figure 4.20. It shows that the power of the

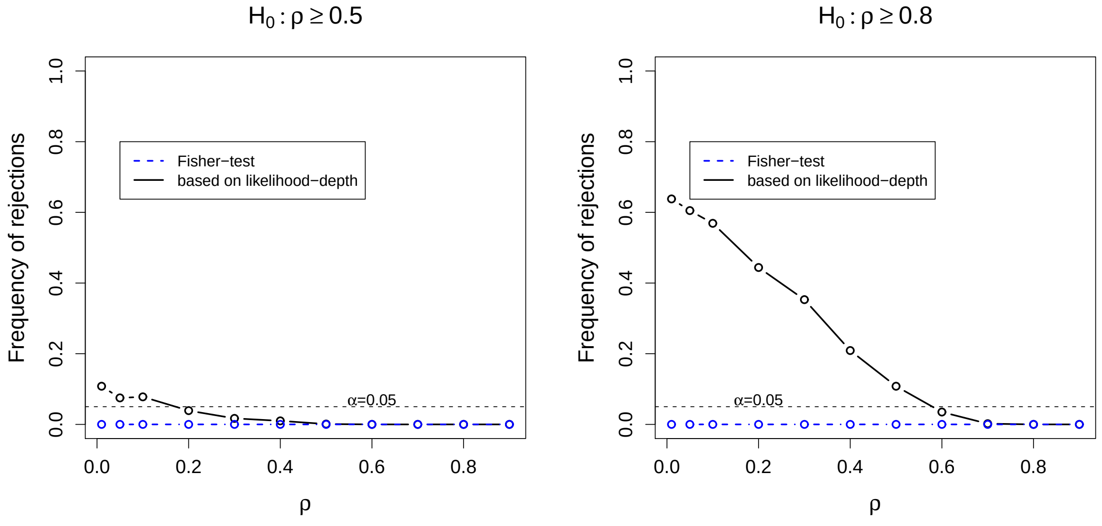

Figure 4.19.: Power-functions of $H_0 : \rho \geq \rho_0$ for data with 10% outliers in $(x_0, x_0) = (10^4, 10^4)$.

Fisher(-Samiuddin)-test is better than the power of the new test.

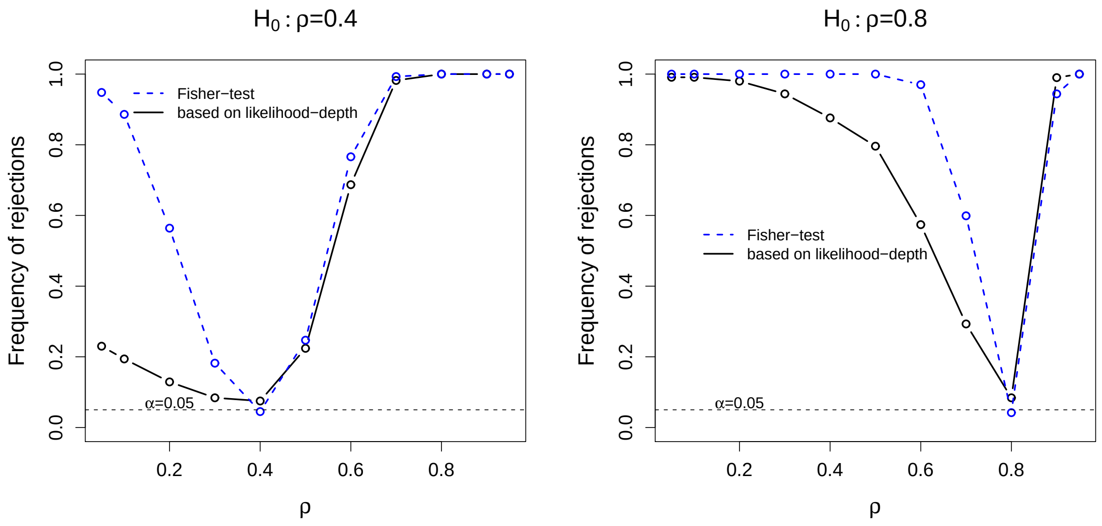

Figure 4.20.: Test for $H_0 : \rho = \rho_0$.

So far we only considered variables with mean $\mu = (0, 0)$ and variance $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. If $\mu = (\mu_1, \mu_2)$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$ with $\mu_1 \neq 0$ or $\mu_2 \neq 0$ or $\sigma_1 \neq 1$ or $\sigma_2 \neq 1$, the test can still be used, only the data has to be transformed before applying the test. Transform $Z = (X, Y)$ to $\widetilde{Z} = (\widetilde{X}, \widetilde{Y})$, where $\widetilde{X} = \frac{X - \mu_1}{\sigma_1}$, $\widetilde{Y} = \frac{Y - \mu_2}{\sigma_2}$. For the data with unknown

$\mu_1, \mu_2, \sigma_1, \sigma_2$, we transform $z = (x, y)$ to $\widetilde{z_i} = (\widetilde{x_i}, \widetilde{y_i})$ with $\widetilde{x_i} = \frac{x_i - \overline{x_*}}{s(x_*)}$ and $\widetilde{y_i} = \frac{y_i - \overline{y_*}}{s(y_*)}$, where $s(w_*)$ is the standard deviation of w_*. The power of the test is not really infected by this transformation, see Figure 4.21. Still the power of the Fisher-test is better.

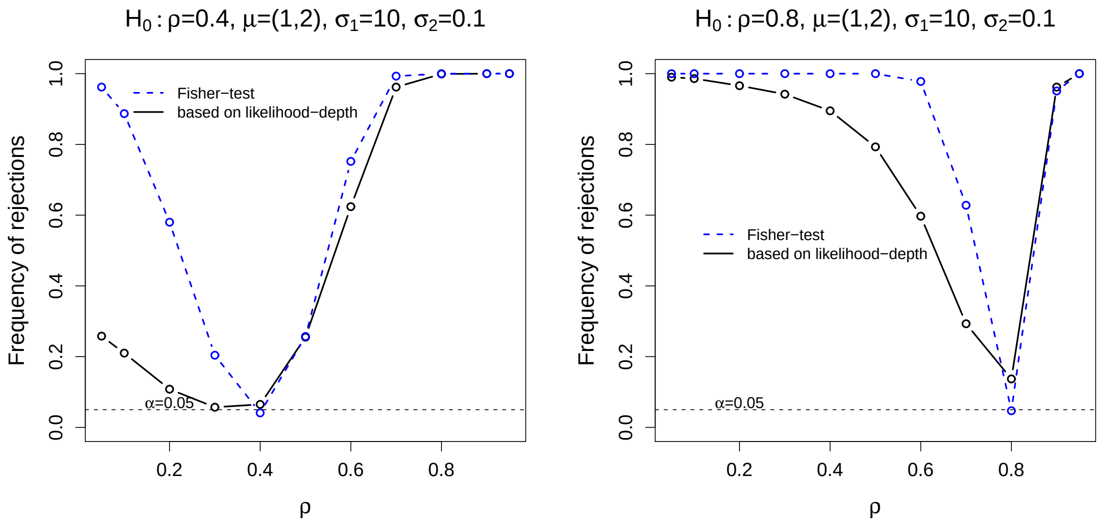

Figure 4.21.: Test for $H_0 : \rho = \rho_0$ with $\mu \neq (0, 0)$ and $\sigma_1 \neq 1 \neq \sigma_2$.

Here too, we consider data with ε-contamination. A ratio of $\varepsilon = 0.1$ of the data is distributed with $\rho_1 = 0.05$. The resulting estimated power-functions are displayed in Figure 4.22. Here the Fisher-test does not keep the level in contrast to the new test based on likelihood-depth.

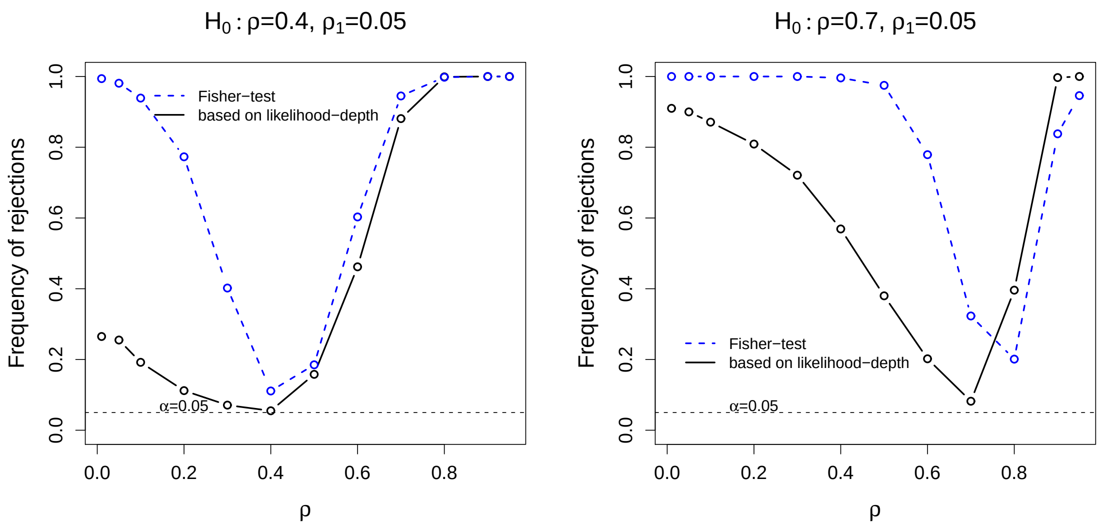

Figure 4.22.: Test for $H_0 : \rho = \rho_0$ with ε-contaminated data ($\varepsilon = 0.1$, $\rho_1 = 0.05$).

All in all we can say that the power of the new test is for the hypotheses $H_0 : \rho \leq \rho_0$ comparable to the power of the Fisher-Samiuddin-test, for the hypotheses of type $H_0 : \rho \geq \rho_0$

it behaves worse, but in contaminated data it is more robust than the power of the Fisher-Samiuddin-test.

4.4. Estimator for the parameter of the Gumbel copula

In this section we use the likelihood-depth to find an estimator for the parameter of the Gumbel copula. Recall the definition of the Gumbel copula, Definition 4.6 on page 134 . We start with $Z_1, \ldots, Z_N$ i.i.d., $Z_i \sim \text{Gum}(\theta)$, $i = 1, \ldots, N$, that means the distribution is given by the Gumbel copula and the margins are uniform on $[0,1]$. First of all we need to determine the density function of a Gumbel copula with parameter θ.

Lemma 4.19. *The density function of the Gumbel copula is*

$$\begin{aligned} f_\theta(z) &= \frac{(-\ln v)^{\theta-1}(-\ln u)^{\theta-1}}{uv} e^{-((-\ln u)^\theta+(-\ln v)^\theta)^{\frac{1}{\theta}}} \\ &\quad \left(\theta - 1 + ((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\right)((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-2} \end{aligned}$$

for $z = (u,v) \in [0,1] \times [0,1]$, $\theta \geq 1$.

Proof: We calculate f_θ as $f_\theta(z) = \frac{\partial}{\partial u}\frac{\partial}{\partial v} C_\theta(u,v)$ and start by determining the partial derivative in v:

$$\begin{aligned} \frac{\partial}{\partial v} C_\theta(u,v) &= \frac{\partial}{\partial v} \exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\} \\ &= -\frac{1}{\theta}\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{\frac{1}{\theta}-1} \theta(-\ln v)^{\theta-1}(-\frac{1}{v}) \\ &\quad \exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\} \\ &= \frac{1}{v}((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-1}(-\ln v)^{\theta-1}\exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\}. \end{aligned}$$

Thus, the density function is

$$\begin{aligned} f_\theta(z) &= \frac{\partial}{\partial u}\left(\frac{1}{v}((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-1}(-\ln v)^{\theta-1}\exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\}\right) \\ &= \frac{1}{v}(-\ln v)^{\theta-1}[((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-2}(\frac{1}{\theta}-1)\theta(-\ln u)^{\theta-1}\left(-\frac{1}{u}\right) \\ &\quad \cdot \exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\} \\ &\quad +((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-1}(\frac{1}{\theta}((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-1}\theta(-\ln u)^{\theta-1}\frac{1}{u}) \\ &\quad \cdot \exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\}] \\ &= \frac{1}{v}(-\ln v)^{\theta-1}\exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\} \end{aligned}$$

$$
\begin{aligned}
& [((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-2}(1-\theta)(-\ln u)^{\theta-1}\left(-\frac{1}{u}\right) \\
& +((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-2}((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}(-\ln u)^{\theta-1}\frac{1}{u}] \\
= \quad & \frac{1}{v}\frac{1}{u}(-\ln v)^{\theta-1}(-\ln u)^{\theta-1}\exp\{-((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\} \\
& \cdot((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}-2}\left(\theta - 1 + ((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}}\right). \qquad \square
\end{aligned}
$$

For the likelihood-depth we also need $h'(\theta, z) = \frac{\partial}{\partial\theta}\ln f_\theta(z)$. Before we calculate this, we introduce a notation for some clearer view. In the following let be $x := -\ln u$ ($\Leftrightarrow u = e^{-x}$) and $y := -\ln v$ ($\Leftrightarrow v = e^{-y}$).

Lemma 4.20. *With the previously defined shortcuts, we receive for* $z := (x, y)$

$$
\begin{aligned}
h'(\theta, z) \quad = \quad & \ln x + \ln y + \left((x^\theta + y^\theta)^{\frac{1}{\theta}} - 1\right)\frac{1}{\theta^2}\ln(x^\theta + y^\theta) \\
& + \left(\frac{1}{\theta} - \frac{(x^\theta + y^\theta)^{\frac{1}{\theta}}}{\theta} - 2\right)\frac{x^\theta \ln x + y^\theta \ln y}{x^\theta + y^\theta} \\
& + \frac{1 + (x^\theta + y^\theta)^{\frac{1}{\theta}}\left(-\frac{\ln(x^\theta + y^\theta)}{\theta^2} + \frac{x^\theta \ln(x) + y^\theta \ln(y)}{\theta(x^\theta + y^\theta)}\right)}{\theta - 1 + (x^\theta + y^\theta)^{\frac{1}{\theta}}}.
\end{aligned}
$$

Proof: According to definition, $h'(\theta, (u, v)) = \frac{\partial}{\partial\theta}\ln(f_\theta(u, v)) = \frac{\partial}{\partial\theta}\ln(f_\theta(e^{-x}, e^{-y}))$. Lemma 4.19 yields

$$
\begin{aligned}
& \frac{\partial}{\partial\theta}\ln(f_\theta(z)) = \frac{\partial}{\partial\theta}\ln\left(e^x e^y x^{\theta-1} y^{\theta-1}\exp\{-(x^\theta + y^\theta)^{\frac{1}{\theta}}\}(x^\theta + y^\theta)^{\frac{1}{\theta}-2}((\theta - 1) + (x^\theta + y^\theta)^{\frac{1}{\theta}})\right) \\
& = \quad \frac{\partial}{\partial\theta}\{x + y + (\theta - 1)\ln(x) + (\theta - 1)\ln(y) - (x^\theta + y^\theta)^{\frac{1}{\theta}} \\
& \qquad +(\frac{1}{\theta} - 2)\ln(x^\theta + y^\theta) + \ln(\theta - 1 + (x^\theta + y^\theta)^{\frac{1}{\theta}})\}.
\end{aligned}
$$

We calculate the derivatives for the individual summands. Here we use that for a differentiable function g, $g > 0$, $\frac{\partial}{\partial t}(g(t)^{\frac{1}{t}}) = g(t)^{\frac{1}{t}}\left(-\frac{\ln g(t)}{t^2} + \frac{g'(t)}{t g(t)}\right)$ holds. For the first four summands we obtain

$$
\frac{\partial}{\partial\theta}(x + y + (\theta - 1)\ln(x) + (\theta - 1)\ln(y)) = 0 + \ln(x) + \ln(y),
$$

for the next two

$$
\begin{aligned}
& \frac{\partial}{\partial\theta}((x^\theta + y^\theta)^{\frac{1}{\theta}}) = (x^\theta + y^\theta)^{\frac{1}{\theta}}\left(-\frac{\ln(x^\theta + y^\theta)}{\theta^2} + \frac{x^\theta \ln(x) + y^\theta \ln(y)}{\theta(x^\theta + y^\theta)}\right) \\
& =: T_1(\theta), \\
& \frac{\partial}{\partial\theta}\left((\frac{1}{\theta} - 2)\ln(x^\theta + y^\theta)\right) = -\frac{1}{\theta^2}\ln(x^\theta + y^\theta) + \left(\frac{1}{\theta} - 2\right)\frac{x^\theta \ln(x) + y^\theta \ln(y)}{(x^\theta + y^\theta)}
\end{aligned}
$$

and finally

$$\frac{\partial}{\partial\theta}\ln(\theta-1+(x^\theta+y^\theta)^{\frac{1}{\theta}}) = \frac{1+T_1(\theta)}{\theta-1+(x^\theta+y^\theta)^{\frac{1}{\theta}}}.$$

In total we receive

$$\begin{aligned}
h'(\theta,(x,y)) &= \ln(x)+\ln(y) \\
&\quad -(x^\theta+y^\theta)^{\frac{1}{\theta}}\left(-\frac{\ln(x^\theta+y^\theta)}{\theta^2}+\frac{x^\theta\ln(x)+y^\theta\ln(y)}{\theta(x^\theta+y^\theta)}\right) \\
&\quad -\frac{1}{\theta^2}\ln(x^\theta+y^\theta)+\left(\frac{1}{\theta}-2\right)\frac{x^\theta\ln(x)+y^\theta\ln(y)}{(x^\theta+y^\theta)} \\
&\quad +\frac{1+(x^\theta+y^\theta)^{\frac{1}{\theta}}\left(-\frac{\ln(x^\theta+y^\theta)}{\theta^2}+\frac{x^\theta\ln(x)+y^\theta\ln(y)}{\theta(x^\theta+y^\theta)}\right)}{\theta-1+(x^\theta+y^\theta)^{\frac{1}{\theta}}} \\
&= \ln x+\ln y+\frac{(x^\theta+y^\theta)^{\frac{1}{\theta}}\ln(x^\theta+y^\theta)}{\theta^2}+(1-(x^\theta+y^\theta)^{\frac{1}{\theta}})\frac{x^\theta\ln x+y^\theta\ln y}{\theta(x^\theta+y^\theta)} \\
&\quad -2\frac{x^\theta\ln(x)+y^\theta\ln(y)}{(x^\theta+y^\theta)}-\frac{1}{\theta^2}\ln(x^\theta+y^\theta) \\
&\quad +\frac{1+(x^\theta+y^\theta)^{\frac{1}{\theta}}\left(-\frac{\ln(x^\theta+y^\theta)}{\theta^2}+\frac{x^\theta\ln(x)+y^\theta\ln(y)}{\theta(x^\theta+y^\theta)}\right)}{\theta-1+(x^\theta+y^\theta)^{\frac{1}{\theta}}} \\
&= \ln x+\ln y+\left((x^\theta+y^\theta)^{\frac{1}{\theta}}-1\right)\frac{1}{\theta^2}\ln(x^\theta+y^\theta) \\
&\quad +\left(\frac{1}{\theta}-\frac{(x^\theta+y^\theta)^{\frac{1}{\theta}}}{\theta}-2\right)\frac{x^\theta\ln x+y^\theta\ln y}{x^\theta+y^\theta} \\
&\quad +\frac{1+(x^\theta+y^\theta)^{\frac{1}{\theta}}\left(-\frac{\ln(x^\theta+y^\theta)}{\theta^2}+\frac{x^\theta\ln(x)+y^\theta\ln(y)}{\theta(x^\theta+y^\theta)}\right)}{\theta-1+(x^\theta+y^\theta)^{\frac{1}{\theta}}}. \qquad \square
\end{aligned}$$

If we choose the estimator $\tilde{\theta}$ of the parameter $\theta \in [1,\infty)$ as the one with maximum likelihood-depth in the data, then $\tilde{\theta}$ is bigger than the real parameter. Here too, the maximum likelihood-depth estimator is biased. We simulated data with different parameter θ, each 1000 times 100 data and determined the parameter with maximum depth. The results are displayed in Figure 4.23, where the mean of the parameters with maximum depth and the standard deviation in every situation is showed. The graphics suggests that the relation between the parameter with maximum depth and the real parameter is linear.

The probability is determined, that one data lies inside the region where $h'(\theta,\cdot) = \frac{\partial}{\partial\theta}\ln f_\theta(\cdot) \geq 0$. With the same notations as in Section 2.1 we get

$$p_\theta := P_\theta(T^\theta_{pos}) = \int\int 1_{T^\theta_{pos}}(u,v) f_\theta(u,v)\,du\,dv.$$

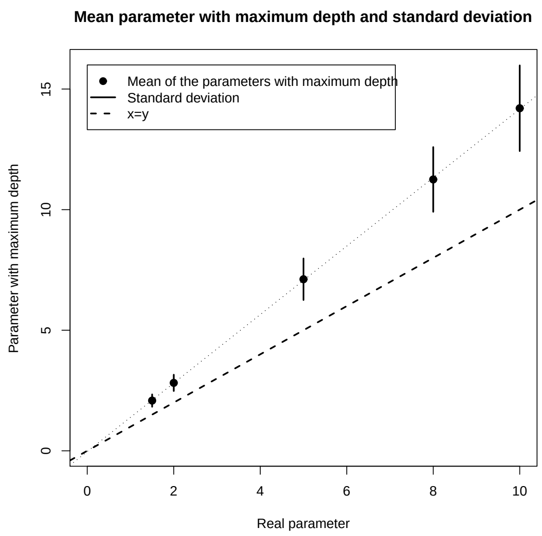

Figure 4.23.: Mean and standard deviation of the parameter with maximum depth for the Gumbel copula with different θ.

To calculate p_θ the zeros of $h'(\theta,\cdot)$ are needed (they are the boundaries of T^θ_{pos} and thereby the limits of the integral). We could not find explicit algebraic expressions for them, so we had a look at the contour-plot of $h'(\theta,u,v)$ for fixed θ and variable $u,v\in[0,1]$. We are led to the assumption, that there exists at most 2 different zeros for each u and that the roots are symmetric to the line $u=v$. The symmetry follows from the symmetry of $h'(\theta,u,v)$ as seen in Lemma 4.20. The graphics in Figures 4.24 and 4.25 show the contour-plots of $h'(\theta,u,v)$ for some values of θ.

Because of the complexity of $h'(\theta,u,v)$, the zeros were only evaluated numerically. As a consequence the integral can also be numerically approximated, only. The points $(0,0)$ and $(1,1)$ are singular points of the density function, thus the evaluation of the integral is made with higher accuracy here. In the following let $r_\theta(u)$ denote the function such that $h'(\theta,(u,r_\theta(u)))=0$ for all $u\in[0,1]$. We have

$$\begin{aligned}
p_\theta &= P_\theta(T^\theta_{pos}) = \int\int 1_{T^{pos}_\theta}(u,v) f_\theta(u,v)dvdu = 2\cdot\int_0^1\int_{r_\theta(u)}^u f_\theta(u,v)dvdu\\
&= 2\int_0^1\int_0^u f_\theta(u,v)dvdu - 2\int_0^1\int_0^{r_\theta(u)} f_\theta(u,v)dvdu\\
&= 2\frac{1}{2} - 2\int_0^1\int_0^{r_\theta(u)} f_\theta(u,v)dvdu\\
&= 1-2\int_0^1\int_0^{r_\theta(u)} f_\theta(u,v)dvdu,
\end{aligned}$$

where the second identity follows from the symmetry of f_θ. Only the remaining integral has to be calculated in the way described above. This leads to the results shown in Table

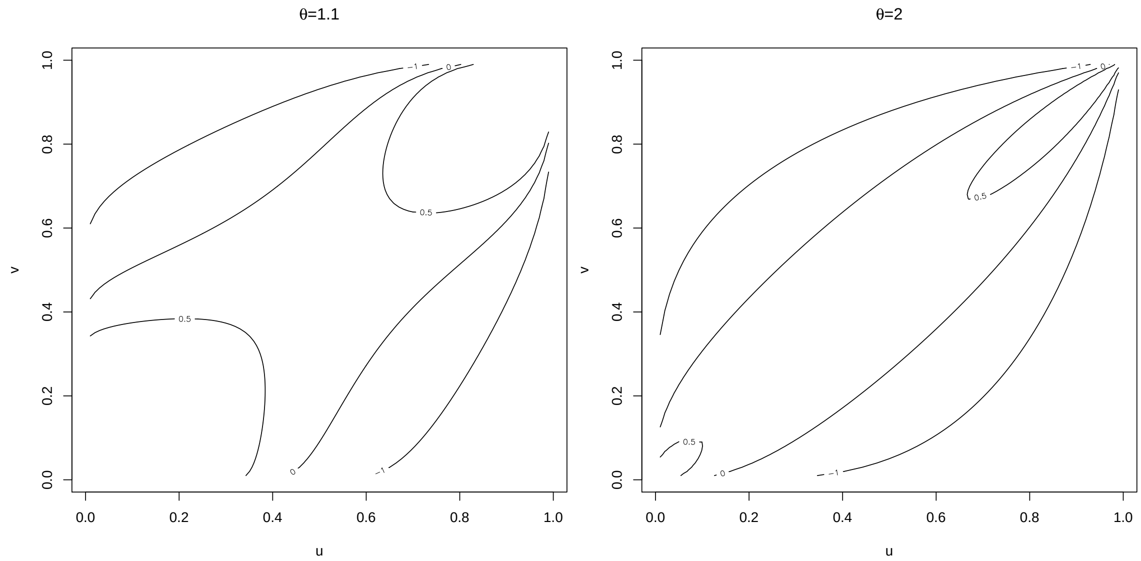

Figure 4.24.: Contour plot of $h'(\theta,(u,v))$ for the Gumbel copula, $\theta = 1.1$ and $\theta = 2$.

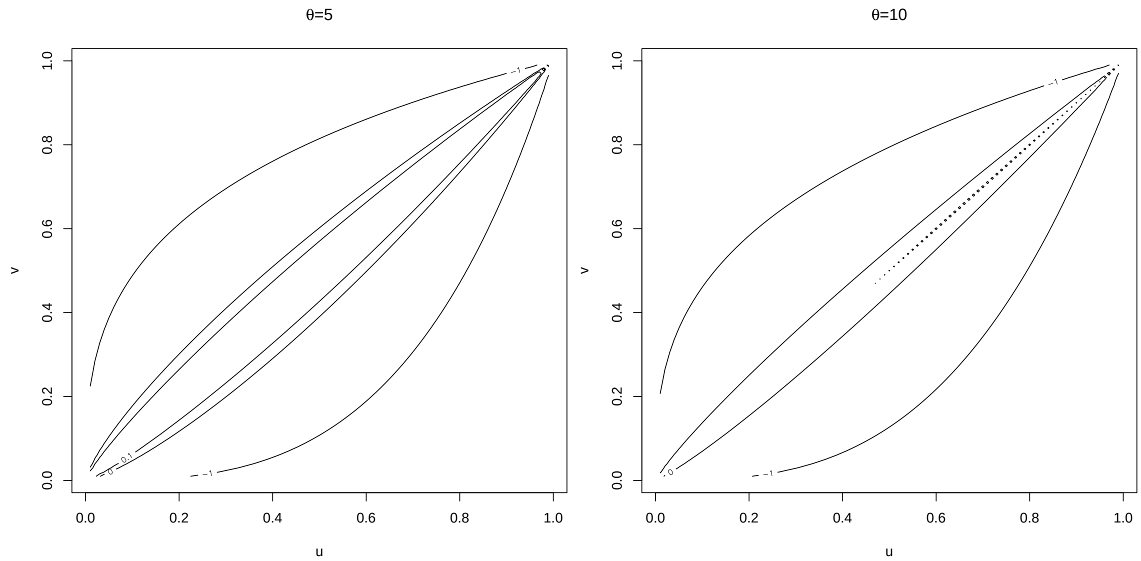

Figure 4.25.: Contour plot of $h'(\theta,(u,v))$ for the Gumbel copula, $\theta = 5$ and $\theta = 10$.

4.10. Alternatively, p_θ could also be calculated by generating a large number of points $z = (u,v)$ from distribution with θ and then determining the part of points for that $\frac{\partial}{\partial\theta} \ln f_\theta(u,v) \geq 0$ is satisfied.

Table 4.10.: p_θ for the Gumbel copula.

θ	p_θ	θ	p_θ	θ	p_θ	θ	p_θ	θ	p_θ
1	0.564399	1.1	0.591775	1.25	0.614965	1.5	0.632269	1.75	0.639433
2	0.642906	2.5	0.645838	3	0.646917	3.5	0.647388	4	0.647617
4.5	0.647738	5	0.647806	5.5	0.647845	6	0.647869	6.5	0.647884
7	0.647893	7.5	0.6479	8	0.647904	8.5	0.647906	9	0.647908
9.5	0.647909	10	0.64791	15	0.647911	20	0.64791	25	0.647909
30	0.647908	35	0.647908	40	0.647907	45	0.647906	50	0.647905
55	0.647904	60	0.647901	65	0.647894	70	0.647882	75	0.647859
80	0.647903	85	0.647766	90	0.647684	95	0.64757	100	0.647418

We see that $p_\theta = P_\theta(T^\theta_{pos})$ is clearly not one half, what shows the biasedness of the maximum likelihood-depth estimator.

The next step is to calculate the shift $s(\theta)$ such that $P_\theta(T^{s(\theta)}_{pos}) = 0.5$. Values are received by first calculating the zeros of $h'(s(\theta), u, v)$ for $s(\theta)$ in equidistant points in $(0, 1)$, then interpolating the zeros to find that θ such that

$$P_\theta(T^{s(\theta)}_{pos}) = 1 - 2\int_0^1 \int_0^{r_\theta(u)} f_\theta(u, v)dvdu \approx 0.5,$$

what means solving

$$\int_0^1 (e^{-((-\ln u)^\theta + (-\ln r_{s(\theta)}(u))^\theta)^{\frac{1}{\theta}}} (\ln u)^{\theta-1}((-\ln u)^\theta + (-\ln r_{s(\theta)}(u))^\theta)^{\frac{1}{\theta}-1}\frac{1}{u})du \approx 0.25.$$

Here we used the bisection method. The results can be found in Table 4.11.

Table 4.11.: Alduration of the parameter with maximum depth.

$s(\theta)$	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
θ	1.133	1.455	1.798	2.148	2.50	2.854	3.208	3.563	3.919	4.274
$s(\theta)$	7.0	8.0	9.0	10.0	15.0	20.0	30.0	40.0	50.0	
θ	4.985	5.696	6.408	7.119	10.677	14.232	21.351	28.467	35.583	

A line of best fit through these points is $0.015 + 0.71s(\theta)$. The graphic in Figure 4.26 shows the points $(s(\theta), \theta)$ and the line of best fit.

Thus, a new estimator can be chosen as

$$\hat{\theta}(z) = \arg\max_{\theta>1} d_T(\theta, z) \cdot 0.71 + 0.015. \tag{4.2}$$

This new estimator was compared to the maximum likelihood estimator (MLE) for datasets with different θ (1000 times, 100 data each). Table 4.12 shows the real parameter (θ), the mean of the new estimator ($\hat{\theta}$) and the mean of the MLE.

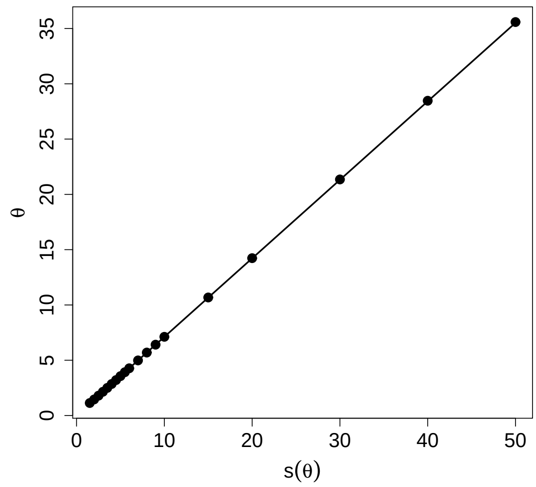

Figure 4.26.: The points $(s(\theta), \theta)$ and the line of best fit.

Table 4.12.: Comparison of $\hat{\theta}$ and MLE for samples with different θ, $N = 100$, 1000 repetitions each.

θ	$\hat{\theta}$	MLE
1.1	1.050	1.137
1.5	1.497	1.504
2.0	2.041	2.020
3.0	3.030	3.023
4.0	4.050	4.031
5.0	5.048	5.028
10.0	10.104	10.109

The table indicates that for simulated datasets without contamination the new estimator is nearly as good as the MLE. The next step is to compare the strength of both for datasets with ε-contamination. Therefore two samples were mixed, a bigger one with parameter θ and a smaller (the contaminated data) with parameter θ_1. Again 1000 samples each were created. For the results see Table 4.13. Here the column $\varepsilon \cdot 100\%$ gives the percentage of the data with parameter θ_1.

Table 4.13.: Comparison of $\hat{\theta}$ and MLE for ε-contaminated data with θ_1, $N = 100$, 1000 repetitions each.

θ	θ_1	$\varepsilon \cdot 100\%$	$\hat{\theta}$	MLE
1.1	10.0	5	1.14	1.15
1.1	10.0	10	1.19	1.18
1.1	100.0	5	1.13	1.64
1.1	100.0	10	1.25	37.78
2.0	10.0	5	2.15	2.07
2.0	1.1	5	1.97	1.94
5.0	1.1	5	4.87	4.28
5.0	20.0	5	5.39	5.23
10.0	1.1	5	9.54	7.17
10.0	1.1	10	9.02	5.47
10.0	2.0	10	9.16	7.14

We also get good results for datasets with contamination for the new estimator, most times even better than for the MLE. Hence,

$$\hat{\theta}(z) = \arg\max_{\theta>1} d_T(\theta, z) \cdot 0.71 + 0.015$$

can be used as an estimator for the parameter of the Gumbel copula.

To make statements about the asymptotic behavior of the estimator for the parameter of the Gumbel copula, we table the simulated mean squared errors of the estimator for growing sample size for data with $\theta = 2$. The results are displayed in Figure 4.27 on the left, in comparison to the MLE. They provide the assumption, that also for the Gumbel copula we receive a consistent estimator, as the mean squared error is tending to zero. To prove consistency, we take again a look on the contour-plot of the zeros of $h'(\theta, \cdot)$. This time we display the zeros for growing θ, see Figure 4.27 on the right. Because the region where $h'(\theta, \cdot)$ is positive is between the zeros, the graphic shows that $\lambda_N^+(\theta, z_*) := \sharp\{n; h'(\theta, z_n) \geq 0\}$ and $\lambda_{\theta_0}^+(\theta) := P_{\theta_0}(h'(\theta, Z) \geq 0)$ are (strictly) decreasing. This is because for $\theta > \theta'$, they lead to the assumption that $\{z; h'(\theta, z) \geq 0\} \subset \{z; h'(\theta', z) \geq 0\}$. With the same arguments we can assume that $\lambda_N^-(\theta, z_*) := \sharp\{n; h'(\theta, z_n) \leq 0\}$ and $\lambda_{\theta_0}^-(\theta) := P_{\theta_0}(h'(\theta, Z) \leq 0)$ are (strictly) increasing and with the strong law of large numbers we also have $\lambda_N^\pm(\theta, Z_*) \to \lambda_{\theta_0}^\pm(\theta)$. The requirements of Proposition 2.7 on page 11 are fulfilled and we would get a strongly consistent estimator for $s(\theta)$ by $\tilde{\theta} = \arg\max d_T(\theta, Z_*)$. As we can further assume, that s^{-1} is continuous, we get that $\hat{\theta}$ given by (4.2) is a strongly consistent estimator for θ.

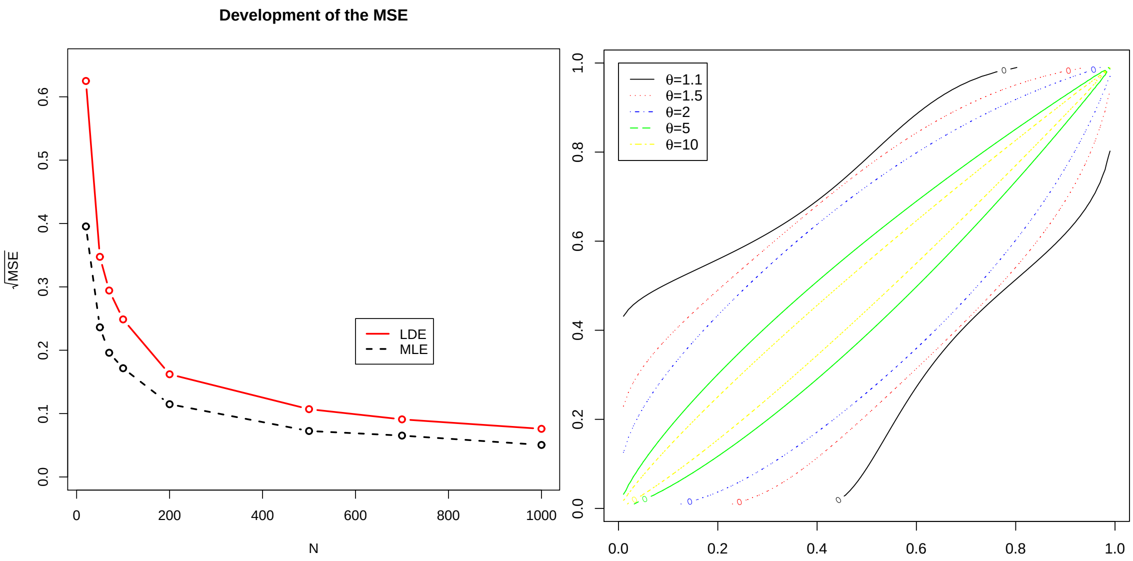

Figure 4.27.: Asymptotic behavior of $\hat{\theta}$ (left-hand side) and zeros of $h'(\theta, \cdot)$ (on the right).

4.4.1. Data with unknown margins

So far we assumed that the data are coming from a distribution, where both marginal distributions are uniform on the interval $[0, 1]$. In practice, however, the marginal distribution must be estimated and the data have to be transformed, so that the distribution of the transformed data is uniform on $[0, 1]$. The marginal distributions F and G can be approximated for example by the cumulative empirical distribution function. The estimated functions shall be denoted by $\hat{F}$ and $\hat{G}$. Then the data are transformed to $\tilde{z}_i = (\hat{F}(x_i), \hat{G}(y_i)) \in [0, 1]$ for $i = 1, \ldots, N$. For this transformed dataset we can have a look at the belonging copula. We have to decide for one family and then (in case we have chosen the Gumbel copula) use the methods from above to estimate the parameter. Table 4.14 shows some results for the estimation as comparison of the MLE and the new estimator $\hat{\theta}$, where we simulated data with different margins and dependence structure given by the Gumbel copula with $\theta = 2$. Even for data without contamination the new estimator achieves most times slightly better results than the MLE. We simulated 1000 datasets with 100 data each, t_n denotes the t-distribution with $df = n$, $\mathcal{N}(0, 1)$ the normal distribution with mean 0 and variance 1, $\mathcal{E}(2)$ the exponential distribution with rate 2 and χ^2_2 the χ^2-distribution with $df = 2$. The graphics in Figure 4.28 show the original data set for margins t_8-distributed and the transformed data.

Thus, the estimation of the margins has no influence on the estimator based on likelihood-depth.

Table 4.14.: Estimation for unknown marginal distribution, $N = 100$, 1000 repetitions each.

F	G	$\hat{\theta}$	MLE
t_3	t_3	2.012	1.977
t_8	t_8	2.027	1.974
t_3	χ^2_2	2.019	1.987
$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	2.016	1.974
$\mathcal{N}(0,1)$	$\mathcal{E}(2)$	2.007	1.971

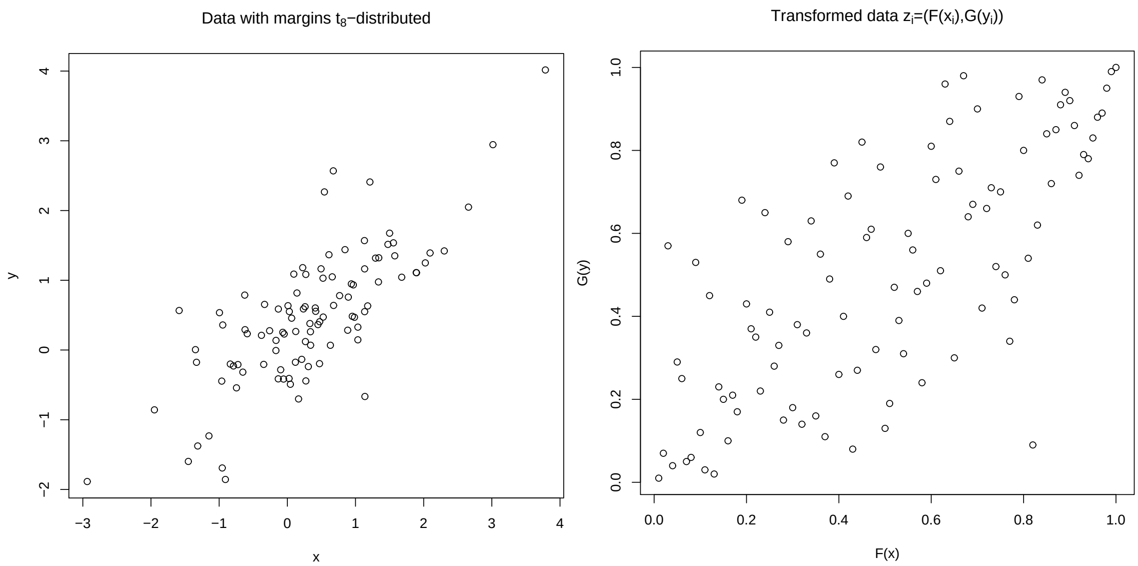

Figure 4.28.: Data with margins t_8-distributed and Gumbel Copula with $\theta = 2$.

4.5. Tests and confidence intervals for the parameter of the Gumbel copula

In this section we construct tests for the hypotheses $H_0 : \theta \leq \theta_0$, $H_0 : \theta \geq \theta_0$ and $H_0 : \theta = \theta_0$, based on likelihood-depth. Therefore we We already showed in the last section about estimation for the Gumbel copula that $p_\theta = P_\theta(T^\theta_{pos}) \neq \frac{1}{2}$ and $s(\theta) > \theta$, see Table 4.10 on page 165 and Table 4.11 on page 165, respectively. We define the test statistic as described in Lemma 2.14, page 17.

Definition 4.21. *Let be $\theta > 1$. The test statistic for testing hypotheses of the parameter*

θ of the Gumbel copula is defined as

$$T(\theta, z_*) := \sqrt{N} \frac{d_S(\theta, z_*) - 2p_\theta(1-p_\theta)}{2\sqrt{p_\theta(1-p_\theta)(1-2p_\theta)^2}}.$$

For values for p_θ see Table 4.10 on page 165. We use Corollary 2.17 on page 19 directly and receive

Corollary 4.22. *The test*

$$\varphi_{\theta_0}^{0,\leq} = 1_{\{\sup_{\theta \leq \theta_0} T(\theta, \cdot) < \Phi^{-1}(\alpha)\}}$$

is an asymptotic α-level test for the hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

The approximated power-function for finite sample-size is displayed in Figure 4.29 for the examples $\theta_0 = 1.5, 2, 5$ and $\theta_0 = 10$. For each θ we simulate 1000 times 100 data each. In the graphics we see, that the test does slightly not keep the level. This should become better, if we increase the sample-size.

In a next step we want to study the robustness of this new test in ε-contaminated data. The graphics in Figure 4.30 show the behavior of the power for datasets with $\theta_0 = 2$ and 5, when the data is contaminated (5% of each dataset are data with distribution $\theta_1 = 1.1$). In Figure 4.31 contamination with a parameter $\theta_1 \geq \theta_0$ is considered, again 5% of each dataset are data with distribution $\theta_1 = 10$. The graphics demonstrate that for contaminated data with a distribution depending on $\theta_1 > \theta_0$, the test does not keep the level for $N = 100$ data, but otherwise the power-function is not affected very much by the contamination. What happens to the power of the test, if we increase the parameter θ_1 of the contaminated data (10%) to infinity? The graphics in Figure 4.32 show the results for $\theta_0 = 2$ and $\theta_1 = 10^3, 10^6$ and 10^9. We see, if we fix the ratio of contamination with distribution θ_1 at 10%, then the frequency of rejections of $H_0 : \theta \leq \theta_0$ does not tend to 1 for $\theta_1 \to \infty$, which indicates that the test is quite robust. If the number of outliers is increasing, the ratio of rejections is growing up to one in θ_0. The graphics in Figure 4.33 show the case for $\theta_0 = 2$, $\theta_1 = 10^6$. We see that $H_0 : \theta \leq \theta_0 = 2$ is always rejected for a ratio of “contamination” of 50%.

Now we take a look at realizations of variables $Z_i = (X_i, Y_i)$, $i = 1, \ldots, N$, with margins F and G (unknown) and dependence structure given by the copula C_θ, where the family of the copula is supposed to be known. The margins are unknown and have to be estimated. For notation and more explanation see the beginning of Section 4.4.1. We test $H_0 : \theta \leq \theta_0$ for the transformed dataset $\tilde{z}_* = ((\hat{F}(x_1), \hat{G}(y_1)), \ldots, (\hat{F}(x_N), \hat{G}(y_N)))$, where $\hat{F}$ and $\hat{G}$ shall denote the estimators of F and G. Figure 4.34 shows exemplary the results for data with dependence given by the Gumbel copula with $\theta = 2$ and different margins, again for 1000 datasets with 100 data each. It indicates that the estimation of the marginal distribution has almost no influence on the power-function. Compare it to Figure 4.29, $\theta = 2$.

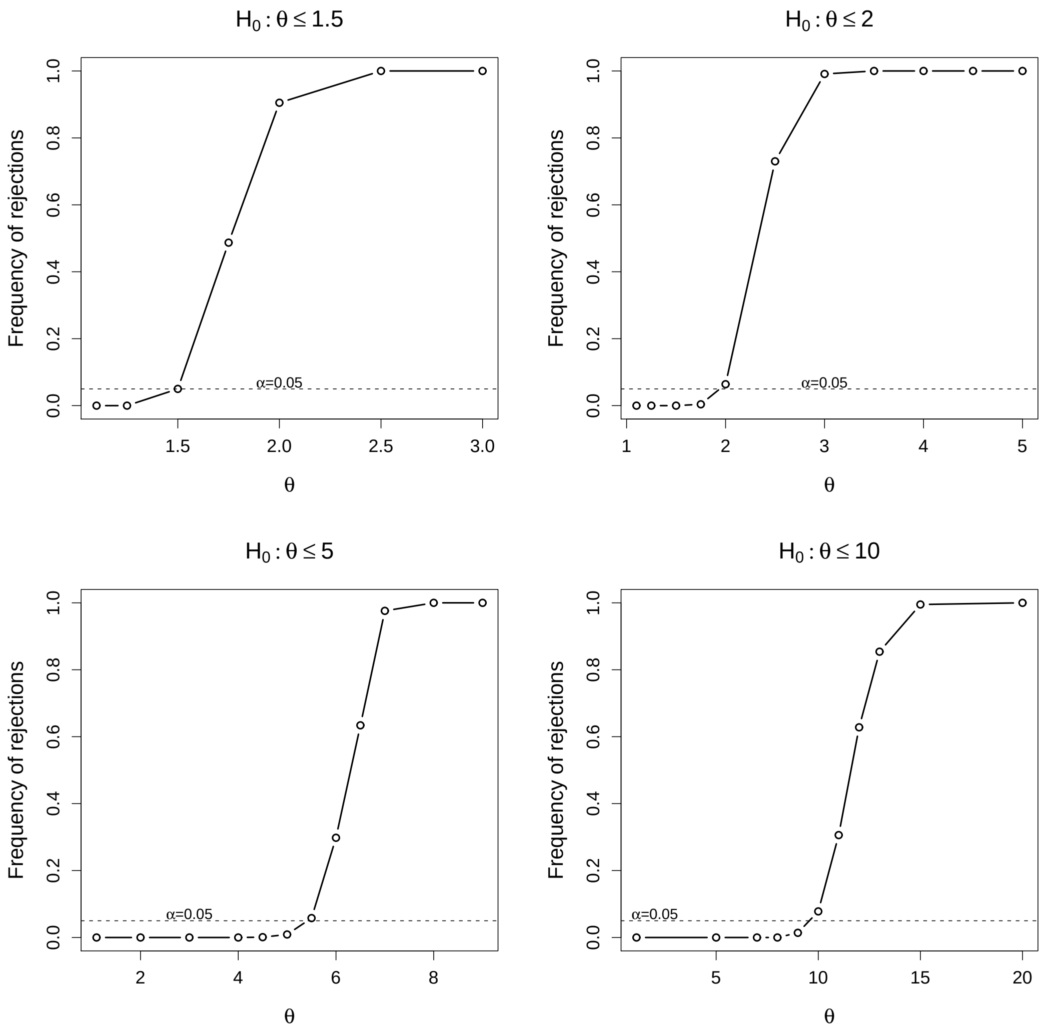

Figure 4.29.: Power of the test for $H_0 : \theta \leq \theta_0$ for Gumbel copulas.

Since we showed in the last section about estimation that $s(\theta) > \theta$, we use Definition 2.18 on page 19 and Definition 2.19 on page 20, to construct a test for the hypothesis $H_0 : \theta \geq \theta_0$. Therefore, we are looking for an approximation of

$$c^1_\alpha(\theta_0) = \max\{\theta \geq 1; \lim_{N\to\infty} P_{\theta_0}(T(\theta, Z_*) < \Phi^{-1}(\alpha)) \leq \alpha\}$$

in the following. As we have no explicit terms for $p_{(\cdot),\theta}$ and $p_{\theta,(\cdot)}$, we can not use Lemma 2.25 on page 23 here. How the power would look like, if we would not improve it, can be seen in the two graphics in Figure 4.35, which show the estimated power-function of the test for $\theta_0 = 2$ and $\theta_0 = 5$. As expected it is very bad.

Like we did in the case of the two-dimensional normal distribution, we simulate 1000 times 1000 data for every $\theta \in \{1.25, 1.5, 2, 2.5, 3, \ldots, 10\}$ and determine the maximum parameter θ', such that $T(\theta', z_*) < \Phi^{-1}(\alpha)$ in maximum $\alpha \cdot 100\%$ of the cases. Table 4.15 illustrates the results of the estimation of c^1_α.

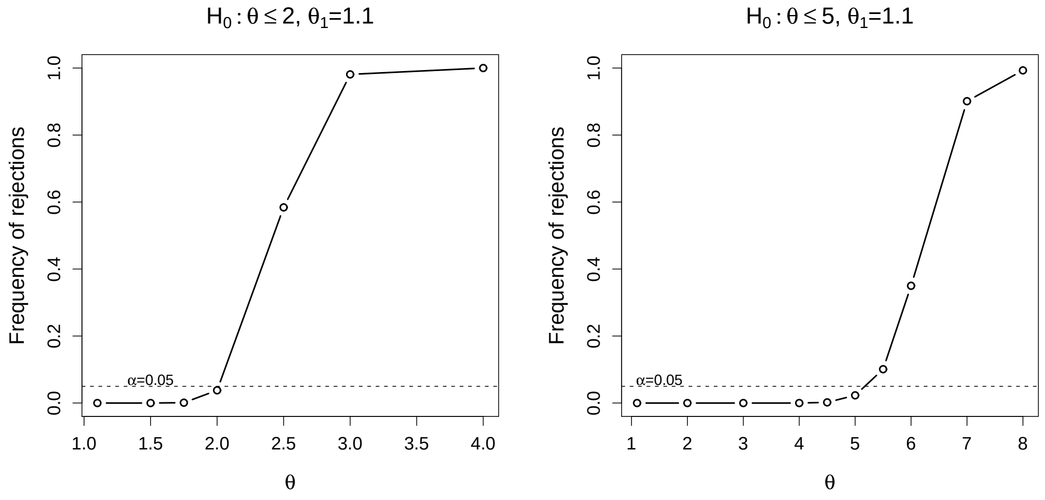

Figure 4.30.: Test for $H_0 : \theta \leq \theta_0$ for data with ε-contamination ($\theta_1 = 1.1$, $\varepsilon = 0.05$).

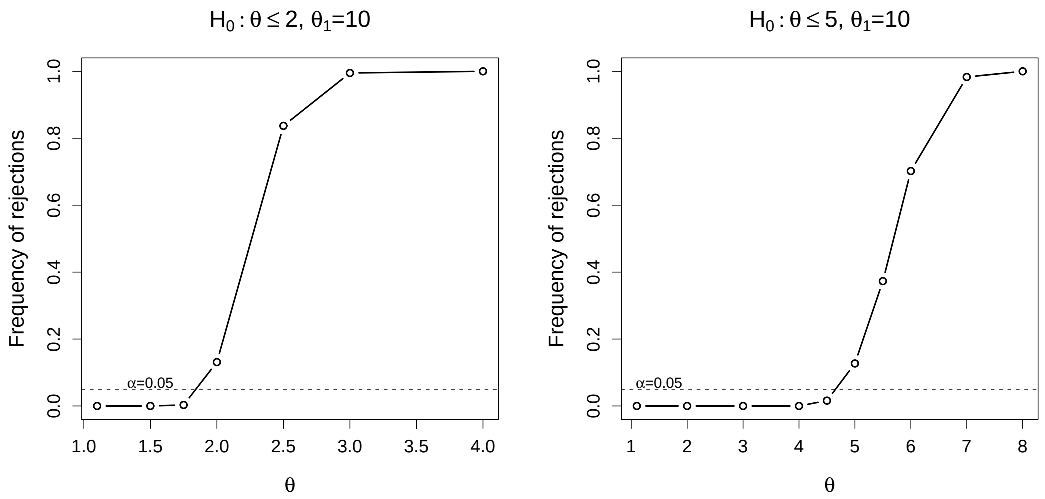

Figure 4.31.: Test for $H_0 : \theta \leq \theta_0$ for data with ε-contamination ($\theta_1 = 10$, $\varepsilon = 0.05$).

Table 4.15.: Values of $c^1_{\alpha=0.05}(\theta)$.

θ_0	1.25	1.5	2	2.5	3	3.5	4	4.5	5	6	7	8	9	10
$\hat{c}^1_{\alpha=0.05}(\theta_0)$	2.5	3	4	5	6	7	8	9	10	12	14	16	18	20

This leads to

Proposition 4.23. *For the Gumbel copula it holds* $c^1_{\alpha=0.05}(\theta) = 2\theta$.

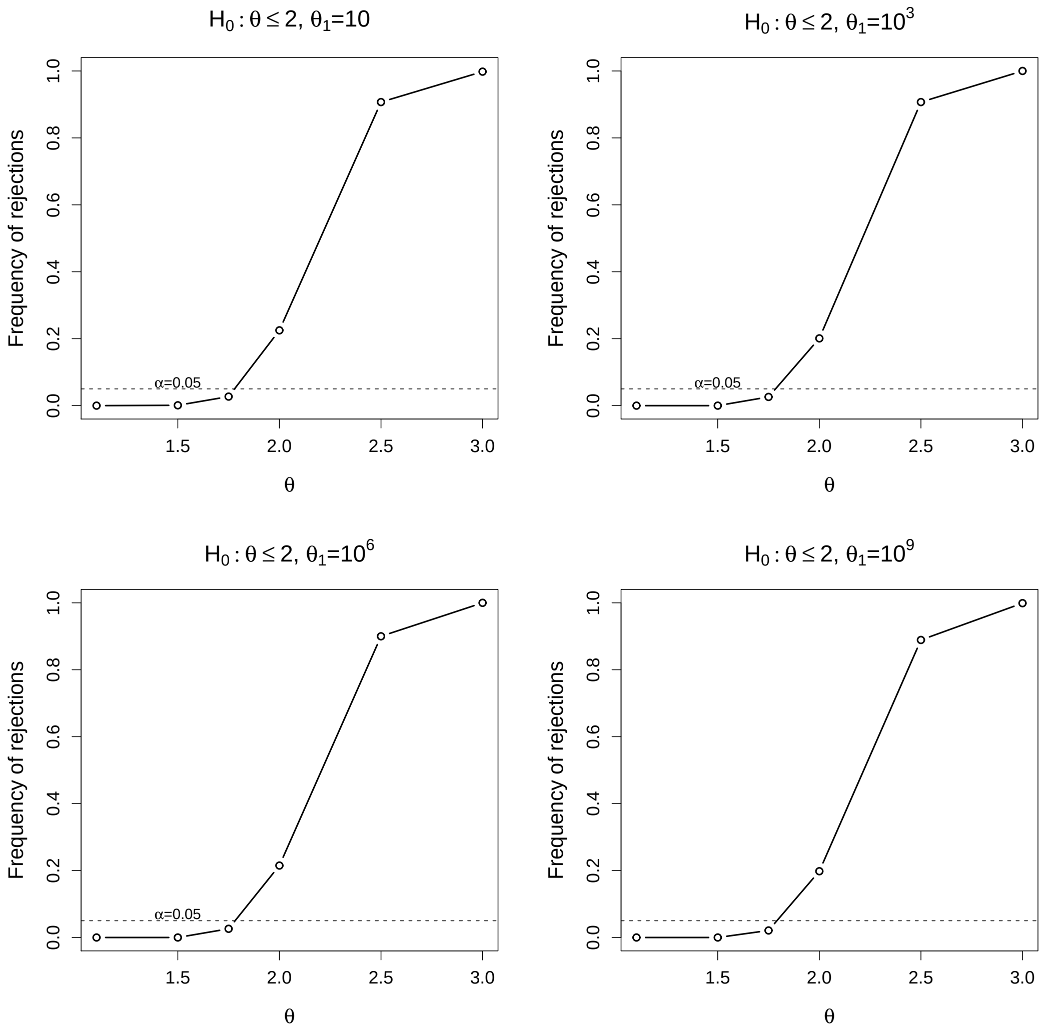

Figure 4.32.: Power for $H_0 : \theta \leq \theta_0$ for ε-contamination with increasing $\theta_1 = 10, 10^3, 10^6, 10^9$, $\varepsilon = 0.1$.

The proof of this Proposition is an open problem.

Hence for c_α^1 the assumptions of Theorem 2.20 on page 20 seem to hold for $\alpha = 0.05$ and it can be used to create a new test.

Corollary 4.24. *Assume the conditions of Theorem 2.20 on page 20 to be true. The test*

$$\varphi_{\theta_0}^{\geq} = 1_{\{\sup_{\theta \geq 2\theta_0} T(\theta,\cdot) < \Phi^{-1}(\alpha)\}}$$

is an asymptotic 0.05*-level test for the hypothesis* $H_0 : \theta \geq \theta_0$ *against* $H_1 : \theta < \theta_0$*.*

For a test with a different level, c_α^1 has to be computed.

The power-function for the new test is estimated and shown for some examples in the graphics of Figure 4.36. They display the improvement of the power-function (compare

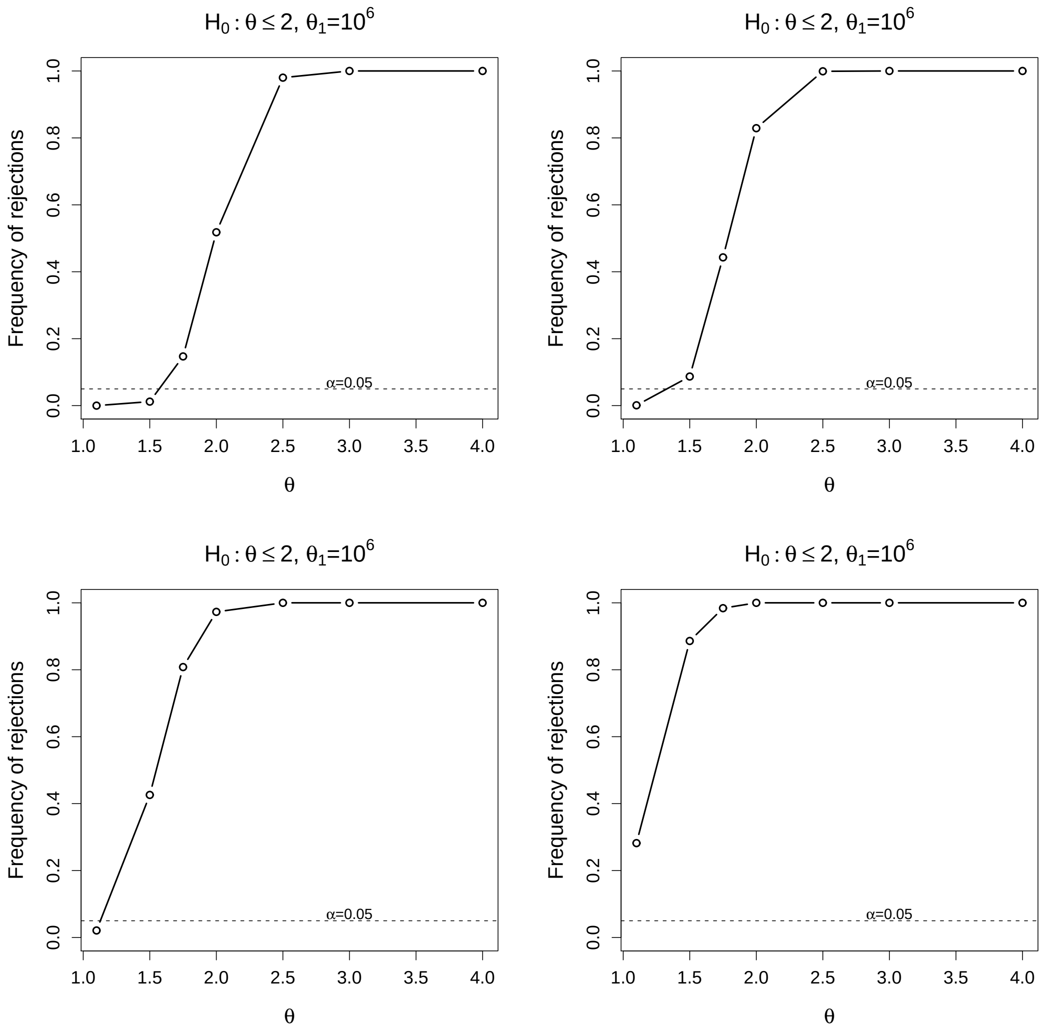

Figure 4.33.: $H_0 : \theta \leq \theta_0$, increasing ratio of contamination $20\% - 50\%$, $\theta_1 = 10^6$.

with Figure 4.35). The power of the test is also estimated for data with contamination. Figure 4.37 shows the results for $\theta_0 = 2$ and $\theta_0 = 5$ with 5% contaminated data with distribution $\theta_1 = 100$ (first row) and $\theta_1 = 1.1$ (second row). We see that the power is also good in case of the data with contamination.

Again we estimate the power-function for data $z_* = ((x_1, y_1), \ldots, (x_N, y_N))$ with unknown margins. First the margins are estimated with the empirical distribution function ($\hat{F}$ and $\hat{G}$), then the test is applied to the data ($u_i = \hat{F}(x_i), v_i = \hat{G}(y_i)$). For more details and notation see subsection 4.4.1. Some of the results for different margins and $\theta_0 = 2$ are shown in Figure 4.38. As before the graphics indicate, when compared with the ones of Figure 4.36 for known margins, that estimation of the margins has almost no influence on the power-function of the improved test.

Using the tests for $H_0 : \theta \leq \theta_0$ and $H_0 : \theta \geq \theta_0$ we can easily give a test for $H_0 : \theta = \theta_0$ and confidence intervals.

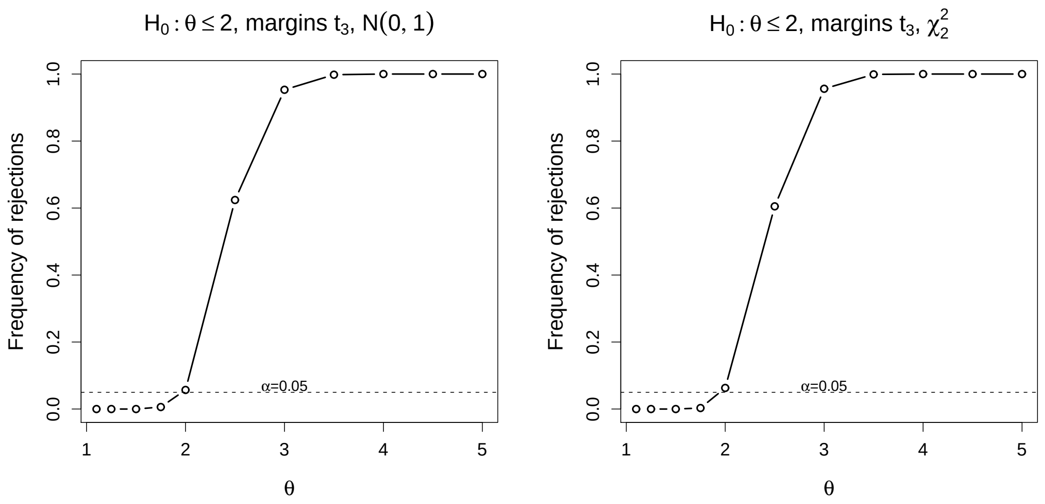

Figure 4.34.: Testing $H_0 : \theta \leq \theta_0$ for data with estimated margins.

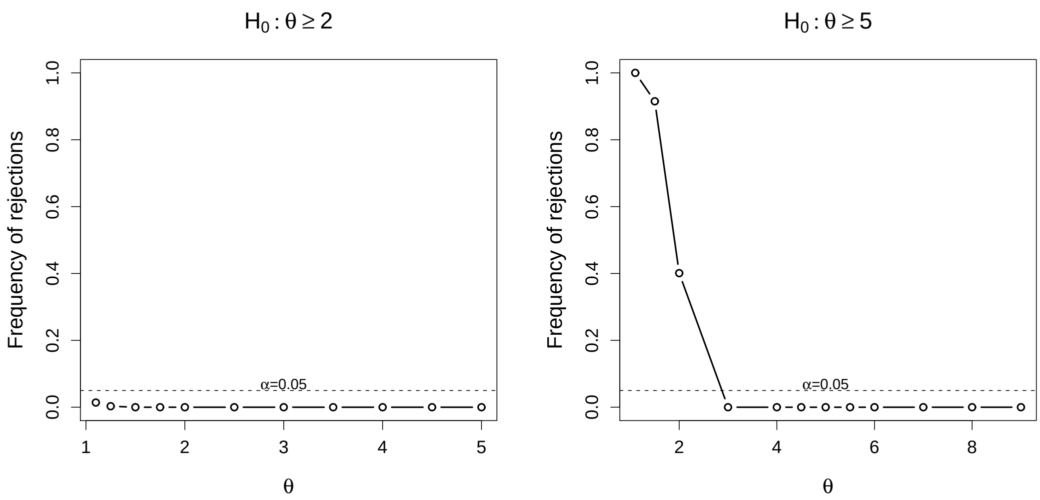

Figure 4.35.: Power of the uncorrected test for $H_0 : \theta \geq \theta_0$.

Corollary 4.25. *Assume the conditions of Theorem 2.20 on page 20 to be true. An asymptotic test with level α for the hypothesis $H_0 : \theta = \theta_0$ is given by*

$$\varphi^{=}_{\theta_0}(z_*) = \max(1_{\{T(\theta_0, z_*) < \Phi^{-1}(\frac{\alpha}{2})\}}(z_*), 1_{\{T(c_{\frac{\alpha}{2}}(\theta_0), z_*) < \Phi^{-1}(\frac{\alpha}{2})\}}(z_*)).$$

Consequently, an asymptotic γ-confidence interval with $\gamma = 1 - \alpha$ in data z_* is

$$\{\theta \geq 1; \varphi^{=}_{\theta}(z_*) = 0\}.$$

For the Gumbel copula we can not prove consistency, as we do not have explicit terms for $p_{\theta,\theta'}$. But analog to the considerations in the section about estimation, we take a look

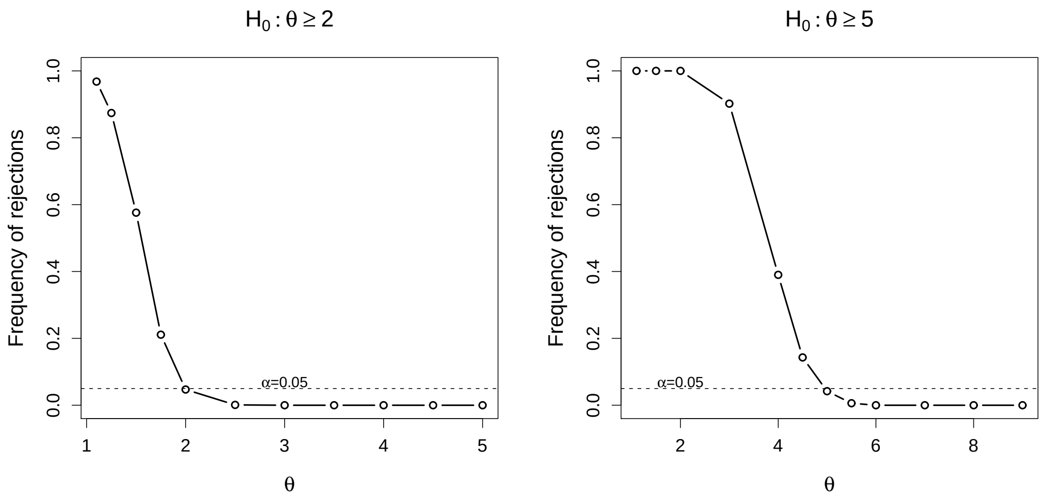

Figure 4.36.: Power of the improved test for $\theta \geq \theta_0$.

at the assumptions of Theorem 2.32, resp. Theorem 2.33. We already motivated in the end of Section 4.4 that we assume $p_{\theta_0,(\cdot)}$ to be strictly increasing. The graphics of Figure 4.39, showing the contours of different Gumbel densities and the plot of the zeros of h' in Figure 4.27, suggest that $p_{(\cdot),\theta}$ is strictly decreasing.

Further it is $c^1_\alpha(\cdot)$ strictly increasing, $c_\alpha(\theta) > \theta$ and $\frac{1}{2} < p_\theta < \frac{1}{2} + \frac{1}{\sqrt{8}} \approx 0.853$, see Table 4.10. This leads to $\varphi^{=}_{\theta_0}, \varphi^{\leq}_{\theta_0}$ and $\varphi^{\leq}_{\theta_0}$ being consistent tests.

4.6. Open problems

We introduced the Gaussian and the Gumbel copula and developed estimators and tests for the unknown parameters of these. In both cases the problem occurs that p_ρ resp. p_θ can not be determined explicitly. Therefore also some of the conditions to prove consistency of the estimators could only be made reasonable and were not proven. The proofs could be the topic of further studies. Considering the Gaussian copula, it would be of interest to prove that only for $\rho = 0$ there exist two solutions ρ' of $p_{\rho,\rho'}$. If this is the case, s^{-1} is unique for $\rho > 0$. Anyway we showed in simulations studies that the new estimators are nearly as good as existing standard methods for uncontaminated data. For contaminated and robust data we showed the robustness of the new estimators based on likelihood-depth. We also showed that the estimation of the margins in case of the Gumbel copula has no influence on the estimation of the parameter.

For the Gumbel copula we also made reasonable that the theorems for the consistency of the tests can be used. But the problem of the unknown explicit forms of $p_\theta, p_{\cdot,\theta}, p_{\theta,\cdot}$ leads to the problem, that c^1_α can only be estimated, the proof that it holds $c^1_\alpha(\theta) = 2\theta$ for $\alpha = 0.05$ is an open problem, see Proposition 4.23. For the Gaussian copula we do not have that $p_{\cdot,\rho}$ and $p_{\rho,\cdot}$ are monotone functions, so here the proof that the tests based

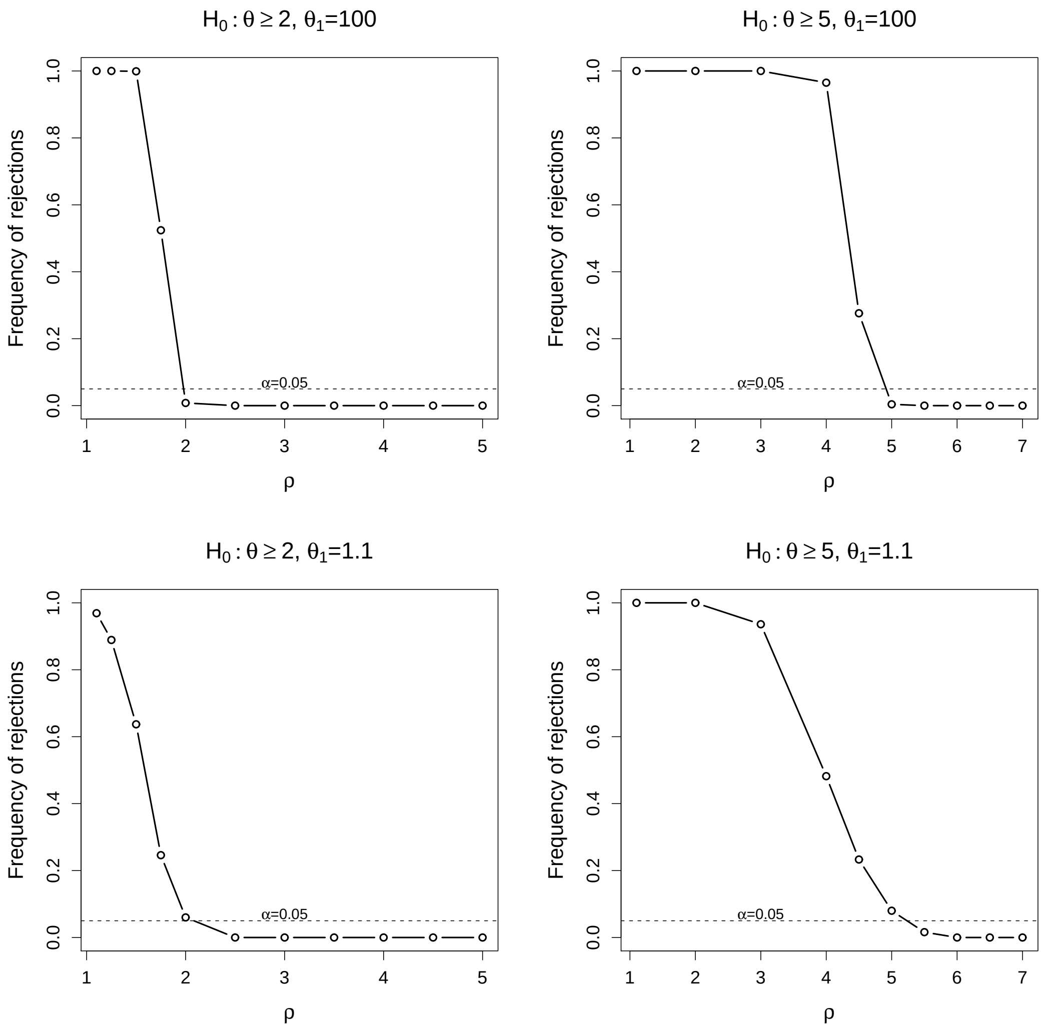

Figure 4.37.: Power of the improved test for $\theta \geq \theta_0$ with ε-contamination ($\theta_1 = 100$ in the first row and $\theta_2 = 1.1$ in the second row, $\varepsilon = 0.05$ each).

on likelihood-depth are consistent is completely missing. This can also be a task for the future. We showed in simulation studies that the tests for both copulas seem to have a good power and are robust against ε-contamination and outliers.

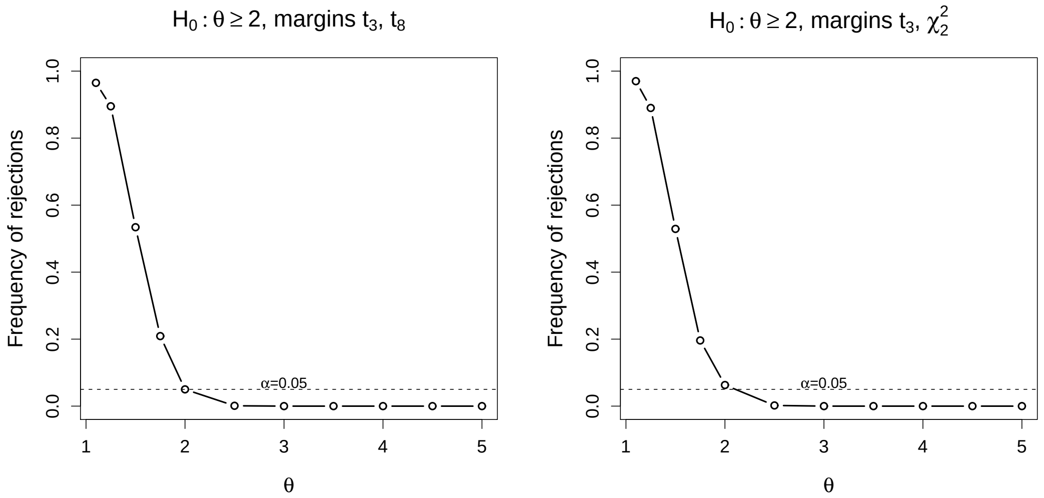

Figure 4.38.: Power of the improved test for unknown margins.

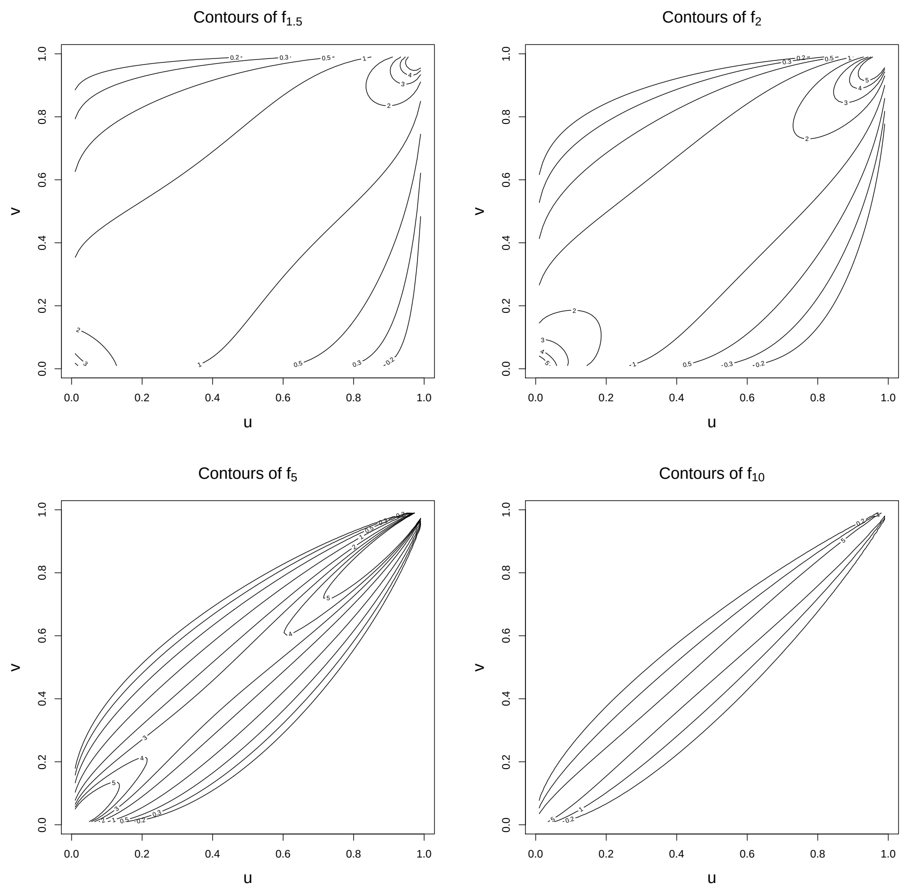

Figure 4.39.: Contours of the density of the Gumbel copula for different θ.

A. Weak convergence and empirical processes

We recall a few definitions and basic statements from the textbook of van der Vaart and Wellner “Weak Convergence and Empirical Processes”, [VaWe 1996], mainly from Chapter 1.2 and 1.9. Then we give the Glivenko-Cantelli theorem from Section 2.4 in [VaWe 1996] and we introduce the Vapnik-$\tilde{\text{C}}$ervonenkis classes and some relevant results from Section 2.6 of [VaWe 1996], especially the statement that Vapnik-$\tilde{\text{C}}$ervonenkis classes are Glivenko-Cantelli. This theory is used to proof consistency of the estimator for the correlation coefficient based on likelihood-depth.

Let be $(\Omega, \mathcal{A}, P)$ a probability space, $T : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ an arbitrary map.

Definition A.1. *[VaWe 1996, Section 1.2]*

(i) The outer integral of T with respect to P is defined as

$$\mathrm{E}^*T := \inf\{\mathrm{E}U; U \geq T, U : \Omega \to \mathbb{R} \cup \{-\infty, \infty\} \text{ measurable and } \mathrm{E}U \text{ exists}\}.$$

(ii) The outer probability of a subset B of Ω is given by

$$P^*(B) := \inf\{P(A); A \supset B, A \in \mathcal{A}\}.$$

(iii) The inner integral is given by $\mathrm{E}_*T = -\mathrm{E}^*(-T)$ and the inner probability by $P_*(B) = 1 - P^*(\Omega \setminus B)$.

Some properties of the outer integral and outer probability are given by the following two lemmas. Here $\vee$ denotes the maximum and $\wedge$ the minimum.

Lemma A.2. *[VaWe 1996, Lemma 1.2.2] Let be $S, T : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ maps. Then it holds*

(i) $(S + T)^* \leq S^* + T^*$, where “=” is true if S is measurable,

(ii) $(S - T)^* \geq S^* - T^*$,

(iii) $|S^* - T^*| \leq |S - T|^*$,

(iv) if S is measurable, then $(ST)^* = S1_{S>0}T^* + S1_{S<0}T_*$,

(v) $(ST)^* \leq S^*T^*1_{S^*>0,T^*>0} + S^*T_*1_{S*<0,T_*>0} + S_*T^*1_{S^*>0,T_*<0} + S_*T_*1_{S_*<0,T_*<0}$,

(vi) $(1_{T>c})^* = 1_{T^*>c}$ *for any* $c \in \mathbb{R}$*,*

(vii) $|T|^* = T^* \vee (-T)^* = T^* \vee (-T_*) = |T_*| \vee |T_*|$*,*

(viii) $(S \vee T)^* = S^* \vee T^*$*,*

(ix) $(S \wedge T)^* = S^* \wedge T^*$*, with equality if* S *is measurable.*

Moreover, $P^*(T > c) = P(T^* > c)$ *for any* $c \in \mathbb{R}$*.*

Lemma A.3. *[VaWe 1996, Lemma 1.2.3] For any subset* B *of* Ω *it holds*

(i) $P^*(B) = \mathrm{E}^* 1_B$*,* $P_*(B) = \mathrm{E}(1_B)_*$*,*

(ii) there exists a measurable set $B^* \supset B$ *with* $P(B^*) = P^*(B)$ *for any such* B^**, it holds that* $1_{B^*} = (1_B)^*$*,*

(iii) $(1_B)^* + (1_{\Omega - B})_* = 1$*.*

Now we give the definitions of the different types of convergence. The convergence in [VaWe 1996] is defined for nets of processes $\{X_\alpha\}_{\alpha \in A}$, with A a directed set, i.e. a set with a partial order. For a sequence, the directed set is the set of natural numbers with the usual ordering. See also [VaWe 1996].

Definition A.4. *[VaWe 1996, Definition 1.9.1] Let be* $X_\alpha, X : \omega \to \mathbb{D}$ *maps, with* $(\mathbb{D}, d)$ *a metric space.*

(i) X_α *converges in outer probability to* X*, if* $d(X_\alpha, X)^* \to 0$ *in probability,i.e.*

$$P(d(X_\alpha, X)^* > \varepsilon) = P^*(d(X_\alpha, X) > \varepsilon) \to 0,$$

for all $\varepsilon > 0$*.*

(ii) X_α *converges almost uniformly to* X*, if for all* $\varepsilon > 0$ *there exists a measurable* A *with* $P(A) \geq 1 - \varepsilon$ *and* $d(X_\alpha, X) \to 0$ *uniformly on* A*.*

(iii) X_α *converges outer almost surely to* X *if* $d(X_\alpha, X)^* \to 0$ *for some versions of* $d(X_\alpha, X)^*$*.*

(iv) X_α *converges almost surely to* X *if* $P_*(\lim d(X_\alpha, X) = 0) = 1$*.*

Definition A.5. *[VaWe 1996, Definition 2.1.5] The covering number* $N(\varepsilon, \mathscr{F}, \|\cdot\|)$ *is the minimal number of balls* $\{g; \|g - f\| < \varepsilon\}$ *of radius* ε *needed to cover the set* $\mathscr{F}$*. The centers of the balls need not belong to* $\mathscr{F}$*, but they should have finite norms. The entropy number is the logarithm of the covering number.*

An envelope function of a class $\mathscr{F}$ is any function $x \mapsto F(x)$, such that $|f(x)| \leq F(x)$ for every x and every f. The minimal envelope function is $x \mapsto \sup_f |f(x)|$.

Now we define Glivenko-Cantelli classes.

Definition A.6. *[VaWe 1996, Section 2.1] Let $\mathscr{F}$ be a collection of measurable functions $f : \mathscr{X} \to \mathbb{R}$, where $(\mathscr{X}, \mathscr{A})$ is a measurable space. $\mathscr{F}$ is a Glivenko-Cantelli class, if*

$$\|\mathbb{P}_N - P\|_{\mathscr{F}} \to 0,$$

in outer probability or outer almost surely where $\|Q\|_{\mathscr{F}} = \sup_{f\in\mathscr{F}} |\int f dQ|$ and $\mathbb{P}_N$ is the empirical measure of a sample of random elements $X_1, \ldots, X_N$ of $\mathscr{X}$, $\mathbb{P}_N(C) = \frac{1}{N}\sharp\{X_i; X_i \in C\}$.

The next theorem is one of the Glivenko-Cantelli theorems given in Section 2.4 of [VaWe 1996]. It shows the connection between the entropy of $\mathscr{F}$ and $\mathscr{F}$ being a Glivenko-Cantelli class. This theorem is for measurable classes, these are defined as follows.

Definition A.7. *[VaWe 1996, Definition 2.3.3] A class $\mathscr{F}$ of measurable functions $f : \mathscr{X} \to \mathbb{R}$ on a probability space $(\mathscr{X}, \mathcal{A}, P)$ is called P-measurable class, if the function*

$$(X_1, \ldots, X_n) \mapsto \|\sum_{i=1}^{n} e_i f(X_i)\|_{\mathscr{F}}$$

is measurable on the completion of $(\mathscr{X}^n, \mathcal{A}^n, P^n)$ for every n and every vector $(e_1, \ldots, e_n) \in \mathbb{R}^n$.

Theorem A.8. *[VaWe 1996, Theorem 2.4.3] Let $\mathscr{F}$ be a P-measurable class of measurable functions with envelope F such that $P^*F < \infty$. Let $\mathscr{F}_M$ be the class of functions $f1_{F\le M}$, when f ranges over $\mathscr{F}$. Further $\mathbb{P}_n$ shall denote the empirical measure. If $\log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_N)) = o_{P^*}(N)$ for every ε and $M > 0$, then $\|\mathbb{P}_N - P\|^* \to 0$ both almost surely and in mean. In particular, $\mathscr{F}$ is Glivenko-Cantelli.*

Remark A.9. *[VaWe 1996, Example 2.1.4] Let $\mathscr{C}$ be a collection of measurable sets in the sample space $(\mathscr{X}, \mathscr{A})$ and $\mathscr{F}$ the set of indicator functions in $\mathscr{C}$. This leads to the empirical distribution indexed by set*

$$C \mapsto \mathbb{P}_N(C) := \frac{1}{N}\sharp\{X_i; X_i \in C\}.$$

In this case, it is convenient to make the identification $C \leftrightarrow 1_C$, both in notation and in terminology. Hence $\mathscr{C}$ is called a Glivenko-Cantelli class, if $\|\mathbb{P}_N - P\|_{\mathscr{C}}$ converges to zero in outer probability or outer almost surely.

The one-dimensional Glivenko-Cantelli-Lemma shows the uniform convergence of the empirical distribution function:

Example A.10. Let be $(X_n)_{n\in\mathbb{N}}$ i.i.d. a sequence of univariate variables with distribution function $F_n = F$ for all $n \in \mathbb{N}$, $F_N : \Omega \to [0,1]$, $F_N(x,\omega) := \frac{1}{N}\sum_{n=1}^{N} 1_{(-\infty,x]}(X_n(\omega))$ and

$$d_N : \Omega \to \mathbb{R}^+, d_N(\omega) := \sup_{x\in\mathbb{R}} |F_N(x,\omega) - F(x)|.$$

Then $d_N \to 0$ almost surely as $N \to \infty$.

One kind of classes that fulfill the hypothesis of this theorem are the so-called Vapnik-Červonenkis classes, or simply VC-classes. The theory can be found in Section 2.6 of [VaWe 1996], we consider only the definition and relevant theorems.

Definition A.11. *[VaWe 1996, Section 2.6] Let be $\mathscr{C}$ a collection of subsets of a set $\mathscr{X}$. An arbitrary set of n points $\{x_1, \dots, x_n\}$ posses 2^n subsets. $\mathscr{C}$ picks out a certain subset from $\{x_1, \dots, x_n\}$, if this can be formed as a set of the form $C \cap \{x_1, \dots, x_n\}$ for $C \in \mathscr{C}$. $\mathscr{C}$ is said to shatter $\{x_1, \dots, x_n\}$ if each of the 2^n subsets can be picked out. The VC-index $V(\mathscr{C})$ of a class $\mathscr{C}$ is the smallest n for which no set of size n is shattered by $\mathscr{C}$. Formally this means*

$$\Delta_n(\mathscr{C}, x_1, \dots, x_n) := \sharp\{C \cap \{x_1, \dots, x_n\}; C \in \mathscr{C}\},$$
$$V(\mathscr{C}) := \inf\{n; \max_{x_1,\dots,x_n} \Delta_n(\mathscr{C}, x_1, \dots, x_n) < 2^n\}.$$

A collection $\mathscr{C}$ of measurable sets C is called VC-class, if its index is finite.

Example A.12. The class of the half-open intervals $\mathscr{C} := \{(-\infty, x]; x \in \mathbb{R}\}$ is a VC-class with VC-index $V(\mathscr{C}) = 2$. $\mathscr{C}$ shatters no set $\{x_1, x_2\}$. Let be $x_1 < x_2$, then there exists no interval $C \in \mathscr{C}$ such that $C \cap \{x_1, x_2\} = \{x_2\}$.

For these classes we have the following lemma and theorem, see van der Vaart and Wellner [VaWe 1996].

Lemma A.13. *[VaWe 1996, Corollary 2.6.3] For a VC-class of sets of index $V(\mathscr{C})$, one has*

$$\max \Delta_n(\mathscr{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathscr{C})-1} \binom{n}{j}.$$

Consequently the numbers on the left side grow polynomially of order at most $O(n^{V(\mathscr{C})-1})$ as $n \to \infty$.

Theorem A.14. *[VaWe 1996, Theorem 2.6.4] There exists a universal constant K such that for any VC-class $\mathscr{C}$ of sets, any probability measure Q, any $r \geq 1$ and $0 < \varepsilon < 1$,*

$$N(\varepsilon, \mathscr{C}, L_r(Q)) \leq KV(\mathscr{C})(4e)^{V(\mathscr{C})} \left(\frac{1}{\varepsilon}\right)^{r(V(\mathscr{C})-1)}.$$

Here the $L_r(Q)$-norm of a function f is $\|f\|_{Q,r} = (\int |f|^r dQ)^{\frac{1}{r}}$. So VC-classes are polynomial classes in the sense that their covering numbers are bounded by a polynomial in ε. With this we end up in

Proposition A.15. *Any appropriately measurable Vapnik-Červonenkis class is Glivenko-Cantelli (provided its envelope function is integrable).*

This statement can be found in [VaWe 1996] in the beginning of Section 2.4.

Example A.16. With Proposition A.15 and Example A.12 we have the uniform convergence of the empirical distribution function of Example A.10.

B. R source code

All simulation studies were done in R, [R 2009].

B.1. Weibull distribution

The likelihood-depth and simplicial likelihood-depth of a parameter `a` in censored or uncensored (`c0=Inf`) data `data` can be calculated using the following functions.

```
depth_a_cens<-function(a,data,c0=Inf)
{
 sol<-numeric(length(a))
 for (i in (1:length(a)))
 {
  ai<-a[i]
  N<-length(data)
  b<-median(data)
  c1<-0.259246
  c2<-2.23998
  uS<-c1^(1/ai)*b
  oS<-min(c2^(1/ai)*b,c0)
  data1<-data[data>=uS]
  data1<-data1[data1<=oS]
  sol[i]<-min((length(data[data<=uS])+length(data[data>=oS])),length(data1))/N
 }
 sol
}

sim_depth_a_cens<-function(a,data,c0=Inf)
{
 sol<-numeric(length(a))
 data_cens<-data[data>=c0]
 data_uncens<-data[data<c0]
 N<-length(data)
 k<-length(data_uncens)
 b<-median(data)
 c1<-0.259246
 c2<-2.23998
 for (i in (1:length(a)))
```

```
 {
  ai<-a[i]
  #calculating t01
  uS<-c1^(1/ai)*b
  #calculating t02
  oS<-c2^(1/ai)*b
  #data between t01 and t02
  data1<-data_uncens[data_uncens>uS]
  data1<-data1[data1<oS]
  Npos<-length(data1)
  Nneg<-length(data_uncens[data_uncens<uS])
        +length(data_uncens[data_uncens>oS])+(N-k)
  Nzero<-length(data[data==uS])+length(data[data==oS])
  sol[i]<-2/(N*(N-1))*(Npos*Nneg+Npos*Nzero+Nneg*Nzero+Nzero*(Nzero-1)/2)
 }
 sol
}
```

The likelihood-depth of the scale parameter is just $\frac{1}{N}\min(\sharp\{n; y_n \geq b\}, \sharp\{n; y_n \leq b\})$, therefore no new function is introduced. The simplicial depth can be calculated using the following function, for censored and uncensored data.

```
#simplicial depth of b in censored data
sim_depth_b_cens<-function(b,data,c0=Inf)
{
 sol<-numeric(length(b))
 N<-length(data)
 for (i in (1:length(b)))
 {
  bi<-b[i]
  if (bi<c0)
  {
   Npos<-length(data[data>bi])
   Nneg<-length(data[data<bi])
   Nzero<-length(data[data==bi])
   sol[i]<-2/(N*(N-1))*(Npos*Nneg+Npos*Nzero+Nneg*Nzero+Nzero*(Nzero-1)/2)
  }
  else
  {
   Npos<-length(data)-length(data[data<c0])
   Nneg<-length(data[data<c0])
   sol[i]<-2/(N*(N-1))*Npos*Nneg
  }
 }
 sol
}
```

B.1.1. Estimators

To find the estimators for the parameters of the Weibull distribution we use Theorem 3.15. The next lines are the source code for the estimation of the parameters of the Weibull distribution for uncensored ($c_0 = \infty$) and type-I right-censored data. The first function determines a parameter with maximum depth in the data. Here we used, that we know the maximum value of the likelihood-depth and that the likelihood-depth is not a continuous function, so we have to use discrete optimization. As a search region for the shape parameter with maximum depth, one can for example start with $[0, 10 \cdot \hat{a}_{MLE}]$(=[lb,ub]), where $\hat{a}_{MLE}$ denotes the maximum likelihood estimator of a.

```
max_depth_cens<-function(data,c0=Inf,lb,ub,steps)
{
 N<-length(data)
 rek<<-0
 rek2<<-0
 rek3<<-0
 success<-0
 maxd<-floor(N/2)/N
 maximize<-function(lb,ub,steps)
 {
  h<-(ub-lb)/steps
  fm<-0
  lbn<-lb
  ubn<-ub
  for (i in (1:steps))
  {
   f1<-depth_a_cens(lb+i*h,data,c0)
   if (f1>fm){fm<-f1;lbn<-lb+i*h}
   if (f1<fm){if ((lb+i*h)<ubn&(lb+i*h)>lbn){ubn<-lb+i*h}}
  }
  if(lbn==lb&ubn==ub&rek<20)
  {
   rek<<-rek+1
   intn<-maximize(lb,ub,steps*10)
   lb<-intn[1]
   ub<-intn[2]
  }
  if (abs(fm-maxd)<10^(-5)|fm>maxd){success<-1}
  else
  {
   if (rek<20)
   {
    rek<<-rek+1
    if (lbn>lb)
    {
     lbn<-lbn-h
```

```
    }
    if (lbn==lb&ubn==ub&rek<20)
    {
     rek<<-rek+1
     intn<-maximize(lb,ub,steps*10)
     lbn<-intn[1]
     ubn<-intn[2]
    }
    intn<-maximize(lbn,ubn,steps)
    lbn<-intn[1]
    ubn<-intn[2]
   }
  }
  c(lbn,ubn,success)
 }
 erg<-maximize(lb,ub,steps)
 #return left bound, because right bound maybe has depth smaller than 1/2
 #left bound on the plateau
 c(erg[1])
}
```

We already considered, that there can be more than one parameter with maximum depth, to find all parameters with maximum depth around the one, that was determined by the function `max_depth`, we use the function `plateau_detection_cens`.

```
plateau_detection_cens<-function(data,maxa,c0=Inf,peps=0.001)
{
 obj<-depth_a_cens(maxa,data,c0)
 ps<-TRUE
 count1<-0
 amax_r<-maxa
 while(ps & count1<100)
 {
  eps<-1
  while (eps>peps)
  {
   eps<-peps #min. plateau-width
   deps<-depth_a_cens(amax_r+eps,data,c0)
   while (obj<=deps)
   {
    count1<-count1+1
    if (obj<deps)
    {
     obj<-deps
     maxa<-amax_r+eps
     break
    }
```

```
    amax_r<-amax_r+eps
    eps<-eps*2
    deps=depth_a_cens(amax_r+eps,data,c0)
   }
  }
  ps=FALSE
 }
 if (count1>=99)
 {
   print("Warning: Plateau-detection stopped on the right")
 }
 ps<-TRUE
 amax_l<-maxa
 count1<-0
 while(ps & count1<100)
 {
  eps<-1
  while (eps>peps)
  {
   eps<-peps
   deps<-depth_a_cens(amax_l-eps,data,c0)
   while (obj<=deps)
   {
    count1<-count1+1
    if (obj<deps)
    {
     obj<-deps
     maxa<-amax_l-eps
     break
    }
    amax_l<-amax_l-eps
    eps<-eps*2
    deps=depth_a_cens(amax_l-eps,data,c0)
   }
  }
  ps=FALSE
 }
 if (count1>=99)
 {print("Warning: Plateau-detection stopped on the left")}
 c(amax_l,amax_r)
}
```

To estimate the shape and the scale parameter of the Weibull distribution as described in Procedure 3.15 and 3.23, we use the three functions from above and the correction functions for the shape and scale parameter. The function `estimate_LDE_cens` returns the lower and the upper bound of the LDE for the shape and the scale parameter. The interval that is returned is not a confidence interval.

```
estimate_LDE_cens<-function(data,c0=Inf,lb=0,ub=100,steps=100)
{
 c1<-0.259246
 c2<-2.23998
 #estimate b as the median
 b1<-median(data)
 #if more than half the data is censored this method is not applicable
 if (b1>=c0){stop("Warning! c0<=median(data), method not applicable!")}
 #maximum depth for a
 maxa<-max_depth_cens(data,c0,lb,ub,steps)
 amax<-plateau_detection_cens(data,maxa,c0)
 amax_r<-amax[2]
 amax_l<-amax[1]
 #we have to distinguish two cases
 #check if upper range of amax so that t02<c0 (then also for lower range)
 if (c2^(1/amax_r)*b1<c0){
 #correction is the same as in uncensored case
 a<-c(0.756714*amax_l,0.756714*amax_r)}
 else
 {
  #the correction for a can be found by solving
  #-2^(-(c0/b1)^a)+2^(-c1^(a/am))=1/2
  #therefore we use newtons-method, this is implemented in corr
  corr<-function(am,a0)
  {
   f<-function(a){-2^(-(c0/b1)^a)+2^(-c1^(a/am))-1/2}
   df<-function(a)
   {
    h<-0.001
    erg<-(f(a+h)-f(a))/h
    erg
   }
   x1<-a0
   x2<-x1-f(x1)/df(x1)
   count<-0
   while(abs(x2-x1)>10^(-6)&count<50)
   {
    x1<-x2
    x2<-x1-f(x1)/df(x1)
    count<-count+1
   }
   if (count>=50){print("Iteration stopped")}
   x2
  }
  if (c2^(1/amax_l)*b1>=c0)
  {
   am<-amax_l
```

```
   a_l<-corr(am,amax_r)
   am<-amax_r
   a_r<-corr(am,am)
   a<-c(a_r,a_l)
  }
  #else the equation has only to be solved for amax_r and
  #for amax_l we can use the same correction as before
  else
  {
   am<-amax_r
   a_r<-corr(am,am)
   a_l<-0.756714*amax_l
   a<-c(a_r,a_l)
  }
  }
  b<-b1/(log(2)^(1/a))
  theta<-c(a,sort(b))
  theta
}
```

B.1.2. Tests

We start with the tests for the shape parameter. For the test statistic the simplicial likelihood-depth is needed, the function is given in the beginning of this section.

```
#if b known, set b=b_0, else b=F, than med=median(dat) must be given
p_a<-function(a,c0=Inf,b=F,med=0)
{
 pa<-numeric(length(a))
 for (i in (1:length(a)))
 {
  ai<-a[i]
  c1<-0.259246
  c2<-2.23998
  if (b)
  {
   #calculating t01
   uS<-c1^(1/ai)*b
   #calculating t02
   oS<-c2^(1/ai)*b
   if (b<c0 & oS<c0)
   {
    pa<-exp(-c1)-exp(-c2)
   }
   if (b<c0 & oS>=c0)
```

```
   {
    pa<-exp(-c1)-exp(-(c0/b)^ai)
   }
   if (b>=c0)
   {
   pa[i]<-exp(-c1)
   }
  }
  if (!b)
  {
   b<-med
   #calculating t01
   uS<-c1^(1/ai)*b
   #calculating t02
   oS<-c2^(1/ai)*b
   if (b>=c0)
   {
    stop("Warning:Median of data should be smaller than censor time")
   }
   if (oS<c0)
   {
    pa[i]<-2^(-c1)-2^(-c2)
   }
   if (oS>=c0)
   {
    pa[i]<-2^(-c1)-2^(-(c0/b)^ai)
   }
  }
 }
 pa
}

#Test for a<=a_0
test_le_a_cens<-function(data,a0,b0=F,c0,level=0.05,lb=0,ub=100,steps=100)
{
 #print(c0)
 c1<-0.259246
 c2<-2.23998
 if (!b0)
 {
  beta<-median(data)
  oS<-c2^(1/a0)*beta
  if (oS>=c0 & a0<0.455/log(c0/beta))
  {
     f<-function(a){
      t1<-1-2^(-c1)+2^(-(c0/beta)^a)
      erg<-t1-2^(-c1^(a0/a))+2^(-(c0/beta)^a0)
```

```
     erg}
   if (f(a0)*f(a0/2)<0){a<-uniroot(f,int=c(a0/2,a0))$root}
   else
   {
    if (f(a0)*f(a0/4)<0){a<-uniroot(f,int=c(a0/4,a0))$root}
    else{a<-a0;print("Warning: No correction c_alpha^2 determined")}
   }
 }
 testa<-seq(0.001,a0,0.001)
}
if(b0)
{
 oS<-c2^(1/a0)*b0
 if (b0<c0&oS>=c0)
 {
  if(a0<0.265/log(c0/b0))
  {
   f<-function(a)
   {
    t1<-1-exp(-c1)+exp(-(c0/b0)^a)
    t1-exp(-c1^(a0/a))+exp(-(c0/b0)^a0)
   }
   if (f(a0)*f(a0/2)<0){a0<-uniroot(f,int=c(a0/2,a0))$root}
   else
    {
     if (f(a0)*f(a0/4)<0){a0<-uniroot(f,int=c(a0/4,a0))$root}
     else{print("Warning: No correction c_alpha^2 determined")}
    }
   #a0<-log(c1)/(log(-log(1-exp(-c1)+2*exp(-(c0/b0)^a0))))*a0
   #print(a0)
  }
 }
 testa<-seq(0.001,a0,0.001)
}
TS<-function(a,b0)
{
if (!b0)
{
 pa<-p_a(a,c0,med=beta)
 depth<-sim_depth_a_cens(a,b=F,data,c0)
 ET<-2*pa*(1-pa)
 VarT<-pa*(1-pa)*(1-2*pa)^2
 TST<-sqrt(length(data))*(depth-ET)/(2*sqrt(VarT))
}
if(b0)
{
 pa<-p_a(a,c0,b0)
```

```
  depth<-sim_depth_a_cens(a,b0,data,c0)
  ET<-2*pa*(1-pa)
  VarT<-pa*(1-pa)*(1-2*pa)^2
  TST<-sqrt(length(data))*(depth-ET)/(2*sqrt(VarT))
 }
 TST
 }
TST<-lapply(testa,TS,b0=b0)
TS<-max(unlist(TST))
if (TS<qnorm(level))
 {H1<-T}
else {H1<-F}
H1
}

#Test for a>=a0

test_ge_a_cens<-function(data,a0,b0=F,c0,level=0.05,lb=0,ub=100,steps=100)
{
 count<<-count+1
 print(count)
 #print(c0)
 c1<-0.259246
 c2<-2.23998
 if (!b0)
 {
  beta<-median(data)
  oS<-c2^(1/a0)*beta
  if(length(data[data==c0])<1)
  {
   calpha<-function(a){1.835*a}
   testa<-seq(calpha(a0),a0+100,0.01*a0)
  }
  else
  {
   if (c0>oS)
   {
    calpha<-function(a){1.835*a}
    testa<-seq(calpha(a0),a0+100,0.01*a0)
   }
   else
   {
    if (a0>=0.455/log(c0/beta))
    {
     f<-function(a){
     t1<-1-2^(-c1)+2^(-(min(c0,c2^(1/a)*beta)/beta)^a)
     erg<-t1-2^(-c1^(a0/a))+2^(-(min(c0,c2^(1/a)*beta)/beta)^a0)
```

```
    erg
    }
    if(f(a0)*f(a0*10)<0)
    {
     as<-uniroot(f,int=c(a0,a0*10))$root
    }
    else
    {
     as<-a0
    }
   testa<-seq(as,as+100,0.01*a0)
   }
   else
   {
    testa<-seq(a0,a0+10,a0)
   }
  }
 }
}
else
{
 oS<-c2^(1/a0)*b0
 if(c0>b0 & c0>oS){a0<-2.275*a0}
 if(c0>b0&c0<=oS)
 {
  if (a0>0.265/log(c0/b0))
  {
    f<-function(a)
    {
     t1<-1-exp(-c1)+exp(-(min(c0,b0*c2^(1/a))/b0)^a)
     t1-exp(-c1^(a0/a))+exp(-(min(c0,c2^(1/a)*b0)/b0)^a0)
    }
    if (f(a0)*f(100*a0)<0){a0<-uniroot(f,int=c(a0,100*a0))$root}
  }
 }
 testa<-seq(a0,a0+100,0.01*a0)
}
TS<-function(a,b0)
{
if (!b0)
{
 pa<-p_a(a,c0,med=beta)
 depth<-sim_depth_a_cens(a,b=F,data,c0)
 ET<-2*pa*(1-pa)
 VarT<-pa*(1-pa)*(1-2*pa)^2
 TST<-sqrt(length(data))*(depth-ET)/(2*sqrt(VarT))
}
```

```
 if(b0)
 {
  pa<-p_a(a,c0,b0)
  depth<-sim_depth_a_cens(a,b0,data,c0)
  ET<-2*pa*(1-pa)
  VarT<-pa*(1-pa)*(1-2*pa)^2
  TST<-sqrt(length(data))*(depth-ET)/(2*sqrt(VarT))
 }
 TST
 }
 TST<-lapply(testa,TS,b0=b0)
 TS<-max(unlist(TST))
 if (TS<qnorm(level))
  {H1<-T}
 else {H1<-F}
 H1
}

#Test for a=a0

test_eq_a_cens<-function(data,a0,b0=F,c0,level=0.05)
{
 c1<-0.259246
 c2<-2.23998
 if (!b0)
 {
  beta<-median(data)
  oS<-c2^(1/a0)*beta
  #print(oS)
  if(length(data[data==c0])<1)
  {
   calpha<-function(a){1.835*a}
   a<-calpha(a0)
  }
  else
  {
   if (c0>oS)
   {
    calpha<-function(a){1.835*a}
    a<-calpha(a0)
   }
   else
   {
    if (a0>=(0.455/log(c0/beta)))
    {
     f<-function(a){
     t1<-1-2^(-c1)+2^(-(min(c0,c2^(1/a)*beta)/beta)^a)
```

```
    erg<-t1-2^(-c1^(a0/a))+2^(-(min(c0,c2^(1/a)*beta)/beta)^a0)
    erg
    }
    if(f(a0)*f(a0*100)<0)
    {
     a<-uniroot(f,int=c(a0,a0*100))$root
    }
    else
    {
     a<-a0
    }
    }
   else
   {
    f<-function(a){
    t1<-1-2^(-c1)+2^(-(c0/beta)^a)
    erg<-t1-2^(-c1^(a0/a))+2^(-(c0/beta)^a0)
    erg
    }
     if (f(a0)*f(a0/2)<0){a<-uniroot(f,int=c(a0/2,a0))$root}
    else
    {
     if (f(a0)*f(a0/4)<0){a<-uniroot(f,int=c(a0/4,a0))$root}
     else{a<-a0;print("Warning: No correction c_alpha^2 determined")}
    }

   }
  }
 }
 TS<-function(a)
 {
  pshape<-p_a(a,c0,med=beta)
  ET<-2*pshape*(1-pshape)
  VarT<-pshape*(1-pshape)*(1-2*pshape)^2
  #print(sim_depth_a_cens(a,b0,data,c0))
  TS<-sqrt(length(data))*(sim_depth_a_cens(a,b0,data,c0)-ET)/(2*sqrt(VarT))
  TS
 }
 #print(c(a,a0))
 #print(TS(a0))
 #print(TS(a))
 if (TS(a0)<qnorm(level/2)|TS(a)<qnorm(level/2)){H1<-T}
 else{H1<-F}
}
if (b0)
{
 oS<-c2^(1/a0)*b0
```

```
#no censored data:
if(length(data[data==c0])<1)
{
 calpha<-function(a){2.275*a}
 a<-calpha(a0)
}
else
{
 if(c0>oS)
 {
 calpha<-function(a){2.275*a}
 a<-calpha(a0)
 }
 else
 {
  if(a0<(0.265/log(c0/b0)))
  {
   f<-function(a)
   {
    t1<-1-exp(-c1)+exp(-(c0/b0)^a)
    t1-exp(-c1^(a0/a))+exp(-(c0/b0)^a0)
   }
   if (f(a0)*f(a0/2)<0){a<-uniroot(f,int=c(a0/2,a0))$root}
   else
   {
    if (f(a0)*f(a0/4)<0){a<-uniroot(f,int=c(a0/4,a0))$root}
    else{a<-a0;print("Warning: No correction c_alpha^2 determined")}
   }
  }
  if(a0>(0.265/log(c0/b0)))
  {
   f<-function(a)
    {
     t1<-1-exp(-c1)+exp(-(min(c0,c2^(1/a)*b0)/b0)^a)
     t1-exp(-c1^(a0/a))+exp(-(min(c0,c2^(1/a)*b0)/b0)^a0)
    }
   if (f(a0)*f(10*a0)<0){a<-uniroot(f,int=c(a0,10*a0))$root}
   else{a<-a0}
  }
 }

}
TS<-function(a)
{
 pshape<-p_a(a,c0,b=b0)
 ET<-2*pshape*(1-pshape)
 VarT<-pshape*(1-pshape)*(1-2*pshape)^2
```

```
   sqrt(length(data))*(sim_depth_a_cens(a,b0,data,c0)-ET)/(2*sqrt(VarT))
  }
  if (TS(a0)<qnorm(level/2)|TS(a)<qnorm(level/2)){H1<-T}
  else{H1<-F}
 }
  H1
}

#confidence-interval for a
ci_shape_cens<-function(dat,b0=F,c0,q=0.95,peps=0.01)
{
 level<-1-q
 at<-0.01
 tst<-test_eq_a_cens(dat,at,b0,c0,level)
 if (!tst){print("Warning: No lower bound")}
 else
 {
  while(tst&at<100)
  {
   at<-at+peps
   tst<-test_eq_a_cens(dat,at,b0,c0,level)
  }
 }
 ar<-at
 if(at>99.9){print("Warning:No confidence interval found!")}
 else
 {
  while(!tst&at<100)
  {
   at<-at+peps
   tst<-test_eq_a_cens(dat,at,b0,c0,level)
  }
  if (at>99.9){print("Warning:No upper bound!")}
 }
 al<-at
 c(ar,al)
}
```

The tests for the scale parameter.

```
#Test for b>=b0
test_ge_b_cens<-function(dat,b0,c0,level=0.05)
{
 TS<-function(b)
 {
  N<-length(dat)
  ps<-exp(-1)
```

```
  TS<-sqrt(N)*(sim_depth_b_cens(b,dat,c0)-2*ps*(1-ps))
  /(2*sqrt(ps*(1-ps)*(1-2*ps)^2))
  TS
 }
 b<-seq(b0,b0+10,0.01)
 TSb<-lapply(b,TS)
 TSb_max<-max(unlist(TSb))
 if (TSb_max<qnorm(level)){H1<-T}
 else {H1<-F}
 H1
}

#test for b<=b0
test_le_b_cens<-function(dat,b0,a0=F,c0,level=0.05)
{
 if(!a0)
 {
  a0<-mean(estimate_LDE_cens(dat,c0)[1:2])
 }
 TS<-function(b)
 {
  N<-length(dat)
  ps<-exp(-1)
  TS<-sqrt(N)*(sim_depth_b_cens(b,dat,c0)-2*ps*(1-ps))
      /(2*sqrt(ps*(1-ps)*(1-2*ps)^2))
  TS
 }
 calpha_b0<-b0*(-log(1-exp(-1)))^(1/a0)
 b<-seq(0.01,calpha_b0,0.01)
 TSb<-lapply(b,TS)
 TSb_max<-max(unlist(TSb))
 if (TSb_max<qnorm(level)){H1<-T}
 else {H1<-F}
 H1
}

#test for b=b0
test_eq_b_cens<-function(dat,b0,a0=F,c0,level=0.05)
{
 if(!a0)
 {
  a0<-mean(estimate_LDE_cens(dat,c0)[1:2])
 }
 TS<-function(b)
 {
  N<-length(dat)
  ps<-exp(-1)
```

```
  TS<-sqrt(N)*(sim_depth_b_cens(b,dat)-2*ps*(1-ps))
      /(2*sqrt(ps*(1-ps)*(1-2*ps)^2))
  TS
 }
 calpha_b0<-b0*(-log(1-exp(-1)))^(1/a0)
 if (TS(b0)<qnorm(level/2)|TS(calpha_b0)<qnorm(level/2)){H1<-T}
 else{H1<-F}
 H1
}

#confidence interval for b
ci_scale_LDE_cens<-function(dat,a0=F,c0,level=0.95,lb=0.01,ub=2000,peps=0.01)
{
 alpha<-1-level
 b<-lb
 tst<-test_eq_b_cens(dat,b,a0,c0,alpha)
 if (!tst){print("Warning:No lower bound!")}
 else
 {
  while (tst&b<ub)
  {
   b<-b+peps
   tst<-test_eq_b_cens(dat,b,a0,c0,alpha)
  }
 }
 bli<-b
 if (b>(ub-peps)){print("Warning: No confidence interval found!")}
 else
 {
  while (!tst &b<ub)
  {
   b<-b+peps
   tst<-test_eq_b_cens(dat,b,a0,c0,alpha)
  }
  if (b>(ub-peps)){print("Warning:No upper bound!")}
 }
 bri<-b
 c(bli,bri)
}
```

B.2. Gaussian copula

Most of the needed procedures in case of the Gauss copula are conform to the ones given for the Weibull distribution, so they can be just copied from there. Just a few changes have to be made, and only the procedures that differ are given here.

To simulate data with 2-dimensional normal distribution and parameters as given in the section about the correlation coefficient, the package "copula", see [Yan 2007], can be used.

```
library(copula)
#the expectation value
mu1<-c(0,0)
#the covariance matrix with correlation r
sigma1<-matrix(c(1,r,r,1),2,2)
#dataset with 100 data and parameter as given above
rmnorm(100,mu1,sigma1)
```

For the calculation of the depth of a parameter `p` in a dataset `data` the function $h_n'(p)$ is needed:

```
hn2<-function(r,z){
-((r^3-r^2*z[1]*z[2]+r*z[1]^2+r*z[2]^2-r-z[1]*z[2])/(-1+r^2)^2)}
```

The tangent depth of a parameter in 2-dim. normal distributed data can be calculated with the function `Tdepth` and the simplicial depth with `Sdepth`.

```
Tdepth<-function(r,data)
{
 gr<-0
 kl<-0
 hnr<-function(dat){h<-hn2(r,dat);h}
 u<-apply(data,1,hnr)
 gr<-sum(u>=0)
 kl<-sum(u<=0)
 if (kl<gr){kl/(length(data[,1]))}
 else {gr/(length(data[,1]))}
}

Sdepth<-function(r,data)
{
 gr<-0
 kl<-0
 l<-length(data[,1])
 hnr<-function(dat){h<-hn2(r,dat);h}
 u<-apply(Daten,1,hnr)
 gr<-sum(u>=0)
 kl<-sum(u<=0)
 depth<-gr*kl*2/((l-1)*l)
 depth
}
```

B.2.1. Estimator

An estimator for the correlation of a dataset can be calculated with analogue functions as for the Weibull distribution, only the depth-functions, the correction and the search-area have to be exchanged.

B.2.2. Tests

The required procedure for the computing of the test (the function SimDepth is also needed):

```
#the test statistic of "r" in "data"
Teststatistic <-function(r,data)
{
 N<-length(data)/2
 p_r<-p[r*1000]
 gamma<-function(pr){2*pr*(1-pr)}
 sigma<-function(pr){sqrt(pr*(1-pr)*(1-2*pr)^2)}
 TS<-sqrt(N)*(Sdepth(r,data)-gamma(p_r))/(2*sigma(p_r))
 TS
}
```

We did not display p_ρ for $\rho = 0.001, \ldots, 0.999$ here, some values can be found in Table 4.1.

The following programs can be used to test the hypotheses $H_0 : \rho \leq \rho_0$, $H_0 : \rho \geq \rho_0$ and $H_0 : \rho = \rho_0$. They return TRUE if the null-hypothesis is rejected. Give ρ_0 as rh, the data as dat and level as level. We start with the test for $H_0 : \rho \leq \rho_0$.

```
gauss.test.le<-function(rh,dat,level=0.05)
{
 rho<-seq(0.001,0.999,0.001)
 rho<-rho[rho<=rh]
 TS<-function(t){TS<-Teststatistic(t,dat)}
 mTS<-max(unlist(lapply(rho,TS)))
 if (mTS<qnorm(level))
 {
  text<-paste("H0 : rho <=",rh<-rh,"is rejected")
  logi<-TRUE
 }
 else
 {
  text<-paste("H0 : rho <=",rh, "can not be rejected")
  logi<-FALSE
  }
 return(list(maxTS=mTS,result=text,Logic=logi))
}
```

The program for testing $H_0 : \rho \leq \rho_0$.

```
gauss.test.ge<-function(rh,dat,level=0.05)
{
 rho<-seq(0.001,0.999,0.001)
 if (level!=0.05 & level!=0.1)
 {

  ("Warning: Use level 0.05 or 0.1! Will go on with level=0.05.")
  level<-0.05
 }
 if (mTS<qnorm(level))
 {
  text<-paste("H0 : rho >=",rh<-rh,"is rejected")
  logi<-TRUE
 }
 if (level==0.1)
 {
  c_rho<-function(rh)
  {
  crho<-min(0.821217+ 0.182348*rh,0.99)
  crho
  }
 if (level==0.05)
 {
  c_rho<-function(rh)
  {
  crho<-min(0.811156+0.19458*rh,0.99)
  crho
  }
 }
 rho<-rho[rho>=c_rho(rh)]
 TS<-function(t){TS<-Teststatistic(t,dat)}
 mTS<-max(unlist(lapply(rho,TS)))
 else
 {
  text<-paste("H0 : rho >=",rh, "can not be rejected");logi<-FALSE}
  return(list(maxTS=mTS,result=text,Logic=logi))
 }
```

And finally the test for $H_0 : \rho = \rho_0$.

```
gauss.test.eq<-function(rh,ds,level=0.05)
{
 ds[,1]<-(ds[,1]-mean(ds[,1]))/sd(ds[,1])
 ds[,2]<-(ds[,2]-mean(ds[,2]))/sd(ds[,2])
 TS<-Teststatistik(rh,ds)
```

```
 if (level!=0.05 & level!=0.1)
 {
  print("Warning: Use level 0.05 or 0.1! Will go on with level=0.05.")
  level<-0.05
 }
 if (level==0.1)
 {
  c_rho<-function(rh)
  {
  crho<-min(0.821217+ 0.182348*rh,0.99)
  crho
  }
 if (level==0.05)
 {
  c_rho<-function(rh)
  {
  crho<-min(0.811156+0.19458*rh,0.99)
  crho
  }
 }
 TS2<-Teststatistic(c_rho(rh),ds)
 if (TS<qnorm(level/2))
 {
  text<-paste("H0 : rho =",rh<-rh,"is rejected")
  logi<-TRUE
 }
 else
 {
  if (TS2<qnorm(level/2))
  {
   text<-paste("H0 : rho =",rh<-rh,"is rejected")
   logi<-TRUE
  }
  else
  {
   text<-paste("H0 : rho =",rh, "can not be rejected");logi<-FALSE
  }
 }
 return(list(TS=TS,TS2=TS2,result=text,Logic=logi))
}
```

B.3. The Gumbel copula

To simulate N Gumbel copula distributed data points with parameter θ, we use the package “copula”, see [Yan 2007], and the enclosed function `rcopula`:

```
library{copula}
r2Gumbel<-function(N,theta)
{
 GuCopula<-archmCopula("gumbel",theta,dim=2)
 GuDaten<-rcopula(GuCopula,N)
 GuDaten
}
```

The following source code was used to calculate the likelihood-depth of a parameter r in a dataset $z_* =$ `data`, which should be transferred as a $n \times 2$- matrix:

```
#function h_n'(r)
hn2<-function(r,x)
{
 xv<-x
 sol<-numeric(length(xv))
 for (i in (1:length(xv[,1])))
 {
  x<-xv[i,]
  sol[i]<-
  (-r*((-log(x[1]))^r*(1-3*((-log(x[1]))^r+(-log(x[2]))^r)^(1/r)+
  ((-log(x[1]))^r+(-log(x[2]))^r)^(2/r))+r^2*((-log(x[1]))^r-
  (-log(x[2]))^r)+r*(-1+((-log(x[1]))^r+(-log(x[2]))^r)^(1/r))*
  (2*(-log(x[1]))^r-(-log(x[2]))^r))*
  log(-log(x[1]))+(1+r*(-1+((-log(x[1]))^r+(-log(x[2]))^r)^(1/r))-
  3*((-log(x[1]))^r+(-log(x[2]))^r)^(1/r)
  +((-log(x[1]))^r+(-log(x[2]))^r)^(2/r))*((-log(x[1]))^r+
  (-log(x[2]))^r)*log((-log(x[1]))^r+(-log(x[2]))^r)+
  r*(r*((-log(x[1]))^r+(-log(x[2]))^r)-
  (-r*(-1+((-log(x[1]))^r+(-log(x[2]))^r)^(1/r))*((-log(x[1]))^r-
  2*(-log(x[2]))^r) +r^2*(-(-log(x[1]))^r+(-log(x[2]))^r)+
  (1-3*((-log(x[1]))^r+(-log(x[2]))^r)^(1/r) +
  ((-log(x[1]))^r+(-log(x[2]))^r)^(2/r))*(-log(x[2]))^r)*
  log(-log(x[2]))))/(r^2*(-1+r+((-log(x[1]))^r+
  (-log(x[2]))^r)^(1/r))*((-log(x[1]))^r+(-log(x[2]))^r))
 }
 sol
}

#simplicial depth of r in the data
SimDepth<-function(r,data)
{
 gr<-0
 kl<-0
 l<-length(data[,1])
 hnr<-function(dat){h<-hn2(r,dat);h}
 u<-apply(data,1,hnr)
```

```
 gr<-sum(u>=0)
 kl<-sum(u<=0)
 depth<-gr*kl*2/((l-1)*l)
 depth
}

#likelihood-depth of rs in the data
TDepth<-function(rs,data)
{
 gr<-0
 kl<-0
 hnr<-function(dat){h<-hn2(rs,dat);h}
 u<-apply(data,1,hnr)
 gr<-sum(u>=0)
 kl<-sum(u<=0)
 if (kl<gr){kl/(length(data[,1]))}
 else {gr/(length(data[,1]))}
}
```

B.3.1. Estimator

The procedure to determine an estimator, as described in this work, for the parameter θ needs the function TDepth and therefore also hn2. For the estimation the same procedures as for the Weibull distribution are used, only the depth functions, the correction of the estimator and the search area (here we can use $[1, 10 \cdot \hat{\theta}_{MLE}]$, with $\hat{\theta}_{MLE}$ the maximum likelihood estimator) have to be exchanged.

B.3.2. Tests

The following procedures are needed to use the programs for the tests:

```
#The value of the test statistic for parameter "t" and dataset "dat"
Teststatistic<-function(t,dat)
{
  N<-length(dat[,1])
sigma_theta<-function(t)
{
 if (t==1.){sigma<-(1-2*p_theta[1])^2*p_theta[1]*(1-p_theta[1])}
   if (t==1.1){sigma<-(1-2*p_theta[2])^2*p_theta[2]*(1-p_theta[2])}
 if (t>1.1&&t<=2)
 {
s<-seq(1.25,t,0.25)
sigma<-
(1-2*p_theta[2+length(s)])^2*p_theta[2+length(s)]*(1-p_theta[2+length(s)])
 }
```

```
 if (t>2&&t<=10)
 {
s<-seq(2.5,t,0.5)
sigma<-
(1-2*p_theta[6+length(s)])^2*p_theta[6+length(s)]*(1-p_theta[6+length(s)])
 }
 else
 {
if (t>10)
{
 s<-seq(10,t,5)
 sigma<-
(1-2*p_theta[22+length(s)])^2*p_theta[22+length(s)]*(1-p_theta[22+length(s)])
}
 }
 sigma<-sqrt(sigma)
 sigma
}
  gamma_theta<-function(t)
  {
if (t==1.){gamma<-2*p_theta[1]*(1-p_theta[1])}
if (t==1.1){gamma<-2*p_theta[2]*(1-p_theta[2])}
if (t>1.1&2>=t)
{
 s<-seq(1.25,t,0.25)
 gamma<-2*p_theta[2+length(s)]*(1-p_theta[2+length(s)])
}
if (t>2&&t<=10)
    {
 s<-seq(2.5,t,0.5)
 gamma<-2*p_theta[6+length(s)]*(1-p_theta[6+length(s)])
    }
else
{
 if (t>10)
 {
s<-seq(10,t,5)
gamma<-2*p_theta[22+length(s)]*(1-p_theta[22+length(s)])
 }
    }
gamma
}
TS<-sqrt(N)*(SimDepth(t,dat)-gamma_theta(t))/(2*sigma_theta(t))
TS
}

#the theta used
```

```
theta<-
c(1, 1.1, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6,
 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 11, 12, 13, 14, 15, 16, 17,
 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85,
 90, 95, 100)

#the belonging p_theta
p_theta<-c(0.564399, 0.591775, 0.614965, 0.632269, 0.639433, 0.642906,
 0.645838, 0.646917, 0.647388, 0.647617, 0.647738, 0.647806, 0.647845,
 0.647869, 0.647884, 0.647893, 0.6479, 0.647904, 0.647906, 0.647908,
 0.647909, 0.64791, 0.647911, 0.647911, 0.647911, 0.647911, 0.647911,
 0.64791, 0.64791, 0.64791, 0.64791, 0.64791, 0.647909, 0.647908,
 0.647908, 0.647907, 0.647906, 0.647905, 0.647904, 0.647901, 0.647894,
 0.647882, 0.647859, 0.647903, 0.647766, 0.647684, 0.64757, 0.647418
 )
```

For the computing of the test $H_0 : \theta \leq \theta_0$ the next program can be used, θ_0 is `th`, the data `dat` should again be a $n \times 2$-matrix. The function returns `TRUE` if the null-hypotheses is rejected.

```
gumbel.test.le <-function(th,dat,level=0.05)
{
 theta<-theta[theta<=th]
 TS<-function(t){TS<-Teststatistic(t,dat)}
 mTS<-max(unlist(lapply(theta,TS)))
 if (mTS<qnorm(level))
 {
  text<-paste("H0 : theta <=",th<-th,"is rejected")
logi<-TRUE
 }
 else
 {
text<-paste("H0 : theta <=",th, "can not be rejected")
logi<-FALSE
 }
 return(list(maxTS=mTS,result=text,Logic=logi))
}
```

And for the computing of the test for $H_0 : \theta \geq \theta_0$ use the lines below, where `oG` gives the biggest $\theta \geq \theta_0$ for that $\sup_{\theta \geq c(\theta_0)} T(\theta, z_*)$ shall be calculated:

```
gumbel.test.ge<-function(th,dat,oG=20,level=0.05)
 {
 theta<-theta[theta>=2*th&theta<=oG]
 TS<-function(t){TS<-Teststatistic(t,dat)}
 mTS<-max(unlist(lapply(theta,TS)))
```

```
 if (mTS<qnorm(level))
 {
  text<-paste("H0 : theta >=",th<-th,"is rejected")
  logi<-TRUE
 }
 else
 {
  text<-paste("H0 : theta >=",th, "can not be rejected")
  logi<-FALSE
 }
 return(list(maxTS=mTS,result=text,Logic=logi))
}
```

List of Symbols

$\mathbb{R}$, $\mathbb{N}$	real numbers, positive integers
Z_i, z_i	variable, its realization
Z_*, $Z_{*,N}$	$(Z_1, \ldots, Z_N)$
$\text{med}(z_1, \ldots, z_N)$	median of $z_1, \ldots, z_N$
f_θ	density function, p. 7
$L(\theta, z)$, $h(\theta, z)$	likelihood, log-likelihood function of f_θ
$h'(\theta, z)$	derivative of $h(\theta, z)$ with respect to θ, p. 7
$\text{E}(X)$, $\text{E}X$	expectation of X
$d_T(\theta, z_*)$, $d_S(\theta, z_*)$	(tangent) likelihood-depth, simplicial likelihood-depth of θ in z_*, p. 9
$\tilde{\theta}$	parameter with maximum likelihood-depth, p.10
T^θ_{pos}	$\{z \in \mathbb{R}^2; h'(\theta, z) \geq 0\}$, p. 10
$p_{\theta,\theta'}$	$P_\theta(Z \in T^{\theta'}_{pos})$, p. 10
p_θ	$p_{\theta,\theta}$
$s(\theta)$	solution of $P_\theta(T^{s(\theta)}_{pos}) = \frac{1}{2}$
$\lambda^+_N(\theta, z_{*,N})$	$\frac{1}{N}\sharp\{n; h'(\theta, z_n) \geq 0\}$
$\lambda^-_N(\theta, z_{*,N})$	$\frac{1}{N}\sharp\{n; h'(\theta, z_n) \leq 0\}$
$\lambda^+_{\theta_0}(\theta)$	$P_{\theta_0}(h'(\theta, Z) \geq 0)$
$\lambda^-_{\theta_0}(\theta)$	$P_{\theta_0}(h'(\theta, Z) \leq 0)$
$T(\theta, z_*)$	test statistic, p. 17
$\varphi(z_*)$	test for $H_0 : \theta \in \Theta_0$, p. 18
$\varphi^{0,=}_{\theta_0}(z_*)$, $\varphi^{=}_{\theta_0}$	(un)corrected test for $H_0 : \theta = \theta_0$, p. 20
$\varphi^{0,\geq}_{\theta_0}(z_*)$, $\varphi^{\geq}_{\theta_0}(z_*)$	(un)corrected test for $H_0 : \theta \geq \theta_0$, p. 20
$\varphi^{0,\leq}_{\theta_0}(z_*)$,$\varphi^{\leq}_{\theta_0}(z_*)$	(un)corrected test for $H_0 : \theta \leq \theta_0$, p. 20
$c^1_\alpha(\theta_0)$, $\check{c}^1_\alpha(\cdot)$	correction and its inverse for the test $H_0 : \theta \geq \theta_0$ if $s(\theta_0) > \theta_0$, p. 19
$c^2_\alpha(\theta_0)$,$\check{c}^2_\alpha(\cdot)$	correction and its inverse for the test $H_0 : \theta \leq \theta_0$ if $s(\theta_0) < \theta_0$, p. 19

$\mathrm{Wei}(a,b)$	Weibull distribution with shape a and scale b, p. 35
$c_{1,2}$	solutions of $\ln(c) = \frac{1}{c-1}$
$t^{a,b}_{01,02}$	zeros of $h'_b(a,\cdot)$, p. 41
κ	solution of $\exp(-c_1^{\kappa}) - \exp(-c_2^{\kappa}) = \frac{1}{2}$
κ_1	solution of $2^{-c_1^{\kappa_1}} - 2^{-c_2^{\kappa_1}} = \frac{1}{2}$
κ_2	solution of $\exp(-c_1^{\kappa_2}) = \frac{1}{2}$
p_{shape}	$p_a = \exp(-c_1) - \exp(-c_2)$
k_0	see p. 76
$\tilde{T}(a,t_*)$	teststatistic in the scale parameter is unknown
$\tilde{p}_{shape}$	$2^{-c_1} - 2^{-c_2}$
$\tilde{b}_N$	median
$\tilde{\varphi}_{a_0}$	tests when the scale parameter is not known
$p^{b_0}_{a_0,c_0}$	$P_{a_0,b_0}(h'_{b_0}(a_0,Y) \geq 0)$
$p^{b_0}_{a_0,a,c_0}$	$P_{a_0,b_0}(T^{a,b_0}_{pos})$
$p^{\tilde{b}_N}_{a,c_0}$	$P_{a,b_0}(h'_{\tilde{b}_N}(a,Y) \geq 0)$
$p^{a_0}_{b_0,b}$	$P_{a_0,b_0}(T^{b}_{pos}) = \exp\left(-\left(\frac{b}{b_0}\right)^{a_0}\right)$
p_{b,c_0}	$P_{a_0,b}(T^{b,c_0}_{pos})$, p.122
k_1	$\ln\left(-\frac{\ln\left(2^{-c_1}-2^{-1}\right)}{\ln(2)}\right) \approx 0.455$
k_2	$\ln(c_1+1) \approx 0.231$
ρ	correlation coefficient

Bibliography

[Aas 2004] Aas, K. (2004) Modeling the dependence structure of financial assets: A survey of four copulas, *SAMBA/22/04*

[And 2005] Andersen, E.W. (2005) Two-stage estimation in copula models used in family studies, *Lifetime Data Analysis 11*, 333-350

[BaSt 2008] Balakrishnan, N, Stehlík, M. (2008) Exact likelihood ratio test of the scale for censored Weibull sample, *IFAS Research Paper Series 35*

[Bed 2006] Bedford, T. (2006) Copulas, degenerate distributions and quantile tests in competing risks problems, *J. Stat. Plann. Inference 136*, 1572-1587

[BCC 2009] Boudt, K., Caliskan, D., Croux, C. (2009) Robust and explicit estimators for Weibull parameters, *Metrika*, DOI: 10.1007/s00184-009-0272-1, Online

[CMMJ 2002] Cacciari, M, Mazzanti, G., Montanari, G-C., Jacquelin, J. (2002) A robust technique for the estimation of the two-parameter Weibull function for complete data sets, *Metron - International Journal of Statistics 3-4*, 64-92

[CFG 1997] Capéraà, P., Fougères,A.-L., Genest, C. (1997) A nonparametric estimation procedure for bivariate extreme value copulas, *Biometrika 84*, No.3, 567-577

[ChCh 1990] Chandra, N.K., Chaudhuri, A. (1990) On testimating the Weibull shape parameter, *Comm. in Stat. B 19*, 637-648

[CPZ 2009] Chen, J., Peng, L., Zhao, Y. (2009) Empirical likelihood based confidence intervals for copulas, *J. Multivariate Anal. 100*, No.1, 137-151

[Che 1997] Chen, Z. (1997) Statistical inference about the shape parameter of the Weibull distribution, *Stat. & Prob. Letters 36*, 85-90

[CHW 2005] Cízek, P., Härdle, W., Weron, R. (2005) *Statistical tools for finance and insurance*, Springer 2005

[Coh 1965] Cohen, C. (1965) Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples, *Technometrics 7*, No. 4, 579-588

[dHNP 2008] de Haan, L., Neves, C., Peng, L. (2008) Parametric tail copula estimation and model testing, *J. Multivariate Anal. 99*, 1260-1275

[Dix 1994] Dixit, U.J. (1994) Bayesian approach to prediction in the presence of outliers for Weibull distribution, *Metrika 41*, 127-136

[DSch 2005] Dobrić, J., Schmid, F. (2005) Testing goodness of fit for parametric families of copulas - Application to financial data, *Commun.Stat.*, Simulation Comput. 34, 1053-1068

[Dub 1966] Dubey, S.D. (1966) Some test functions for the parameters of the Weibull distribution, *Nav. Res. Logist. Q. 13*, 113-128

[DNR 2000] Durrleman, V., Nikeghbali,A., Roncalli,T. (2000) *Which copula is the right one?*, Groupe de Recherche Opérationelle, Crédit Lyonnais, *Working paper*

[Fer 2005] Fermian, J.-D. (2005) Goodness-of-fit test for copulas, *J. Multivariate Anal. 95*, No.1, 119-152

[GGR 1995] Genest, C., Ghoudi, K., Rivest, L.-P. (1995) A semiparametric estimation procedure in multivariate families of distribution, *Biometrika 82*, 543-552

[GeSe 2009] Genest, C., Segers, J. (2009) Rank-based inference for bivariate extreme-value copulas, *Ann. Statist. 37*, 2990-3022

[HeFu 1999] He, X., Fung, W.K. (1999) Method of medians for lifetime data with Weibull models, *Statist. Med. 18*, 1993-2009

[Hoff 2007] Hoff, P.D. (2007) Extending the rank likelihood for semiparametric copula estimation, *The Annals of Applied Statistics*, Vol.1, No. 1, 265-283

[HoLa 1986] Homan, S.M., Lachenbruch, P.A. (1986) Robust estimation of the exponential mean parameter for small samples: Complete and censored data, *Comm.Stat., Simulation Comput. 15*, Issue 4, 1087-1108

[Joe 1997] Joe, H. (1997) *Multivariate Models and Dependence Concepts*, Monographs on Statistics an Applied Probability 73, Chapman and Hall, London

[Ka 1996] Kahle, W. (1996) Estimation of the Parameters of the Weibull distribution for censored samples, *Metrika 44*, 27-40

[KSS 2007] Kim, G., Silvapulle, M.J., Silvapulle, P. (2007) Comparison of semiparametric and parametric methods for estimating copulas, *Computational Statistics & Data Analysis 51*, 2836-2850

[Law 2003] Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, Wiley Series in Probability and Statistics, Second Ed.

[Lee 1990] Lee, A.J. (1990) *U-Statistics. Theory and Practice*, Marcel Dekker, New York

[LeWa 2003] Lee, E. T., Wang, J. W. (2003) *Statistical Methods for survival data analysis*, Wiley 2003, New Jersey, Third Ed.

[Liu 1988] Liu, R.Y. (1988) On a notion of simplicial depth, *Proc. Nat. Acad. Sci. USA 8*5, 1732-1734

[Liu 1990] Liu, R.Y. (1990) On a notion of data depth based on random simplices, *Ann. Statist. 18*, 405-414

[MaSo 2006] Malvergne, Y., Sornette, D. (2006) *Extreme finance risks. From dependence to risk management*, Springer

[Ma 2005] Marks, N. (2005) Estimation of Weibull parameters from common percentiles, *Journal of Applied Statistics 32*, 17-24

[MiMu 2004] Mizera, I., Müller, Ch.H. (2004) Location-Scale Depth, *J.Am.Stat.Assoc. 99*, No. 468, 949-989

[Miz 2002] Mizera, I. (2002) On depth and deep points: a calculus, *The Annals of Statistics*, Vol. 30, No. 6, 1681-1736

[Mos 2002] Mosler, K. (2002) Multivariate Dispersion, Central Regions and Depth, The Lift Zonoid Approach, *Lecture Notes in Statistics 165*, Springer

[Mue 2005] Müller, Ch.H. (2005) Depth estimators and tests based on the likelihood principle with applications to regression, *J.Multivariate Anal. 95*, No.1, 153-181

[MXJ 2004] Murhty, D.N.P., Xie, M., Jiang, R. (2004) *Weibull Models*, Wiley

[Nel 2006] Nelsen, R.B. (2006) *An Introduction to Copulas*, Springer Series in Statistics, Springer, Second Ed.

[Pan 2005] Panchenko, V. (2005) Goodness-of-fit test for copulas, *Physica A 355*, 176-182

[R 2009] R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna

[Rin 2009] Rinne, H. (2009) *The Weibull Distribution- A Handbook*, CRC Press

[RH 1999] Rousseeuw, P.J., Hubert, M. (1999)Regression depth (with discussion), *J. Amer. Statist. Assoc. 94*, 388-433.

[Sac 2004] Sachs, L. (2004) *Angewandte Statistik. Anwendung statistischer Methoden*, Springer, 8. Aufl.

[SeYo 1996] Seki, T., Yokoyama, S. (1996) Robust parameter estimation using the bootstrap method for the 2-parameter Weibull distribution, *IEEE Transactions on Reliability Vol. 45*, No. 1, 34-41

[ShLa 1984] Shier, D.R., Lawrence, K.D. (1984) A comparison of robust regression techniques for estimation of Weibull parameters, *Comm. in Stat. B 13*, 743-750

[Sam 1970] Samiuddin, M. (1970) On a test for assigned value of correlation in a bivariate normal distribution, *Biometrika 57*, 461-464

[Tuk 1975] Tukey, J.W. (1975) Mathematics and the picturing of data, *Proc. International Congress of Mathematicians 2*, Canad.Math.Congress, Montreal, 523-531

[VaWe 1996] van der Vaart, A.W., Wellner, J.A. (1996) *Weak Convergence and Empirical Processes, With Applications to Statistics* (Corrected second printing), Springer

[Wei 1951] Weibull, W. (1951) *A statistical distribution function of wide applicability*, ASME Journal of Applied Mechanics, Transactions of the American Society of Mechanical Engineers, 293-297

[WMF 1995] Witting, H., Müller-Funk, U. (1995) *Mathematische Statistik II*, Teubner, Stuttgart

[WeHaMu 2009] Wellmann, R., Harmand, P., Müller, Ch.H. (2009) Distribution-free tests for polynomial regression based on simplicial depth, *J. Multivariate Anal.* 100, 622-635

[WeMu 2010] Wellmann, R., Müller, Ch.H. (2010) Tests for multiple regression based on simplicial depth, *J. Multivariate Anal.* 101, 824-838

[WoWo 1982] Wong, P.G., Wong, S.P. (1982) A curtailed test for the shape Parameter of the Weibull distribution, *Metrika 29*, 203-209

[WuTs 2006] Wu, J.-W., Tseng, H.-C. (2006) Statistical inference about the shape parameter of the Weibull distribution by upper record values, *Statistical Papers 48*, 95-129

[Yan 2007] Yan, J. (2007) Enjoy the joy of copulas: with a package copula, *Journal of Statistical Software 21*, No. 4, 1-21

[ZoSe 2000a] Zou, Y., Serfling, R. (2000) General notions of statistical depth functions, *Ann. Statist. 28*, 461-482

[ZoSe 2000b] Zou, Y., Serfling, R. (2000) Structural properties and convergence results for contours of sample statistical depth functions, *Ann. Statist. 28*, 483-499

Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig und ohne unerlaubte Hilfe angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden. Teile der Dissertation, Ausschnitte aus Kapitel 4, wurden im November 2009 zur Veröffentlichung bei der Zeitschrift **Computiational Statistics & Data Analysis** eingereicht.

Dortmund, im März 2010

Liesa Denecke