Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TECHNISCHE UNIVERSITÄT WIEN
Vienna | Austria

DISSERTATION

Statistical Inference Based on Empirical Integral Transforms

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften unter der Leitung von

Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Karl Grill
Institut für Stochastik und Wirtschaftsmathematik (E105)

eingereicht an der Technischen Universität Wien,
an der Fakultät für Mathematik and Geoinformation von

Dipl.-Ing. Tomasz Hołyński
Matrikelnummer 0427531

1. Begutachter: **Prof. Keith Knight PhD**
Department of Statistical Sciences, University of Toronto, Canada

2. Begutachter: **Prof. Apostolos Batsidis PhD**
Department of Mathematics, University of Ioannina, Greece

Wien, 09.04.2024

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Abstract

Empirical integral transform methods are powerful yet relatively little-known tools of statistical inference. They offer a framework for designing estimators and tests in parametric and non-parametric settings. The procedures typically rely on a distance measure between the transform of a model and its empirical counterpart, or employ empirical versions of some unique transform properties that hold under null hypotheses.

This work advances parametric estimation of probability distributions using their Laplace transforms and characteristic functions. Specifically, differential equations satisfied by the transforms are used to construct estimators that are both robust and explicit, while retaining comparatively high efficiency - a rare feature among diverse types of existing estimators. A method is presented for deriving the equations, enabling applications to distributions with intractable transform expressions.

The main analytical effort lies in establishing the asymptotic normality and robustness theory of the proposed estimators, with robustness examined through influence functions. Expressions for asymptotic covariance matrices and influence functions often involve intricate integrals, which appears to be the cost for achieving explicitness in the estimators themselves.

The thesis places equal emphasis on empirical evidence. Extensive simulations are conducted to compare the proposed estimators with popular robust and non-robust techniques. Various distribution types are considered, including symmetric and skewed ones, with light and heavy tails, further in the presence of outliers and model misspecifications. The combination of experiments and theoretical analysis reveals a crucial finding: an optimal trade-off between efficiency and robustness of the estimators, along with their numerical reliability, can be consistently achieved by pre-estimating the scale of the estimators' weight function from the sample. The author contends that this aspect has been overlooked in early constructions of transform-based and other minimum distance estimators relying on weighted integrated distances.

The transform methods, especially based on the differential-equations, moreover enable inference for a variety of non-standard distributions. Prominent instances include mixed, compound, and non-normalized distributions encountered across diverse application fields. In addition, families of distributions, such as the Pearson or Katz family, are often characterized by differential equations, either in the variable's or transform domains. Therefore, the presented methods support estimation and identification within

entire families. While non-standard models and the families are not our primary focus, numerous examples are given to demonstrate a wider spectrum of applications and motivate further research.

A separate chapter is dedicated to goodness-of-fit testing, introducing a novel test for the log-normal distribution based on the Laplace transform. In particular, the procedure utilizes a functional differential equation satisfied by the transform. The test compares well in terms of power with several famous tests (e.g., Shapiro-Wilk, Jarque-Bera, Anderson-Darling). Importantly, it tends to be uniformly powerful across distributional alternatives and remains quite stable concerning its tuning parameter's value. Aspects such as consistency, asymptotic distribution of the test statistic, and bootstrap-based determination of critical points, are addressed.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Kurzfassung

Empirische Integraltransformationstechniken sind leistungsstarke, jedoch vergleichsweise wenig bekannte Werkzeuge der statistischen Inferenz. Sie bieten einen Rahmen für die Entwicklung von Schätzern und Tests in parametrischen und nicht-parametrischen Szenarien. Die Verfahren stützen sich in der Regel auf ein Distanzmaß zwischen der Transformation eines Modells und ihrer empirischen Entsprechung oder verwenden empirische Versionen einiger Transformationseigenschaften, die unter der Nullhypothese gelten.

Diese Arbeit zielt darauf ab, die parametrische Schätzung voranzutreiben, indem sie speziell die Differentialgleichungen nutzt, die von den Transformationen von Wahrscheinlichkeitsverteilungen erfüllt werden, insbesondere der Laplace-Transformation und der charakteristischen Funktion. Der Ansatz liefert Schätzer, die sowohl robust als auch explizit sind und dabei eine relativ hohe Effizienz beibehalten - eine seltene Eigenschaft unter verschiedenen Arten von bestehenden Schätzern. Es wird eine Methode zur Ableitung der Gleichungen präsentiert, die Anwendungen auf Verteilungen ermöglicht, deren Transformationen keine Ausdrücke in geschlossener Form aufweisen.

Unser primärer analytischer Fokus liegt darauf, die asymptotische Normalität und Robustheitstheorie der vorgeschlagenen Schätzer zu etablieren, wobei die Robustheit durch Einflussfunktionen untersucht wird. Ausdrücke für asymptotische Kovarianzmatrizen und Einflussfunktionen beinhalten oft aufwendige Integrale, was als Preis für die Explizitheit der Schätzer selbst erscheint.

Die Dissertation legt gleichen Wert auf empirische Evidenz. Umfangreiche Simulationen werden durchgeführt, um die Leistung der vorgeschlagenen Schätzer mit bekannten robusten und nicht-robusten Schätztechniken zu vergleichen. Verschiedene Verteilungstypen werden berücksichtigt, darunter symmetrische und schief verteilte, mit leichten und schweren Rändern, außerdem in Anwesenheit von Ausreißern und bei Modellspezifikationsfehlern. Die Kombination von Experimenten und theoretischer Analyse enthüllt eine wichtige Erkenntnis: Ein optimaler Ausgleich zwischen Effizienz und Robustheit sowie die numerische Zuverlässigkeit der Schätzer können durch vorherige Schätzung der Skalierung der Gewichtsfunktion der Schätzer konsequent erreicht werden. Der Autor räumt ein, dass dieser Aspekt bei früheren Konstruktionen von transformationsbasierten und anderen Minimum-Abstands-Schätzern, die auf gewichteten integrierten Abständen beruhen, übersehen wurde.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die Transformationsmethoden, insbesondere jene, die auf Differentialgleichungen basieren, ermöglichen zudem Schlussfolgerungen für verschiedene weniger gebräuchliche Verteilungen. Prominente Beispiele umfassen gemischte, zusammengesetzte und nichtnormalisierte Verteilungen, die in verschiedenen Anwendungsbereichen vorkommen. Darüber hinaus werden Verteilungsfamilien wie die Pearson- oder Katz-Familie oft durch Differentialgleichungen charakterisiert, entweder im Verteilungs- oder im Transformationsbereich. Daher unterstützen die vorgestellten Methoden Schätzungen und Identifikationen innerhalb ganzer Familien. Obwohl nicht-standard Modelle und Familien nicht unser Hauptaugenmerk sind, werden zahlreiche Beispiele präsentiert, um ein breiteres Anwendungsspektrum zu veranschaulichen und weitere Forschung zu motivieren.

Ein eigenes Kapitel ist Anpassungstests gewidmet und führt einen neuartigen Test für die log-normale Verteilung ein. Dieser Test schneidet in Bezug auf die Teststärke gut ab im Vergleich zu mehreren etablierten Tests (z. B. Shapiro-Wilks, Jarque-Bera, Anderson-Darling). Besonders wichtig ist, dass er dazu neigt, gleichmäßig leistungsstark über verschiedene Verteilungsalternativen zu sein und stabil in Bezug auf den Wert seines Einstellparameters bleibt. Aspekte wie Konsistenz, asymptotische Verteilung und bootstrap-basierte Bestimmung von kritischen Werten werden behandelt.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Acknowledgements

The research contained in and related to this thesis could not have emerged without the support of several persons, whom I would like to thank sincerely.

Foremost among them is my supervisor, Professor Karl Grill from the Institute of Statistics and Mathematical Methods in Economics at the Vienna University of Technology. His guidance, expertise, and encouragement have been invaluable throughout this endeavor. While affording me much freedom in shaping the topic, he has rigorously overseen the work, attended to crucial details, and helped with intricacies. I am grateful for the time and patience he has always offered me, for our numerous meetings, talks, and consultations.

Further, I want to express my deep gratitude to the external reviewers of this thesis: Professor Keith Knight from the University of Toronto, Canada, and Professor Apostolos Batsidis from the University of Ioannina, Greece. I am indebted for their interest and care in reviewing the text, as well as for their comments, corrections, and constructive feedback.

Empirical transform methods is a specialized research area, exercised by a narrow group of mathematical statisticians. Contacts with them were precious and motivating. I owe this networking opportunity to Professor Simos Meintanis from the National and Kapodistrian University of Athens. I thank him for inviting me to focused sessions on non-parametric and goodness-of-fit problems. The two conferences held in the beautiful Italian cities of Salerno and Trento (see Bibliography) allowed me to present my contributions and engage with the community.

This thesis also inscribes itself into the realm of robust statistics. I am grateful to an authority in this field, Professor Peter Filzmoser, head of the Computational Statistics Research Unit in our institute, for evaluating my early proposals on robust estimators at his seminar, and for his important hints.

To me, no research can be fruitfully done without friends around — dear people who accompany us, discuss, and exchange ideas. Closest to my work, in terms of location, were my three companions from the Faculty of Mathematics and Geoinformation: Łukasz Fabrykowski, Lucas Fertl, and Dogan Günes. Each brought his own unique perspective on probability and statistics. We have spent many moments together, and their camaraderie and insights have been priceless. Thank you, guys!

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

A separate mention goes to my long-time friend in Vienna, Jacek Haneczok. I am very fortunate to have met him during my post-graduation years. Jacek, a gifted mathematician, machine learning specialist, and versatile thinker, accompanied me in my gradual transition from probability modeling to the domain of data science. Many venues in Vienna witnessed our lively discussions on diverse topics, too numerous to name, from which I always benefited greatly. What united us was our shared dedication to methodology. I want to thank Jacek especially for our collaboration on minimum distance estimators in 2018, which I am pleased to cite in this thesis.

The most intense phase of my work coincided with the Covid-19 pandemic, overshadowing university life and isolating scholars. It would have been much harder to go through it without the company of my other dear friends: Monika Marczak, Izabela Wojdanowska, Judyta Woźniak, and Wojciech Tarkowski. Their presence and keeping in touch were inestimable during the hard lockdowns, and I deeply appreciate their friendship beyond.

Finally, last but definitely not least, my heartfelt gratitude goes to my whole family, simply for their love and support throughout my academic and research years.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Erklärung zur Verfassung der Arbeit

Dipl.-Ing. Tomasz Hołyński
Johannagasse 42/15
1050 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 9. April 2024

Tomasz Hołyński

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Contents

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

CHAPTER 1

Introduction

Let $(\Omega,\ \mathcal{F},\ P_{\boldsymbol{\theta}})$ be a family of probability spaces indexed by a vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$. Let X be a random variable defined on one (but unspecified) member of this family, described by the distribution function $F_{\boldsymbol{\theta}}(x) = P_{\boldsymbol{\theta}}(X \leq x)$.

One of the fundamental goals of mathematical statistics is to estimate $\boldsymbol{\theta}$ from a random sample $X_1, ..., X_n$ of independent copies of X, assuming a specific distribution[1]. This is accomplished by computing a statistic $\widehat{\boldsymbol{\theta}}_n = T_n(X_1, ..., X_n)$, called an *estimator* of $\boldsymbol{\theta}$. A closely related question involves determining whether a sample $X_1, ..., X_n$ stems from a member of the postulated distributional family $F_{\boldsymbol{\theta}}(x)$ for *some* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Testing such a hypothesis is referred to as a *goodness-of-fit test* (GoF). In practice, the test often includes the estimation of $\boldsymbol{\theta}$ because its value is unknown a priori.

Many methods for deriving estimators rely on the aforementioned distribution function or the associated density $f_{\boldsymbol{\theta}}(x)$. Specifically, the *maximum likelihood estimator* (ML) is the value of parameter that maximizes the joint density of an i.i.d. sample, that is

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_n^{\mathrm{ML}} &= \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \prod_{j=1}^{n} f_{\boldsymbol{\theta}}(X_j) \\ &= \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{j=1}^{n} \log f_{\boldsymbol{\theta}}(X_j)\,. \end{aligned} \tag{1.1}$$

[1]The mode of statistical inference in which a parametric model of a distribution (or other function of data) is assumed to be known before estimation is termed the *Fisherian paradigm*, due to the famous British statistician R.A. Fisher (1890-1962). In turn, the paradigm that treats the parameters as constants is referred to as *frequentism*, contrasting with the *Bayesian paradigm* that treats them as random variables. While this thesis avoids delving into the philosophical foundations of statistics, it is crucial to clarify that our work is exclusively embedded within the Fisherian and frequentist paradigms. For in-depth comparisons of these and other paradigms see Barnett (1999) or Spanos (2019).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

In turn, the *minimum distance estimators* (MD) are constructed as minimizers of various integral divergences involving $F_{\boldsymbol{\theta}}(x)$ or $f_{\boldsymbol{\theta}}(x)$ and their non-parametric estimates, denoted by $\widehat{F}_n(x)$ and $\widehat{f}_n(x)$; see Cao et al. (1995), Henze and Klar (2002), Scott (2001) or Basu et al. (2011). One well-known example of this class is the minimum Cramér-von Mises distance estimator expressed as

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{CM}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \int \Big[F_{\boldsymbol{\theta}}(x) - \widehat{F}_n(x)\Big]^2 \mathrm{d}F_{\boldsymbol{\theta}}(x)\,, \tag{1.2}$$

where $\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \leq x)$ represents the empirical distribution function. Other types of the distribution-based estimators encompass the *maximum spacing estimators*, introduced by Ranneby (1984), and estimators the leverage certain distributional characterizations, see e.g. Betsch et al. (2021).

Yet other popular estimators rest on partial characteristics of the distributions. For example, the *method-of-moments estimator* (MM), denoted by $\widehat{\boldsymbol{\theta}}_n^{\mathrm{MM}}$, is defined as the parameter value that equates the first p theoretical and sample moments of X. That is, $\widehat{\boldsymbol{\theta}}_n^{\mathrm{MM}}$ is the solver of

$$\begin{cases} \mathbb{E}_{\boldsymbol{\theta}}[X] = \displaystyle\int x \,\mathrm{d}F_{\boldsymbol{\theta}}(x) = \frac{1}{n}\sum_{j=1}^{n} X_j\,, \\ \vdots \\ \mathbb{E}_{\boldsymbol{\theta}}[X^p] = \displaystyle\int x^p \,\mathrm{d}F_{\boldsymbol{\theta}}(x) = \frac{1}{n}\sum_{j=1}^{n} X_j^p\,, \end{cases} \tag{1.3}$$

where $p = \dim(\boldsymbol{\theta})$. Extensions and modifications of this approach, including the *generalized method of moments* and *probability-weighted moments*, as well as similar techniques matching the quantiles have been developed, especially in econometrics and risk management; see, e.g., Hansen (1982), Hosking et al. (1985), Hosking and Wallis (1987), McNeil et al. (2015).

All the distribution-based estimators come with their pros and cons in terms of precision (bias and variance), robustness, and computational costs. For example, the ML estimators are often preferred for their asymptotic optimality and, quite frequently, for their explicit expressions. The MM estimators, while also often explicit, tend to have higher variances and be less efficient asymptotically. In contrast, the MD and quantile-based estimators are nearly never explicit and necessitate numerical procedures. However, they are typically robust against data contamination and outliers.

An alternative and fairly general estimation method, being subject of this thesis, is offered by *integral transforms* of the probability distributions. Such (parametric) transforms are defined generally through the following Lebesque-Stieltjes integral:

$$\mathcal{T}_{\boldsymbol{\theta}}(s) = \int K(s,x)\mathrm{d}F_{\boldsymbol{\theta}}(x)\,, \tag{1.4}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

where $K(s, x)$ is a real or complex-valued *kernel*, and s is the *transform variable*. The interval of s-values $\mathcal{D} \subseteq \mathbb{R}$ for which the integral converges (i.e. is finite) is called the *interval of convergence*.

The three most popular choices for the kernels are

- $K(s, x) = e^{-sx}$, yielding the *Laplace transform* (LT);
- $K(s, x) = e^{isx}$, $(i^2 = -1)$, yielding the *characteristic function* (CF);
- $K(s, x) = s^x$, yielding the *probability generating function* (PGF).

These three transforms[2] are primarily powerful tools in probability theory, each with slightly different domain of application[3]. Firstly, as known from probability textbooks, they are used to prove fundamental convergence theorems for sums of random variables; see e.g. Feller (1968) and Feller (1971). Secondly, they aid in solving diverse distributional problems that arise in the algebra of random variables, as seen in Springer (1979), and in more complex stochastic models such as counting, compound and mixed distributions, queueing systems, flow-graphs, and risk models. Numerous related references include Cox (1970), Kleinrock (1975), Kleinrock (1976), Prabhu (1980), Grandell (1997), Rolski et al. (1999), Huzurbazar (2005) or Trivedi (2016). Noteworthy applications can also be found in the field of mathematical statistics. Specifically, the famous saddle-point approximation, which provides approximations for distributions of estimators and test statistics in finite samples, is based on LT; see Field and Ronchetti (1990) or Butler (2007). In turn, CF, in form of the Fourier transform, enables the spectral analysis of time series; see, e.g., Box and Jenkins (1970), Madsen (2007).

In all these applications, the transforms play a highly important yet auxiliary role. However, in this thesis, we employ the transforms as the *direct tool* of statistical inference. To

[2]Note that the transform qualitatively identical with LT (with kernel $K(s, x) = e^{sx}$) is referred to in the literature as the *moment generating function* (MGF). In turn, the transform with $K(s, x) = e^{-isx}$ is widely known in probability and engineering as the *Fourier transform*. Also, any r-th raw moment of a random variable X also fits within the framework of transforms with kernel $K(r, x) = x^r$. Putting $r = s \in \mathbb{R}$ gives the *Mellin transform*, applied in studying products of independent random variables; see Springer (1979). Transforms with non-exponential kernels, such as the *Hilbert* or *Laguerre transform*, are covered in Ditkin and Prudnikov (1965), Sumita and Kijima (1988) and Poularikas (2010).

[3]As discussed in more detail later, the CF is the most universal transform, existing for any distribution for all $s \in \mathbb{R}$. In contrast, the LT and PGF, used for either continuous or discrete distributions, have typically a restricted interval of existence or even may not exist for some two-sided distributions with heavy tails, e.g. the Cauchy distribution. Interested readers are referred to excellent sources on theory of these three integral transforms that go beyond the scope of this thesis, including Widder (1959), Feller (1968), Lukacs (1970), Feller (1971), Kawata (1972), Ushakov (1999), Abate and Whitt (1996). A large body of literature is devoted to the important problem of transform inversion; for a comprehensive overview see Cohen (2007), and for applications in probability refer to Abate and Whitt (1992) and Abate and Whitt (1995).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

achieve this, the notion of an *empirical integral transform* of a probability distribution is required. Given a random sample $(X_1, \dots, X_n)$, it is defined as

$$\widehat{\mathcal{T}}_n(s) = \int K(s,x) \mathrm{d}\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} K(X_j, s)\,, \tag{1.5}$$

being a non-parametric and unbiased estimator of the model version (1.4).

With the model transform and its empirical counterpart, the *transform-based estimator* of a distribution parameter $\boldsymbol{\theta}$ is defined, in its most general form, as

$$\widehat{\boldsymbol{\theta}}_n^{\mathcal{T}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ \rho\Big(\mathcal{T}_{\boldsymbol{\theta}}(s), \widehat{\mathcal{T}}_n(s)\Big)\,, \tag{1.6}$$

where $\rho(\cdot,\cdot) \mapsto \mathbb{R}_0^+$ is a suitable divergence measure in s-domain. This measure can be computed over an interval of $\mathcal{D}$ (as an integral) or over a finite grid of points $s_1, \dots, s_k \in \mathcal{D}$ (as a sum).

Statistical problems that motivate the transform-based estimators include (but are not limited to) the three following situations:

1. The density and/or distribution function is unknown but, instead, the corresponding transform is available. This commonly arises in financial and actuarial statistics, particularly with stable and compound distributions. Relevant estimators were developed, among others, by Press (1972), Paulson et al. (1975), Koutrouvelis (1980), Csörgő (1984), Epps and Pulley (1985), Koutrouvelis and Meintanis (2002), Braun et al. (2008), Marcheselli et al. (2008) and Van Zyl (2016).

2. The density is known but the likelihood has singularities and the ML estimator breaks down. The best known example of this situation occurs in estimating finite mixtures of Gaussian distributions; see Quandt and Ramsey (1978), Schmidt (1982), Titterington et al. (1985), Besbeas and Morgan (2003) and Xu and Knight (2011) for several variants of transform-based estimators proposed in this scenario.

3. The support of the distribution depends on the estimated parameter. In this case the ML estimator may not be asymptotically optimal whereas the MD estimators, like (1.2), are hard to implement due to unknown limits of integration; see Koutrouvelis and Canavos (1997) and Koutrouvelis et al. (2005).

1.1 Motivation and objectives of the thesis

Naturally, the transform-based estimators also apply to the standard distributions for which the traditional approaches (likelihood-, moment- and quantile-based) are fully feasible. One motivating factor is their robustness to outliers. As a result, researchers have examined this quality and others in more general settings, by considering the type of transform (LT/CF/PGF) and functional forms of the divergence $\rho(\cdot,\cdot)$; see Heathcote

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

(1977), Feuerverger and McDunnough (1984), Campbell (1993) and Jiménez-Gamero et al. (2016) for the CF- and LT-based estimators, and Chergui (1996), Dowling and Nakamura (1997), Sharifdoust et al. (2016), Jiménez-Gamero and Batsidis (2017) and Hołyński (2019a) for the PGF-based ones.

However, a wider use of the transform-based estimators seems to inhibited by computational problems which we now explain. Namely, most of them (considered so far in literature) have form of weighted $\mathcal{L}^2$-type estimators that can be written as

$$\widehat{\boldsymbol{\theta}}_n^{\mathcal{T}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \int\limits_{\mathcal{D}'} \left|\mathcal{T}_{\boldsymbol{\theta}}(s) - \widehat{\mathcal{T}}_n(s)\right|^2 \mathrm{d}W(s)\,, \tag{1.7}$$

where $\mathcal{D}' \subseteq \mathcal{D}$ and $W(s)$ is a weight function that makes the integral finite and controls the estimator's properties. Note that very rarely this $\mathcal{L}^2$ distance is explicitly computable so that the estimates (as the minimizers) could be obtained in closed forms. The problem is even worse when the transform itself has no tractable analytical expression. And this happens quite often, for example for the popular distributions with non-exponential tails like Pareto, Weibull, Fréchet or log-normal, which have intractable LT and CF; see, e.g., Nadarajah and Kotz (2006), Penson and Górska (2014) or Asmussen et al. (2016). Computing the estimates turns then into an intensive and nested procedure, in which both the transform and the distance must be computed by numerical integrations for every iteration of the optimization loop. All this makes the estimators of type (1.7) computationally demanding, prone to numerical errors (particularly at the boundaries of the parameter space and/or in the presence of outliers), difficult to re-sample (whether for bootstrap or jacknife) or and challenging to simulate extensively by Monte Carlo method.

The computational efforts may be smaller if for $W(s)$ we take a stepwise function mapping $\mathcal{D}'$ to [0,1] and having k (equal or unequal) steps at several points $s_1, ..., s_k$. Then, (1.7) becomes a weighed least-squares type estimator. In addition, if one takes $k = p = \dim(\boldsymbol{\theta})$, computation of $\widehat{\boldsymbol{\theta}}_n^{\mathcal{T}}$ reduces to solving for $\boldsymbol{\theta}$ the system of equations

$$\begin{cases} \mathcal{T}_{\boldsymbol{\theta}}(s_1) = \widehat{\mathcal{T}}_n(s_1)\,, \\ \vdots \\ \mathcal{T}_{\boldsymbol{\theta}}(s_p) = \widehat{\mathcal{T}}_n(s_p)\,. \end{cases} \tag{1.8}$$

The resultant M-estimator can be called the *transform-matching estimator* since it matches (equates) the model- and the empirical transform at a set of points. However, even if a tractable expression for the transform is available, then, typically, the above system also requires a numerical solution due to non-linearity of $\mathcal{T}_{\boldsymbol{\theta}}(s)$ in $\boldsymbol{\theta}$.

One of the main goals of this thesis is to alleviate these computational problems by developing transform-based estimators which admit explicit expressions for a wide range of standard and derived distributions. For this purpose, instead of using the direct $\mathcal{L}^2$ distances as in (1.7), we base the estimators on differential equations satisfied by specific

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

transforms. To illustrate the approach, suppose that a transform satisfies the first-order differential equation

$$\mathcal{T}_{\boldsymbol{\theta}}(s) + h(\boldsymbol{\theta}, s)\mathcal{T}_{\boldsymbol{\theta}}'(s) = 0\,, \tag{1.9}$$

where $\mathcal{T}_{\boldsymbol{\theta}}'(s)$ denotes derivative w.r.t. to s and $h(\boldsymbol{\theta}, s)$ is a known function. Replacing the transform and the derivative by their sample versions, $\widehat{\mathcal{T}}_n(s)$ and $\widehat{\mathcal{T}}_n'(s) = \frac{1}{n}\sum_{j=1}^{n} K'(X_j, s)$, we obtain $\widehat{\mathcal{T}}_n(s) + h(\boldsymbol{\theta}, s)\widehat{\mathcal{T}}_n'(s) \approx 0$. The $\mathcal{L}^2$ estimator may be now defined as the minimizer of the left side of such empirical differential equation, namely

$$\widehat{\boldsymbol{\theta}}_n^{\mathcal{T}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \int\limits_{\mathcal{D}'} \left|\widehat{\mathcal{T}}_n(s) + h(\boldsymbol{\theta}, s)\widehat{\mathcal{T}}_n'(s)\right|^2 \mathrm{d}W(s)\,. \tag{1.10}$$

Thanks to the differential equation the possibly complicated or unavailable expression for $\mathcal{T}_{\boldsymbol{\theta}}(s)$ is not needed any more. Inference about the parameter is now conducted solely through the function $h(\boldsymbol{\theta}, s)$ - typically having a simpler form. Moreover, if $h(\boldsymbol{\theta}, s)$ is a linear function of the elements of $\boldsymbol{\theta}$, or it is a quotient of such functions, then, upon appropriate choice of $W(s)$, the estimator admits explicit formulas. Similar considerations hold for the 'transform-matching' version of (1.10), that is for the M-estimator defined by

$$\begin{cases} \widehat{\mathcal{T}}_n(s_1) + h(\boldsymbol{\theta}, s_1)\widehat{\mathcal{T}}_n'(s_1) = 0\,, \\ \vdots \\ \widehat{\mathcal{T}}_n(s_p) + h(\boldsymbol{\theta}, s_p)\widehat{\mathcal{T}}_n'(s_p) = 0\,. \end{cases} \tag{1.11}$$

It should be acknowledged that the use of differential equations of transforms is not novel in statistical applications. For example, the integrated statistics based on such equations have already been used in several goodness-of-fit tests, as seen in Henze and Klar (2002), Henze et al. (2012), Batsidis et al. (2020) and Goffard et al. (2022). Also, estimating equations of type (1.11), employing PGF as the transform, have been proposed in Kemp and Kemp (1988) and Dowling and Nakamura (1997) for the estimation of discrete distributions. However, the two latter papers are rather limited in studying the estimators' properties and simulations. Importantly, they did not cover the general case of (1.10) with continuous weight function $W(s)$. Notably, there has been neither follow-up research concerning continuous distributions in terms of LTs and CFs, which we focus on in the present work.

An important contribution of this thesis also lies in how the differential equations (and, consequently, the estimators) are derived. All the papers cited herein utilized the differential equations only for distributions with explicitly known transforms, often expressed in terms of elementary functions. In those cases, the equations could readily be found by differentiating the transform expression w.r.t. s. But, among the continuous distributions, this is possible only for several such as the normal, Cauchy, gamma, inverse

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Gaussian and logistic distributions. Here we go beyond this limited collection. Motivated by the proof in Gaunt (2021)[4], we present a method for deriving the equations when transform expressions are unknown in closed forms, such as those for beta of 1st and 2nd kind, Pareto, Rayleigh, Maxwell and distributions of the inverted variables. Equations obtained using this method can have order higher than one, may involve more than two derivatives, and may be non-homogenous. To apply these equations for statistical purposes, we must examine uniqueness of their solutions, as this aspect will condition consistency of the resulting estimators and tests.

The primary focus of this thesis is on continuous distributions estimated or tested using LTs and CFs. Due to space limitations, we do not cover inference for the discrete laws by means of the PGF transform. Nevertheless, it is worth emphasizing that during preparatory research some PGF-based estimators were also derived and demonstrated promising results in simulations. Besides, one should keep in mind that a PGF-based differential equation (if available) can be converted into the LT- or CF-based forms, by applying the change of variables $s = e^{-z}$ or $s = e^{iz}$. Thus, the presented theory extends to discrete scenarios as well.

The main analytical effort of the work is to prove the consistency and asymptotic normality of the new estimators in the $\mathcal{L}^2$ settings. The proofs require asymptotic results from the theory of V- and U-statistics. Upon obtaining the asymptotic variances or covariance matrices of the estimators, we can compute their asymptotic relative efficiencies w.r.t. to the ML and other estimators.

Our second motivation for pursuing the new estimators is their robustness to outliers. We will derive their influence functions which are the most popular theoretical measures of robustness; see Hampel et al. (1986), Jurečková et al. (2019). Despite being in complicated integral forms, these functions exhibit a crucial characteristic - they decay to zero at infinity. This feature is ideally required for robust estimators. In contrast, the influence functions of the already studied transform-based $\mathcal{L}^2$ estimators of type (1.7) are bounded but do not fall to zero at infinity, as shown by Campbell (1993).

Having obtained the influence functions, we can study the *robustness-efficiency tradeoffs* of the proposed estimators. Interestingly, we discover that, for specific distributions, their high robustness can be often coupled with significant efficiency. However, achieving this desired quality uniformly across the parameter space requires the weight function $W(s)$ to be driven by the sample. We discuss how to control this function by pre-estimating its scale parameter. In the author's view, the necessity of the data-driven tuning of the transform-based estimators was possibly overlooked in the early studies such as Heathcote (1977) or Campbell (1993). Our objective is to investigate thoroughly this issue by both analysis and simulation.

An important yet often neglected practical aspect of parameter estimators is their behavior under model misspecification. This occurs when estimation is performed despite

[4]Available as preprint since December 2019.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

the assumed distribution $F_{\boldsymbol{\theta}}(x)$ is different from the true one that generates the sample. Unfortunately, assessing the impact of wrong assumptions in analytical way is very hard, except for simple functionals like the sample mean or median. Also in our case the complexity of the proposed estimators precludes such analyses, even in the asymptotic conditions. We therefore resort to Monte Carlo simulations to examine performances under misspecification.

Goodness-of-fit testing

The latter, shorter portion of the thesis is devoted to goodness-of-fit (GoF) testing. The classical GoF tests utilize the distribution function $F_{\boldsymbol{\theta}}(x) = F(x|\boldsymbol{\theta})$. The test statistic, say T_n, takes usually the form of a divergence between the empirical distribution $\widehat{F}_n(x)$ and its parametrically estimated version $F(x|\widehat{\boldsymbol{\theta}}_n)$, where $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator of the parameter. A large realized value of T_n rejects the hypothesis that given sample originates from $F(x|\boldsymbol{\theta})$; see D'Agostino and Stephens (1986) and Thas (2010).

The three most popular distribution-based GoF procedures are the *Kolmogorov-Smirnov* (KS), the *Cramér-von Mises* (CM) and the *Anderson-Darling* (AD) tests. Their test statistics are given by

$$T_n^{\mathrm{KS}} = n \sup_x \left| F(x|\widehat{\boldsymbol{\theta}}_n) - \widehat{F}_n(x) \right| , \tag{1.12}$$

$$T_n^{\mathrm{CM}} = n \int \left[F(x|\widehat{\boldsymbol{\theta}}_n) - \widehat{F}_n(x) \right]^2 \mathrm{d}F(x|\widehat{\boldsymbol{\theta}}_n) , \tag{1.13}$$

$$T_n^{\mathrm{AD}} = n \int \frac{[F(x|\widehat{\boldsymbol{\theta}}_n) - \widehat{F}_n(x)]^2}{F(x|\widehat{\boldsymbol{\theta}}_n)[1 - F(x|\widehat{\boldsymbol{\theta}}_n)]} \mathrm{d}F(x|\widehat{\boldsymbol{\theta}}_n) . \tag{1.14}$$

The transform-based GoF tests, like the transform-based estimators, employ either the direct $\mathcal{L}^2$ distances or the $\mathcal{L}^2$ norms of the empirical differential equations. That is, the normalized test statistics are

$$T_n = n \int_{\mathcal{D}'} \left| \mathcal{T}(s|\widehat{\boldsymbol{\theta}}_n) - \widehat{\mathcal{T}}_n(s) \right|^2 \mathrm{d}W(s) , \tag{1.15}$$

or, for example,

$$T_n = n \int_{\mathcal{D}'} \left| \widehat{\mathcal{T}}_n(s) + h(\widehat{\boldsymbol{\theta}}_n, s) \widehat{\mathcal{T}}_n'(s) \right|^2 \mathrm{d}W(s) . \tag{1.16}$$

The role of the weight function $W(s)$ in the testing context is to control the power of the procedure against specific distributional alternatives. In case of using LT and CF, the choice of $W(s)$ relies on the well-known Tauberian theorems saying that the behavior of the transform $\mathcal{T}(s)$ as $s \to 0$ reflects that of $F(x|\boldsymbol{\theta})$ as $x \to \infty$, and vice versa. In addition, $W(s)$ should be chosen so that T_n has a closed-form expression.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The transform-based tests serve as flexible omnibus procedures and have already been proposed for many continuous and discrete laws. The CF-based tests, utilizing the statistic of type (1.15) trace back to the normality test by Epps and Pulley (1983). Later, Gürtler and Henze (2000), Matsui and Takemura (2005) and Pudełko (2012) continued the approach to test the Cauchy hypothesis, while Epps (2005) and Jiménez-Gamero et al. (2009) extended the method to various location-scale families, incorporating distributions such as exponential, uniform, Laplace, logistic, and extreme-value. Along the same lines, tests have beed constructed using LT, basing either on (1.15) or on (1.16), for the following distributions: normal by Zghoul (2010) and Henze and Koch (2020), exponential by Baringhaus and Henze (1991) and Henze (1992), inverse Gaussian by Henze and Klar (2002), gamma by Henze et al. (2012) and continuous compound Poisson by Goffard et al. (2022). Additionally, the PGF-based tests with (1.15) or (1.16) were developed by Rueda et al. (1991), Baringhaus and Henze (1992), Meintanis (2003), Meintanis (2008), Jiménez-Gamero and Alba-Fernández (2019) and Batsidis et al. (2020) for the following discrete laws: Poisson, generalized Poisson, negative binomial, Poisson-Tweedie and Bell.

Simulations conducted in the mentioned papers show that, very often, the empirical power of the transform-based tests is higher than that of the KS, CM or AD counterparts. Encouraged by this fact, we also develop a new LT-based test for the log-normal distribution. Its density is expressed by

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \qquad x, \sigma > 0,\ \mu \in \mathbb{R}, \tag{1.17}$$

where $\boldsymbol{\theta} = (\mu, \sigma)^\top$ stands for the mean and standard deviation of the underlying normal variable.

This test has been hindered by the absence of tractable expression for the associated LT or CF; see Asmussen et al. (2016). In fact, our test exploits a *functional differential equation* satisfied by the transform, which may be written in general as

$$\mathcal{T}_{\boldsymbol{\theta}}(s \cdot g(\boldsymbol{\theta}, s)) + h(\boldsymbol{\theta}, s)\mathcal{T}_{\boldsymbol{\theta}}'(s) = 0\,. \tag{1.18}$$

Equations of this form are obviously not in the class of ordinary differential equations which we will use to obtain the estimators. In particular, Leipnik (1991) showed that for the log-normal distribution the LT of (1.17) satisfies

$$L(se^{\sigma^2}) + \exp\{-\mu - \sigma^2/2\}L'(s) = 0\,, \tag{1.19}$$

and so $h(\boldsymbol{\theta}, s) = e^{-\mu-(\sigma)^2/2}$ and $g(\boldsymbol{\theta}, s) = e^{\sigma^2}$.

The proposed procedure is based on the weighted $\mathcal{L}^2$ statistic of the type (1.16) that employs sample version of (1.19) and is derived in a closed form. We discuss the aspects of the test including consistency, distribution of the statistic under null-hypothesis and bootstrap-based estimation of the critical points. We conduct finite-sample simulations where we compare it with the CM and AD tests as well with the famous Shapiro-Wilk and Jarque-Bera tests for normality applied to log-transformed samples.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

1.2 Organization of the thesis

The remainder of the thesis consists of five chapters, structured as follows. Chapter 2 provides an overview analytical and stochastic properties of LT and CF and their empirical counterparts. While this constitutes a broad subject, we focus only on the properties relevant to statistical inference. The chapter also includes a brief introduction to robustness and influence functions in the context of transform functionals.

The subject of Chapter 3 is differential equations satisfied by LTs and CFs of various probability distributions. This chapter explains how to derive the equations for particular standard distributions, with special attention given to cases where the transforms are not known explicitly. The aforementioned alternative method of differentiation, necessary to handle these cases, is described and illustrated. Subsequently, the chapter addresses the requirements related to the uniqueness of solutions for the equations, which, in a slightly different manner, govern the consistency of resulting estimators and goodness-of-fit tests. Finally, the chapter explores several examples of non-standard (mixed and compound) distributions to demonstrate the applicability of the differential equation approach in broader modeling contexts.

Chapter 4 first deals with the asymptotic and robustness theories of the proposed $\mathcal{L}^2$ estimators. The related theorems, which prove the asymptotic normality (with specific forms of covariance matrices) and establish the influence functions, pertain to three specific forms of the differential equations of LTs and CFs. In the remaining and most extensive portion of the chapter, we apply the theory to the estimators derived for five standard distributions (normal, exponential, Rayleigh, gamma, Pareto) and one mixed distribution (normal variance-gamma). These distributions are carefully selected to cover: (i) estimators based on all the three forms of differential equations considered in the proofs, (ii) various families of distributions, including two two-sided location-scale family, two one-sided scale families, and two one-sided scale-shape families, (iii) distributions whose transform is known explicitly and/or given through its differential equation, (iv) a distribution whose density is unknown but the associated transform is available, and (v) robustness against two kinds of outliers (large and small). Closed-form expressions are derived for nearly all of the studied estimators. A detailed robustness-efficiency analysis is provided and a data-driven tuning of the weight function is proposed in each case. Also, for each distribution, extensive simulations and comparisons with traditional robust and non-robust estimators are conducted. For the normal and gamma distributions, we also study the impact of model misspecification on our estimators and their popular competitors. The wealth of empirical evidence provides a foundation for drawing detailed, practice-oriented findings and recommendations which close this chapter.

Chapter 5 presents the GoF test for the log-normal distribution which was already outlined above.

Chapter 6 concludes the thesis, providing a high-level summary of key observations and comparisons between the proposed methods and conventional approaches. It outlines as-

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

pects that demand further analysis. Finally, it discusses statistical models and problems not covered in this thesis, to which, however, the differential equations of transforms can be applied, at least potentially. These include estimation of discrete and non-normalized distributions, estimation within the Pearson's family of distributions, regression models, along with multivariate and non-parametric problems. Some further research directions are suggested, with consideration to literature.

The thesis also includes four appendices. Appendix A summarizes important results from the theory of V- and U-statistics, primarily utilized in Chapters 4 and 5. Appendix B outlines the derivations of various formulas and expressions obtained in this thesis. Appendix C compiles the improper definite integrals used throughout the text. Appendix D enumerates mathematical symbols and abbreviations. Most of the standard probability distributions studied or mentioned in the thesis are consolidated in Tables 2.1 and 2.2 on p. 16 and 17, and in sec. 5.4.3 on p. 148.

1.2.1 Notational conventions

All vectors and matrices are emphasized using a bold font. The derivatives of transforms w.r.t. transform variable s are denoted by prime notation. That is, for example, the first four derivatives of $L_{\boldsymbol{\theta}}(s)$ w.r.t. s are denoted as $L'_{\boldsymbol{\theta}}(s)$, $L''_{\boldsymbol{\theta}}(s)$, $L'''_{\boldsymbol{\theta}}(s)$ and $L^{(4)}_{\boldsymbol{\theta}}(s)$. In contrast, the derivatives of the same functions w.r.t. distribution parameter are denoted by partial derivative operators, such as $\frac{\partial L_{\boldsymbol{\theta}}(s)}{\partial \theta}$, $\frac{\partial^2 L_{\boldsymbol{\theta}}(s)}{\partial \theta^2}$, $\frac{\partial^2 L_{\boldsymbol{\theta}}(s)}{\partial \theta_i \partial \theta_j}$, and so on. An exception is made in the sec. 6.1.5 of the last chapter where partial derivatives of multivariate characteristic functions are discussed: they are denoted by, for instance, $\frac{\partial \phi_{\boldsymbol{\theta}}(\boldsymbol{s})}{\partial s_1}$, where $\boldsymbol{s} = (s_1, ..., s_d)^\top$ is a vector of transform variables.

To simplify notation, we omit the dependence of distributions, densities, and transforms on $\boldsymbol{\theta}$ unless necessary. This convention is followed in Chapter 2 and parts of Chapter 3.

Given the multitude of functions, constants, parameters, and indices employed, it was inevitable that certain symbols are occasionally reused at various places (for example, p, q, α, γ, ψ, k, K, c, C, $w(\cdot)$). However, their meanings remain clear from the context.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

CHAPTER 2

Integral transforms and their empirical counterparts

2.1 Integral transforms

2.1.1 Laplace transforms and their properties

The Laplace transform (LT) of a probability distribution $F(x)$ is defined as

$$L(s) = \mathbb{E}[e^{-sX}] = \int_{-\infty}^{\infty} e^{-sx} \mathrm{d}F(x)\,. \tag{2.1}$$

Its analytical properties, which we describe below, depend on the support of $F(x)$ and heaviness of the tails.

2.1.1.1 Existence, uniqueness

For any distribution the LT exists (converges) on interval $\mathcal{D}$ of the form (σ_-, σ_+), $[\sigma_-, \sigma_+]$, $(\sigma_-, \sigma_+]$ or $[\sigma_-, \sigma_+)$, where $-\infty \leq \sigma_- \leq \sigma_+ \leq \infty$. In any case, $\mathcal{D}$ contains $s = 0$ and $L(0) = 1$. The values of σ_- and σ_+ are determined by the limits of $e^{sx}F(x)$ as $x \to \infty$ and $e^{sx}(1-F(x))$ as $x \to -\infty$, respectively. For distributions with exponential tails we have $-\infty < \sigma_- < 0$ and $0 < \sigma_+ < \infty$, while for those with sub-exponential tails $\sigma_- = -\infty$ and $\sigma_+ = \infty$. If either the positive or negative tail is heavy, then $\sigma_- = 0$ or $\sigma_+ = 0$. Finally, if $\sigma_- = \sigma_+ = 0$ we say that LT does not exist (e.g. Cauchy distribution).

Distinct probability distributions have unique LTs, and any distribution is uniquely determined by values of its LT on some interval $I \in \mathcal{D}$; see Feller (1971).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.1.1.2 Convexity, boundedness, monotonicity

As expectation of a convex and positive function, $L(s)$ is always convex and positive. Regarding boundedness and monotonicity, the two following cases can be observed:

- $F(x)$ is supported both on the positive and the negative real axis. Then $L(s)$ is unbounded for all $s \in \mathcal{D}$ except for $s = 0$ and $\lim_{s\to\sigma_-} L(s) = \lim_{s\to\sigma_+} L(s) = \infty$. Moreover, if $F(x)$ is symmetric about $x = 0$, then $L(s)$ is even and has minimum at $s = 0$.

- $F(x)$ is supported only on the positive (resp. negative) real axis. In this case, $0 \leq L(s) \leq 1$ for $s \geq 0$ (resp. $s \leq 0$) and $L(s)$ is monotone decreasing (resp. increasing) and $\lim_{s\to\infty} L(s) = 0$ (resp. $\lim_{s\to-\infty} L(s) = 0$). Moreover, $L(s)$ is completely monotone on $(0, \infty)$ (resp. on $(-\infty, 0)$), which means that derivatives $L^{(k)}(s)$ of all orders $k = 1, 2, ...$ exist on this interval and $(-1)^k L^{(k)}(s) \geq 0$; see Feller (1971) or Abate and Whitt (1996).

2.1.1.3 Affine transformations

Let X be a random variable with LT $L_{\mathrm{X}}(s)$. For any two $a, b \in \mathbb{R}$ the LT of $aX + b$ is given by

$$L_{a\mathrm{X}+b}(s) = e^{-sb} L_{\mathrm{X}}(as)\,. \tag{2.2}$$

This holds since $L_{a\mathrm{X}+b}(s) = \mathbb{E}[e^{-s(a\mathrm{X}+b)}] = \mathbb{E}[e^{-saX} e^{-sb}] = \mathbb{E}[e^{-saX}]\mathbb{E}[e^{-sb}]$.

2.1.1.4 Moment generation

If X has moments of all orders, that is $\mathbb{E}[X^k] < \infty$, $k \geq 1$, then $L(s)$ admits the Taylor expansion about $s = 0$,

$$L(s) = \sum_{k=0}^{\infty} \frac{s^k}{k!}\,\mathbb{E}[X^k]\,, \qquad |s| < \min(|\sigma_-|, \sigma_+), \tag{2.3}$$

with $\mathbb{E}[X^0] = 1$. If σ_- or $\sigma_+ = 0$, the expansion is not possible, although moments of all orders may exist. Moreover,

$$\mathbb{E}[X^k] = \lim_{s\to 0} (-1)^k L^{(k)}(s)\,, \tag{2.4}$$

meaning that finite k-th derivative of LT at $s = 0$ implies finiteness of $\mathbb{E}[X^k]$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

2.1.1.5 Tauberian relations

The so-called *Tauberian theorems* stand for various theorems describing the asymptotic behavior of a distribution in terms of the asymptotic behavior of its transforms; see e.g. Widder (1959) and Feller (1971). For most statistical applications, qualitative understanding of these relations is sufficient. In particular, when using LTs, it is important to keep in mind that

(i) the tail behavior of $F(x)$ as $|x| \to \infty$ reflected by the behavior of $L(s)$ as $|s| \to 0$;

(ii) the behavior of $L(s)$ for large $|s|$ describes that of $F(x)$ at small values of $|x|$.

This is explained as follows: when $|s|$ is small it is only when $|x|$ is large that variations in s significantly change the values of e^{-sx} and thus of $L(s) = \int e^{-sx}\mathrm{d}F(x)$. On the other hand, for large $|s|$, $L(s)$ is greatly influenced by changes of $F(x)$ close to $x = 0$. Moreover, (2.4) is in agreement with (i) since the values of the moments (or their non-existence) also reflect the tail behavior of $F(x)$.

2.1.2 Examples of Laplace transforms

The properties discussed above are now demonstrated through examples. Tables 2.1 and 2.2 on p. 16 and 17 list the most popular continuous distributions and their LTs. The tables include both two-tailed (symmetric) and one-tailed (skewed) distributions, with tails of different heaviness and LTs converging on different intervals. Included is also the Cauchy distribution whose LT does not exist except for $s = 0$. We see that in some cases the LTs admit closed form expressions, while in others, they do not. Elegant expressions in terms of elementary functions are, in fact, available only for the normal, Laplace, logistic, uniform, gamma and inverse Gaussian laws. In yet other cases, the transforms may be expressed with the help of a special function (as seen for the Gumbel and Pareto distributions), or, if all the moments exist, represented by the Taylor expansion (as in case of the beta distribution of 1st kind). Nevertheless, for the remaining distributions (and many more not listed in the table), tractable expressions are unknown. This fact is an important motivation behind the differential-equation methodology proposed in this thesis.

Fig. 2.1 plots the LTs for eight selected distributions, each under three parameter settings. The normal and logistic distributions demonstrate the cases with unbounded LTs. In contrast, for the gamma, inverse Gaussian, Pareto, log-normal, and (one-sided) uniform distributions, we observe that $0 < L(s) \leq 1$ for all $s \in [0, \infty]$, no matter if $\mathcal{D}$ coincides with that interval (the heavy-tailed cases) or if $\mathcal{D}$ extends on some values negative s-values (the exponentially-tailed cases). Additionally, note that, as indicated by (2.4), the steepness of $L(s)$ near $s = 0$ is higher for parameter settings that yield larger moment values.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Table 2.1: Standard probability distributions and their Laplace transforms.

distribution	density $f(x)$	Laplace transform $L(s)$	i.o.c $\mathcal{D}$
normal(μ,σ^2) $x,\mu\in\mathbb{R},\ \sigma^2>0$	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$	$\exp\{-\mu s+\sigma^2 s^2/2\}$	$\mathbb{R}$
Laplace(μ,β) $x,\mu\in\mathbb{R},\ \beta>0$	$\frac{1}{2\beta}\exp\left\{\frac{-\lvert x-\mu\rvert}{\beta}\right\}$	$\frac{\exp\{-\mu s\}}{1-\beta^2 s^2}$	$\left(-\frac{1}{\beta},\frac{1}{\beta}\right)$
logistic(μ,β) $x,\mu\in\mathbb{R},\ \beta>0$	$\frac{\exp\left\{\frac{-(x-\mu)}{\beta}\right\}}{\beta\left(1+\exp\left\{\frac{-(x-\mu)}{\beta}\right\}\right)^2}$	$\frac{\exp\{-\mu s\}\pi\beta s}{\sinh(\pi\beta s)}$	$\left(-\frac{1}{\beta},\frac{1}{\beta}\right)$
exponential(λ) $x\geq 0,\lambda>0$	$\lambda\exp\{-\lambda x\}$	$\frac{\lambda}{\lambda+s}$	$(-\lambda,\infty)$
gamma(α,β) $x\geq 0\ \ \alpha,\beta>0$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp\{-\beta x\}$	$\left(\frac{\beta}{\beta+s}\right)^\alpha$	$(-\beta,\infty)$
inv.Gauss.(λ,μ) $x,\lambda,\mu>0$	$\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left\{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right\}$	$\exp\left\{\frac{\lambda}{\mu}\left(1-\sqrt{1+\frac{2\mu^2 s}{\lambda}}\right)\right\}$	$\left[-\frac{\lambda}{2\mu^2},\infty\right)$
uniform(a,b) $x\in(a,b),\ a,b\in\mathbb{R},\ b>a$	$\frac{1}{b-a}$	$\frac{\exp\{-as\}-\exp\{-bs\}}{s(b-a)}$	$\mathbb{R}$
triangular(a,b,c) $x\in(a,c),\ a,b,c\in\mathbb{R},$ $c>b>a$	$\begin{cases}\frac{2(x-a)}{(b-a)(c-a)} & a<x<b\\ 2/(c-a) & x=b\\ \frac{2(c-x)}{(c-a)(c-b)} & b<x<c.\end{cases}$	$\frac{\left[(c-b)e^{-as}-(c-a)e^{-bs}+(b-a)e^{-cs}\right]}{(c-a)(b-a)(c-b)s^2/2}$	$\mathbb{R}$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Table 2.2: Standard probability distributions and their Laplace transforms (cont.)

distribution	density $f(x)$	Laplace transform $L(s)$	i.o.c $\mathcal{D}$
Gumbel(μ,β) $x,\mu\in\mathbb{R},\beta>0$	$\exp\Big\{-\Big(\frac{x-\mu}{\beta}+e^{-(x-\mu)/\beta}\Big)\Big\}$	$\Gamma(1+\beta s)\exp\{-\mu s\}$	$\big(-\frac{1}{\beta},\infty\big)$
Pareto$(\alpha,\beta)^*$ $x\geq 0\ \ \alpha,\beta>0$	$\frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}$	$\alpha(\beta s)^\alpha\exp\{s\beta\}\Gamma(\beta s,-\alpha)$	$[0,\infty)$
Rayleigh$(\vartheta)^{**}$ $x,\vartheta>0$	$\frac{x}{\vartheta^2}\exp\{-x^2/2\vartheta^2\}$	$1-s\vartheta\sqrt{\frac{\pi}{2}}\exp\Big(\frac{\vartheta^2s^2}{2}\Big)\,\mathrm{erfc}\Big(\frac{\vartheta s\sqrt{2}}{2}\Big)$	$\mathbb{R}$
Beta(α,β) $0<x<1,\ \alpha,\beta>0$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$	$1+\sum_{k=1}^{\infty}\frac{(\text{-}s)^k}{k}\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}$	$\mathbb{R}$
Beta-2(α,β) $x\geq 0,\ \alpha,\beta>0$	$\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{\mathrm{B}(\alpha,\beta)}$	no tractable form	$[0,\infty)$
log-normal(μ,σ) $x,\sigma>0,\ \mu\in\mathbb{R}$	$\frac{1}{\sqrt{2\pi}\sigma x}\exp\Big\{-\frac{(\log x-\mu)^2}{2\sigma^2}\Big\}$	no tractable form	$[0,\infty)$
Weibull(λ,c) $x\geq 0,\ \lambda,c>0$	$\lambda c(\lambda x)^{c-1}\exp\{-(\lambda x)^c\}$	no tractable form	$[0,+\infty)$ for $c<1$
Maxwell(σ) $x,\sigma>0$	$\sqrt{\frac{2}{\pi}}\frac{x^2}{\sigma^3}\exp\{-x^2/2\sigma^2\}$	no tractable form	$\mathbb{R}$
GIG$(\lambda,\mu,p)^{***}$ $x,\lambda,\mu>0,\ p\in\mathbb{R}$	$\frac{\mu^p x^{p-1}}{2\mathrm{K}_p(\lambda/\mu)}\exp\Big\{\frac{-\lambda(x-\mu)^2}{2\mu^2x}\Big\}$	no tractable form	$\mathbb{R}$
Cauchy(μ,β) $x,\mu\in\mathbb{R},\beta>0$	$\frac{1}{\pi\beta}\Big(\frac{\beta^2}{\beta^2+(x-\mu)^2}\Big)$	does not exist	–

$\mathrm{B}(\cdot,\cdot)$ - beta function; $\Gamma(\cdot)$ - gamma function; $\Gamma(\cdot,\cdot)$ - lower incomplete gamma functions;
$\mathrm{erfc}(\cdot)$ - complementary error function; $\mathrm{K}_p(\cdot)$ - modified Bessel function of the 2nd kind.
* Pareto type II (Lomax); this LT expression is derived in Nadarajah and Kotz (2006).
** This LT expression is given in Meintanis and Iliopoulos (2003b).
*** GIG = generalized inverse Gaussian distribution; see sec. 3.2.1.2 for details.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

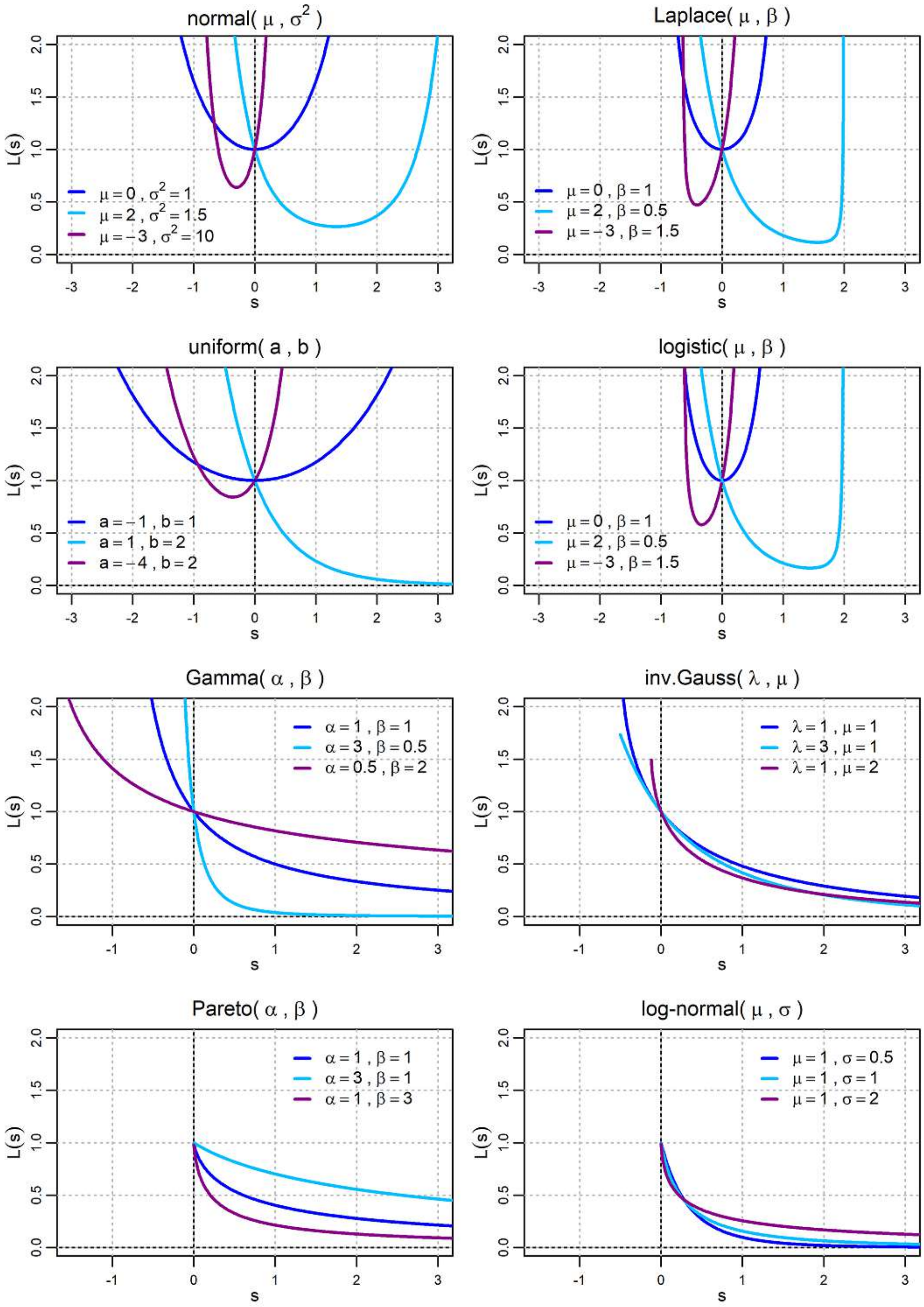


Figure 2.1: Laplace transforms of selected probability distributions.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

2.1.3 Characteristic functions and their properties

The problems related to unboundedness, various convergence regions or non-existence of LTs are resolved by employing the *characteristic function* (CF). The CF is defined as

$$\phi(s) = \mathbb{E}[e^{isX}] = \int_{-\infty}^{\infty} e^{isx}\mathrm{d}F(x)\,, \qquad i^2 = -1\,, \tag{2.5}$$

and exists for any distribution for all $s \in \mathbb{R}$. However, this convenience comes at the cost of working with a complex-valued quantity. Namely, CF must be decomposed into the real and imaginary part, that is $\phi(s) = \mathrm{Re}\{\phi(s)\} + i\mathrm{Im}\{\phi(s)\} = u(s) + iv(s)$, where

$$u(s) = \mathbb{E}[\cos(sX)] = \int_{-\infty}^{\infty} \cos(sx)\mathrm{d}F(x)\,, \tag{2.6}$$

and

$$v(s) = \mathbb{E}[\sin(sX)] = \int_{-\infty}^{\infty} \sin(sx)\mathrm{d}F(x)\,. \tag{2.7}$$

Alternatively, CF can be represented in the polar form as $\phi(s) = |\phi(s)|e^{i\arg\phi(s)}$, where $|\phi(s)| = \sqrt{u^2(s)+v^2(s)}$ is the *complex modulus* and $\arg\phi(s) = \arctan(v(s)/u(s))$ stands for the *polar argument*. Often, it is easier to work with the *squared modulus*, namely

$$|\phi(s)|^2 = u^2(s) + v^2(s)\,. \tag{2.8}$$

In this thesis, we refer to $u(s)$, $v(s)$ and $|\phi(s)|^2$ as the *CF components*.

2.1.3.1 Boundedness, monotonicity, uniqueness

CF of any distribution is a uniformly continuous function and has the following properties:

(i) $\phi(0) = 1$, $u(0) = 1$ and $v(0) = 0$;

(ii) $|\phi(s)| \leq 1$ for all $s \in \mathbb{R}$;

(iii) $\phi(-s) = \overline{\phi(s)}$.

The properties (i) and (iii) follow directly from the definition (2.5), whereas (ii) results from boundedness of the sine and cosine functions. Moreover, if $F(x)$ is symmetric about $x=0$ then

(iv) $v(s) = 0$,

so that $|\phi(s)|^2 = u^2(s)$. This property may be applied to test symmetry of a distribution.

A certain drawback of CF (when compared to LT) is that $|\phi(s)|$, $u(s)$ and $v(s)$ are non-monotone due to periodic factors. Specifically, for any CF $\phi(s)$ there exists $s_0 > 0$ such that $\phi(s_0) = 1$ and s_0 is period of $\phi(s)$; see Epps (1993), Ushakov (1999).

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Regarding the behavior of $|\phi(s)|$ away from zero, there are two cases: if $F(x)$ is absolutely continuous then $\lim_{|s|\to\infty}|\phi(s)| = 0$, whereas if $F(x)$ is discrete $\limsup_{|s|\to\infty}|\phi(s)| = 1$.

The values of $\phi(s)$ on the entire real s-line uniquely determine the underlying distribution. However, unlike for LT, CFs of two different distributions may coincide over finite interval $(-a, a)$, $a \in \mathbb{R}$. Still, such cases are exceedingly rare and remain of purely academic interest. For examples, see Ushakov (1999) p. 263-264, and Feller (1971) p. 506.

2.1.3.2 Affine transformations

Let X be a random variable with CF $\phi_{\mathrm{X}}(s)$. Then, for any two $a, b \in \mathbb{R}$, the CF of $aX + b$ is given by

$$\phi_{a\mathrm{X}+b}(s) = e^{isb}\phi_{\mathrm{X}}(as)\,, \tag{2.9}$$

which follows from the expectation property.

Another important and useful fact is the invariance of the modulus to changes of location of X. Setting $a = 1$ in (2.9), it is not difficult to show that for any $b \in \mathbb{R}$

$$|\phi_{\mathrm{X}+b}(s)| = |\phi_{\mathrm{X}}(s)|\,. \tag{2.10}$$

2.1.3.3 Moment generation. Tauberian relations

Similarly to LT, we now highlight the connection between CF to moments and tails of distributions. If the k-th moment of $F(x)$ exists, then $\phi(s)$ is k times differentiable around zero, allowing for the Taylor expansion,

$$\phi(s) = \sum_{j=0}^{k} \frac{(is)^j}{j!}\mathbb{E}[X^j] + o(|s|^j)\,, \qquad \text{as } |s| \to 0\,, \tag{2.11}$$

where

$$\mathbb{E}[X^k] = (-i)^k\phi^{(k)}(0)\,. \tag{2.12}$$

The Tauberian relations for CFs and the underlying distributions are qualitatively identical to those for LT and have the same implications for statistical procedures. Namely, the tail behavior of $F(x)$ is determined by the behavior of $\phi(s)$ near the origin $s = 0$, and vice versa, the behavior of $\phi(s)$ for large s reflects that of $F(x)$ in the neighborhood of $x = 0$. The specific theorems on these relations can be found, for example, in Pitman (1968).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

2.1.4 Examples of characteristic functions

Finally, we exemplify CFs of some concrete parametric distributions. Symbolic computation of CF is usually as easy or difficult as the computation of the corresponding LT. Expressions for CF in particular cases follow from LT by the change of variable $s \to -is$, $i^2 = -1$. Lack of a tractable expression for LT means the same for CF. An important exception is the Cauchy(μ,β) model whose CF is known and expressed as $\phi(s) = \exp\{\mu is - \beta|s|\}$, although the LT does not exist.

Unfortunately, the use of CF poses analytical problems because in statistical applications one often needs the expressions for $|\phi(s)|$, $|\phi(s)|^2$, $u(s)$ and $v(s)$, and these are not always derivable. For example, the CFs of the gamma(α,β) distribution,

$$\phi(s) = \left(\frac{\beta}{\beta - is}\right)^{\alpha}, \tag{2.13}$$

cannot be decomposed into the real and imaginary parts due to non-integer power α. The same applies to the inverse Gaussian, Pareto and Gumbel distributions. This limitation diminishes usefulness of CF for several life-time distributions and favors LT.

Table 2.3 on p. 22 lists standard distributions for which all the CF components have tractable expressions. The table also includes the discrete Poisson(λ) distribution, with probability mass function, $p(k) = \lambda^k e^{-\lambda}/k!$, $k = 0, 1, 2...$, to exemplify a CF which is purely periodic.

Fig. 2.2 plots the functions $|\phi(s)|^2$, $u(s)$ and $v(s)$ in four representative cases: normal, Cauchy, exponential and Poisson. We have there examples of distributions symmetric around $x = 0$, i.e. normal(0,1) and Cauchy(0,1), in which case always $v(s) = 0$ for all $s \in \mathbb{R}$. The Cauchy examples, in general, demonstrate the scenario where CF is not differentiable at $s = 0$, indicating a lack of moments for that random variable. In turn, for the Poisson law, as just mentioned, all the CF components are periodic. Finally, the exponential distribution is an instance of skewed distributions for which $|\phi(s)|^2 = u(s)$. Refer to Meintanis and Iliopoulos (2003a) for more details on this interesting property of exponentiality and its statistical uses.

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Table 2.3: Characteristic functions of standard probability distributions; $s \in \mathbb{R}$

distribution	CF and squared modulus	real and imaginary part
normal(μ, σ^2) $x, \mu \in \mathbb{R},\ \sigma^2 > 0$	$\phi(s) = \exp\{i\mu s - \sigma^2 s^2/2\}$ $\lvert\phi(s)\rvert^2 = \exp\{-\sigma^2 s^2\}$	$u(s) = \cos(\mu s)\exp\{-\sigma^2 s^2/2\}$ $v(s) = \sin(\mu s)\exp\{-\sigma^2 s^2/2\}$
Laplace(μ, β) $x, \mu \in \mathbb{R},\ \beta > 0$	$\phi(s) = \frac{e^{i\mu s}}{1+\beta^2 s^2}$ $\lvert\phi(s)\rvert^2 = \frac{1}{(1+\beta^2 s^2)^2}$	$u(s) = \frac{\cos(s\mu)}{1+\beta^2 s^2}$ $v(s) = \frac{\sin(s\mu)}{1+\beta^2 s^2}$
logistic(μ, β) $x, \mu \in \mathbb{R},\ \beta > 0$	$\phi(s) = \frac{\exp\{is\mu\}\pi\beta s}{\sinh(\pi\beta s)}$ $\lvert\phi(s)\rvert^2 = \frac{(\pi\beta s)^2}{\sinh^2(\pi\beta s)}$	$u(s) = \frac{\cos(s\mu)\pi\beta s}{\sinh(\pi\beta s)}$ $v(s) = \frac{\sin(s\mu)\pi\beta s}{\sinh(\pi\beta s)}$
Cauchy(μ, β) $x, \mu \in \mathbb{R},\ \beta > 0$	$\phi(s) = \exp\{\mu i s - \beta\lvert s\rvert\}$ $\lvert\phi(s)\rvert^2 = \exp\{-\beta\lvert s\rvert\}$	$u(s) = \cos(\mu s)e^{-\beta\lvert s\rvert}$ $v(s) = \sin(\mu s)e^{-2\beta\lvert s\rvert}$
stable(μ, γ, α) $x, \mu \in \mathbb{R},\ \gamma > 0,\ \alpha \in (0{,}2]$	$\phi(s) = \exp\{\mu i s - \gamma\lvert s\rvert^\alpha\}$ $\lvert\phi(s)\rvert^2 = \exp\{-2\gamma\lvert s\rvert^\alpha\}$	$u(s) = \cos(\mu s)e^{-\gamma\lvert s\rvert^\alpha}$ $v(s) = \sin(\mu s)e^{-\gamma\lvert s\rvert^\alpha}$
uniform(a, b) $x \in (a{,}b),\ a{,}b \in \mathbb{R},\ b > a$	$\phi(s) = \frac{\exp\{ibs\} - \exp\{ias\}}{is(b-a)}$ $\lvert\phi(s)\rvert^2 = \frac{4\sin^2(s(b-a)/2)}{s^2(b-a)^2}$	$u(s) = \frac{\sin(sb) - \sin(sa)}{s(b-a)}$ $v(s) = \frac{\cos(sa) - \cos(sb)}{s(b-a)}$
exponential(λ) $x \geq 0,\ \lambda > 0$	$\phi(s) = \frac{\lambda}{\lambda - is}$ $\lvert\phi(s)\rvert^2 = \frac{\lambda^2}{\lambda^2 + s^2}$	$u(s) = \frac{\lambda^2}{\lambda^2 + s^2}$ $v(s) = \frac{\lambda s}{\lambda^2 + s^2}$
Poisson(λ) $k = 0{,}1{,}2{,}\ldots,\ \lambda > 0$	$\phi(s) = \exp\{-\lambda(1 - e^{is})\}$ $\lvert\phi(s)\rvert^2 = \exp\{-2\lambda(1 - \cos(s))\}$	$u(s) = e^{-\lambda(1-\cos(s))}\cos\{\lambda\sin(s)\}$ $v(s) = e^{-\lambda(1-\cos(s))}\sin\{\lambda\sin(s)\}$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

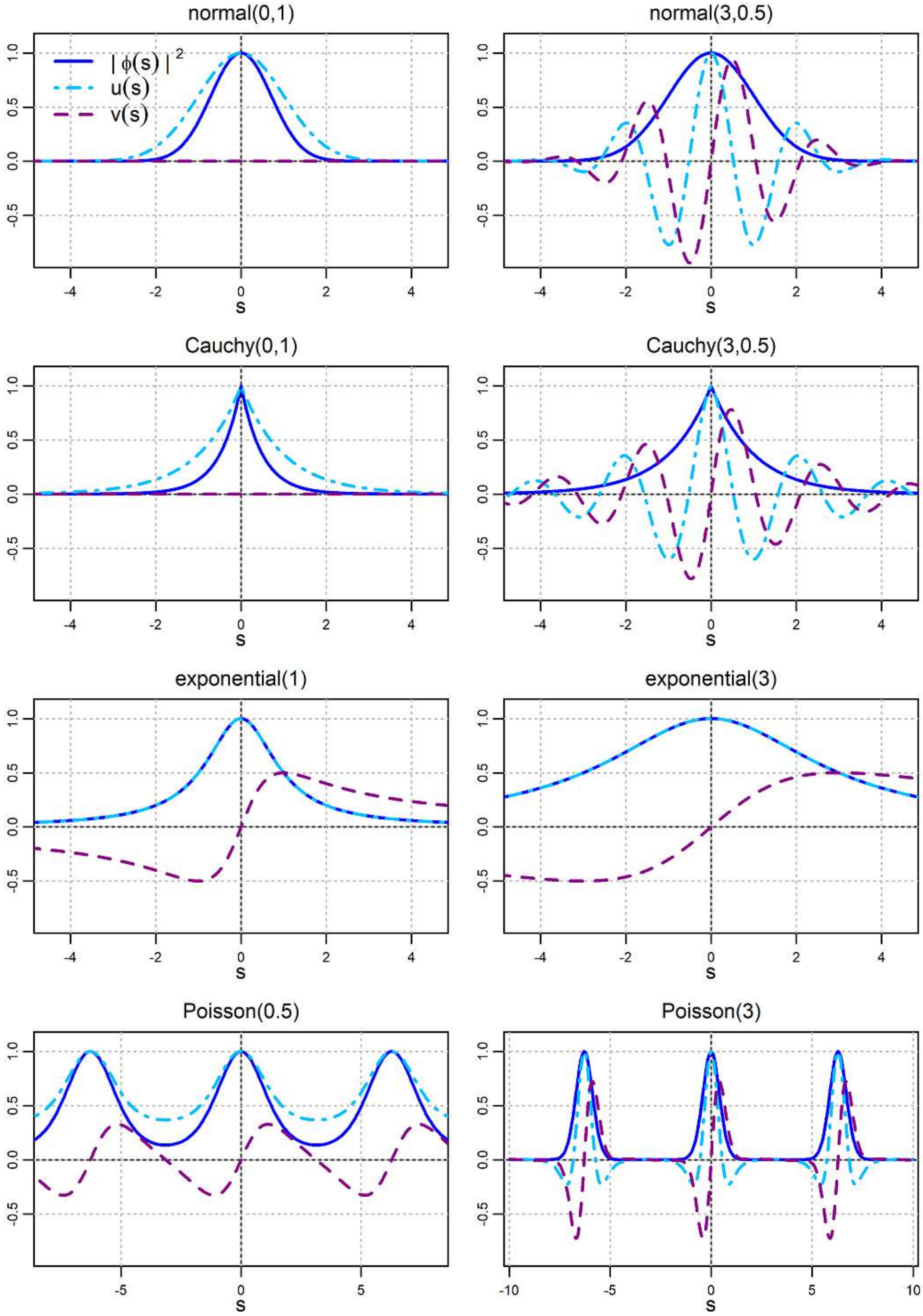


Figure 2.2: Components of CF, $|\phi(s)|^2$, $u(s)$ and $v(s)$, for selected standard distributions.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

2.2 Empirical integral transforms

The *empirical integral transform* is defined as

$$\widehat{\mathcal{T}}_n(s) = \int K(s,x)\mathrm{d}\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} K(s,X_j)\,, \tag{2.14}$$

where $\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{I}\{X_j \leq x\}$ is empirical distribution function of the sample $X_1, ..., X_n$. Under i.i.d. assumption it is an unbiased non-parametric estimator of the model, that is

$$\mathbb{E}[\widehat{\mathcal{T}}_n(s)] = \mathcal{T}(s)\,, \tag{2.15}$$

and, by the Strong Law of Large Numbers (SLLN),

$$\widehat{\mathcal{T}}_n(s) \xrightarrow{\text{a.s.}} \mathcal{T}(s)\,. \tag{2.16}$$

The variance of the empirical transform for any n is

$$\mathbb{V}[\widehat{\mathcal{T}}_n(s)] = \frac{\mathbb{E}[K(s,X)^2] - \mathbb{E}^2[K(s,X)]}{n}\,. \tag{2.17}$$

If $\mathbb{E}[K(s,X)^2]$ is finite, the Central Limit Theorem (CLT) implies asymptotic normality of the functional as $n \to \infty$, namely

$$\sqrt{n}\Big(\widehat{\mathcal{T}}_n(s) - \mathcal{T}(s)\Big) \xrightarrow{\text{d}} \mathcal{N}\Big(0\,, \mathbb{E}[K(s,X)^2] - \mathbb{E}^2[K(s,X)]\Big). \tag{2.18}$$

Then, considered as a random variable indexed by s, $\widehat{\mathcal{T}}_n(s)$ is a non-stationary continuous-time random process converging to a Gaussian process with certain covariance function.

2.2.1 Empirical Laplace transform

The empirical counterpart of (2.1), called the *empirical Laplace transform* (ELT), is

$$\widehat{L}_n(s) = \frac{1}{n}\sum_{j=1}^{n} e^{-sX_j}\,. \tag{2.19}$$

While $L(s)$ exists often only on some interval $\mathcal{D} \subseteq \mathbb{R}$, $\widehat{L}_n(s)$ exists and is differentiable always for all $s \in \mathbb{R}$. Also, in any case, $\widehat{L}_n(0) = L(0) = 1$, and, according to (2.4), $(-1)^k \widehat{L}_n^{(k)}(0) = \frac{1}{n}\sum\limits_{j=1}^{n} X_j^k$, $k \geq 0$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

2.2.1.1 Finite-sample and asymptotic behavior

We have obviously $\mathbb{E}[\widehat{L}_n(s)] = L(s)$, and, using (2.17), we obtain

$$\mathbb{V}[\widehat{L}_n(s)] = \frac{L(2s) - L^2(s)}{n}\,. \tag{2.20}$$

Note that the variance is finite for all s such that $2s \in \mathcal{D}$. That is, it is finite for all $s \geq 0$ for the distributions on $\mathbb{R}_+$ or $\mathbb{R}_+^0$ (e.g. gamma, Pareto). However, for the distributions supported on $\mathbb{R}_+$, whose LT exists on a bounded interval (σ_-, σ_+) (e.g. Laplace, logistic), the variance is finite merely for $s \in (\sigma_-/2, \sigma_+/2)$. For these s-points, by CLT,

$$\sqrt{n}\Big(\widehat{L}_n(s) - L(s)\Big) \ \xrightarrow{\ \mathrm{d}\ }\ \mathcal{N}\Big(0\,, L(2s) - L^2(s)\Big)\,. \tag{2.21}$$

For any two points s_1 and s_2 such that $s_1 + s_2 \in \mathcal{D}$, $\widehat{L}_n(s)$ converges to a Gaussian process with mean $L(s)$ and covariance

$$\mathbb{C}\mathrm{ov}[\widehat{L}_n(s_1), \widehat{L}_n(s_2)] = \frac{L(s_1 + s_2) - L(s_1)L(s_2)}{n}\,. \tag{2.22}$$

Derivation of (2.22) is shown in Appendix B.

Finally, by SLLN we have that $\widehat{L}_n(s) \xrightarrow{\mathrm{a.s.}} L(s)$ point-wise. However, as proved by Csörgő (1982), the strong convergence holds also uniformly, that is $\sup_{s\in\mathcal{D}^I} |\widehat{L}_n(s) - L(s)| \xrightarrow{\mathrm{a.s.}} 0\,$, where $\mathcal{D}^I$ is the interior of $\mathcal{D}$.

2.2.1.2 Finite-sample behavior: examples

The accuracy of $\widehat{L}_n(s)$ as an estimator of $L(s)$ is now illustrated in two cases: (i) for a one-tailed skew distribution on $\mathbb{R}_+$, and (ii) for a two-tailed symmetric distribution on $\mathbb{R}$, taking the exponential(λ) and normal(μ,σ^2) distributions as examples. For these two distributions, from (2.20) we obtain

$$\mathbb{V}[\widehat{L}_n(s)] = \frac{\lambda s^2}{n(\lambda + 2s)(\lambda + s)^2}\,, \tag{2.23}$$

and

$$\mathbb{V}[\widehat{L}_n(s)] = \exp\{-2\mu s + \sigma^2 s^2\}(\exp\{\sigma^2 s^2\} - 1)/n\,, \tag{2.24}$$

respectively. Fig. 2.3 displays the two variances and simulated 95% confidence bands of $\widehat{L}_n(s)$, as an estimator of $L(s)$, for $n = 50$. We see that in both cases the precision of the estimation is highest around $s = 0$. At any other s-point, boundedness of $\mathbb{V}[\widehat{L}_n(s)]$ is contingent on the boundedness of $L(s)$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

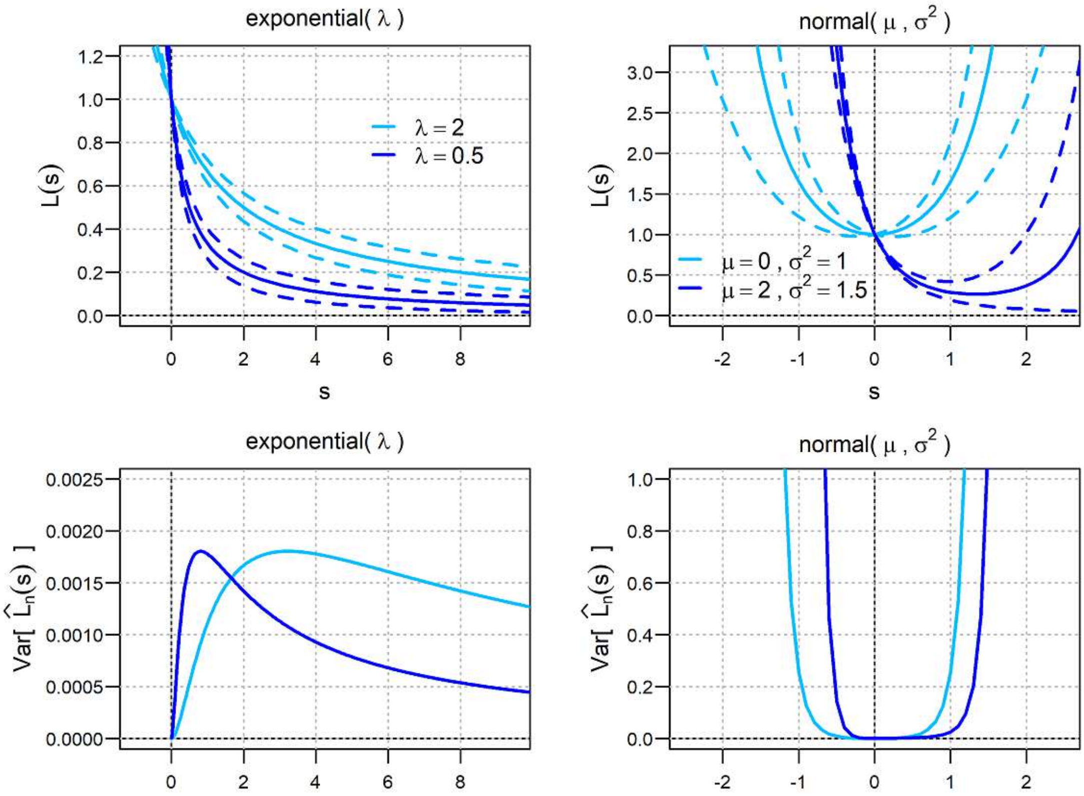


Figure 2.3: Variances and simulated 95% confidence bands of $\widehat{L}_n(s)$ as estimator of $L(s)$ for the normal and exponential distributions; sample size $n = 50$, 3000 replications.

2.2.2 Empirical characteristic function

The sample counterpart of (2.5), called the *empirical characteristic function* (ECF), is given by

$$\widehat{\phi}_n(s) \quad = \quad \frac{1}{n}\sum_{j=1}^{n} e^{isX_j}\,. \tag{2.25}$$

Its real and imaginary parts are given, respectively, by

$$\widehat{u}_n(s) = \frac{1}{n}\sum_{j=1}^{n}\cos(sX_j)\,, \qquad \text{and} \qquad \widehat{v}_n(s) = \frac{1}{n}\sum_{j=1}^{n}\sin(sX_j)\,. \tag{2.26}$$

The squared modulus of CF is by definition

$$|\widehat{\phi}_n(s)|^2 = \widehat{u}_n^2(s) + \widehat{v}_n^2(s)\,. \tag{2.27}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Applying trigonometric identities, it can also be expressed as

$$|\widehat{\phi}_n(s)|^2 = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\cos(s(X_j - X_k))\,, \tag{2.28}$$

which is a degree-2 V-statistic with symmetric kernel; see Lee (1990) and Appendix A.

Since $F_n(x)$ is a legitimate probability distribution, it follows from sec. 2.1.3.1 that: (i) $\widehat{\phi}_n(0) = 1$; (ii) $|\widehat{\phi}_n(s)| \leq 1,\ s\in\mathbb{R}$; (iii) $\widehat{\phi}_n(-s) = \overline{\widehat{\phi}_n(s)}$; and (iv) $\limsup_{|s|\to\infty}|\widehat{\phi}_n(s)| = 1$.

2.2.2.1 Finite-sample and asymptotic behavior

We will now examine some stochastic properties of the ECF components. Of course, $\mathbb{E}[\widehat{u}_n(s)] = u(s)$ and $\mathbb{E}[\widehat{v}_n(s)] = v(s)$. The finite-sample variances

$$\mathbb{V}[\widehat{u}_n(s)] = \frac{1/2 + u(2s)/2 - u^2(s)}{n}\,, \tag{2.29}$$

$$\mathbb{V}[\widehat{v}_n(s)] = \frac{1/2 - u(2s)/2 - v^2(s)}{n}\,, \tag{2.30}$$

are obtained from (2.17) by applying trigonometric power-reduction formulas. By CLT, we have then

$$\sqrt{n}\Big(\widehat{u}_n(s) - u(s)\Big) \xrightarrow{\ \mathrm{d}\ } \mathcal{N}\Big(0\,, 1/2 + u(2s)/2 - u^2(s)\Big)\,, \tag{2.31}$$

and

$$\sqrt{n}\Big(\widehat{v}_n(s) - v(s)\Big) \xrightarrow{\ \mathrm{d}\ } \mathcal{N}\Big(0\,, 1/2 - u(2s)/2 - v^2(s)\Big)\,, \tag{2.32}$$

both of which hold for all $s \in \mathbb{R}$ due to boundedness of $u(s)$ and $v(s)$.

The squared modulus of ECF, as a degree-2 V-statistics, is only asymptotically unbiased:

$$\mathbb{E}[|\widehat{\phi}_n(s)|^2] \quad = \quad |\phi(s)|^2 + \frac{1 - |\phi(s)|^2}{n}\,, \tag{2.33}$$

which is derived in Appendix B. Also, its variance has more complicated expression, namely

$$\begin{aligned}\mathbb{V}[|\widehat{\phi}_n(s)|^2] \quad &= \quad \frac{2(n-1)(n-2)}{n^3}\Big(\mathrm{Re}\{\phi^2(s)\overline{\phi(2s)}\} + |\phi(s)|^2 - 2|\phi(s)|^4\Big) \\ &+ \quad \frac{n-1}{n^3}\Big(|\phi(2s)|^2 - 2|\phi(s)|^4 + 1\Big).\end{aligned} \tag{2.34}$$

This expression was initially derived in Thaler (1974); see also Koutrouvelis (1980).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

The asymptotic normality of the squared modulus does not appear to be reported in the literature. By using the theory of U-statistic, see Appendix A, we can show that

$$\sqrt{n}\Big(|\widehat{\phi}_n(s)|^2 - |\phi(s)|^2\Big) \xrightarrow{\text{d}} \mathcal{N}\Big(0\,, \mathbb{V}_{\text{AS}}\Big), \tag{2.35}$$

where

$$\mathbb{V}_{\text{AS}} = 2u(2s)[u^2(s) - v^2(s)] + 4v(2s)u(s)v(s) + 2|\phi(s)|^2 - 4|\phi(s)|^4\,. \tag{2.36}$$

The statistics $\widehat{u}_n(s)$, $\widehat{v}_n(s)$ and $|\widehat{\phi}_n(s)|^2$ are asymptotically Gaussian processes in s. Their covariance functions for any finite n are given in Koutrouvelis (1980) and Ushakov (1999). The cross-covariance between $\widehat{u}_n(s)$ and $\widehat{v}_n(s)$, which will be needed later, is

$$\text{Cov}[\widehat{u}_n(s_1), \widehat{v}_n(s_2)] \;=\; \frac{1}{2n}\Big(v(s_1+s_2) - v(s_1-s_2) - 2u(s_1)v(s_2)\Big)\,. \tag{2.37}$$

Let us now look at the limiting behavior of the ECF components as $|s| \to 0$ and $|s| \to \infty$, for fixed n. By unbiasedness of $\widehat{u}_n(s)$ and $\widehat{v}_n(s)$ and by (2.29), (2.30), (2.33) and (2.34), it follows that for any distribution

$$\begin{cases} \lim\limits_{|s|\to 0} \Big(\mathbb{E}[\widehat{u}_n(s)], \mathbb{E}[\widehat{v}_n(s)]\Big) \;=\; (1,0)\,, \\ \lim\limits_{|s|\to 0} \Big(\mathbb{V}[\widehat{u}_n(s)], \mathbb{V}[\widehat{v}_n(s)]\Big) \;=\; (0,0)\,, \\ \lim\limits_{|s|\to 0} \mathbb{E}[|\widehat{\phi}_n(s)|^2] = 1, \qquad \lim\limits_{|s|\to 0} \mathbb{V}[|\widehat{\phi}_n(s)|^2] = 0\,. \end{cases} \tag{2.38}$$

As $|s| \to \infty$, the situation is more involved. For absolutely continuous distributions, based on the properties discussed in sec. 2.1.3.1 and given that $\lim_{|s|\to\infty} u(s) = 0$, $\lim_{|s|\to\infty} v(s) = 0$, one finds that

$$\begin{cases} \lim\limits_{|s|\to \infty} \Big(\mathbb{E}[\widehat{u}_n(s)], \mathbb{E}[\widehat{v}_n(s)]\Big) \;=\; (0,0)\,, \\ \lim\limits_{|s|\to \infty} \Big(\mathbb{V}[\widehat{u}_n(s)], \mathbb{V}[\widehat{v}_n(s)]\Big) \;=\; (1/2n, 1/2n)\,, \\ \lim\limits_{|s|\to \infty} \mathbb{E}[|\widehat{\phi}_n(s)|^2] = 1/n, \qquad \lim\limits_{|s|\to \infty} \mathbb{V}[|\widehat{\phi}_n(s)|^2] = (n-1)/n^3\,. \end{cases} \tag{2.39}$$

However, if the distribution is discrete (or has discrete masses), the finite-sample means and variances of $\widehat{u}_n(s)$, $\widehat{v}_n(s)$ and $|\widehat{\phi}_n(s)|^2$ oscillate in s and have no limits; see Koutrouvelis (1980).

Finally, we should mention the strong stochastic convergence of ECF as $n \to \infty$. According to SLLN, we know that $\widehat{\phi}_n(s) \xrightarrow{\text{a.s.}} \phi(s)$ point-wise. However, unlike for ELT, this property does not hold uniformly for all $s \in \mathbb{R}$. The strong convergence holds uniformly only over a *fixed* interval, i.e., $\sup_{|s|\le s_0} |\widehat{\phi}_n(s) - \phi(s)| \xrightarrow{\text{a.s.}} 0$ for any $s_0 < \infty$, as proved in Feuerverger and Mureika (1977). See also Marcus (1981) and Epps (1993).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

2.2.2.2 Finite-sample behavior: examples

We now illustrate the variances of $\widehat{u}_n(s)$, $\widehat{v}_n(s)$ and $|\widehat{\phi}_n(s)|^2$ taking as examples the exponential(λ) and normal(μ,σ^2) distributions. The variances are plotted in Fig. 2.4 for $n = 50$ along with simulated 95% confidence bands. To compute them, we used the formulas (2.29), (2.30) and (2.34) and inserted there relevant expressions for $u(s)$, $v(s)$ and $|\phi(s)|^2$ from Table 2.2. We observe that the components of ECF are most precise as estimators of the CF near $s = 0$. The variance curves of $|\widehat{\phi}_n(s)|^2$ for both types of distributions resemble those of $\widehat{L}_n(s)$ for one-tailed distributions; see Fig. 2.3. Namely, as $|s| \to \infty$, $|\widehat{\phi}_n(s)|^2$ attains a maximum and drops to a very low level: $(n-1)/n^3$. In contrast, the variances of $\widehat{u}_n(s)$ and $\widehat{v}_n(s)$ monotonically increase from zero to a higher level: $2/n$. This suggests that, at least for continuous distributions, $|\widehat{\phi}_n(s)|^2$ is more reliable inference tool than $\widehat{u}_n(s)$ or $\widehat{v}_n(s)$ when they are used alone.

2.3 Robustness of empirical transforms

Robustness in statistics refers to the resistance of a statistical functional against contamination of the (postulated) model distribution. Typically, interest lies in robustness against *outliers* which are excessively large or small observations in the sample. These untypical or erroneous observations can disrupt the outcomes of estimation or testing. In this thesis, we specifically analyze the robustness of estimators based on empirical transforms. Therefore, we first address robustness of the empirical transforms themselves as estimators of the model transforms. It must be stressed that robustness is a very multifaceted area, covered broadly, on a different level of mathematical abstraction, in monographs such as Hampel et al. (1986), Staudte and Sheather (1990), or Jurečková et al. (2019). For a special discussion on the nature and sources of outliers in statistical data, the reader is referred to Barnett and Lewis (1994). This thesis places emphasis on the *infinitesimal robustness* of the estimators, expressed by their influence function, which is by far the most popular robustness measure.

2.3.1 Influence function

Influence function (IF) describes the impact of infinitesimal contamination at point x exerted on a statistical functional $T(F)$, where $F = F(x)$ is the model distribution. Specifically, the IF is defined as

$$\mathrm{IF}[x; T(F)] = \lim_{\epsilon \to 0^+} \frac{T\Big((1-\epsilon)F + \epsilon\Delta_x\Big) - T(F)}{\epsilon}, \tag{2.40}$$

where Δ_x is the distribution function assigning probability 1 to single point x. The IF is thus a special case of the Gâteau derivative; see Jurečková et al. (2019).

Although one often says 'IF of an estimator' (that is of a function of data), the IF refers to the corresponding functional. In other words, when studying robustness of an estimator $\widehat{\theta}_n = T(\widehat{F}_n)$ we consider $\mathrm{IF}[x; \theta]$ where $\theta = T(F)$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

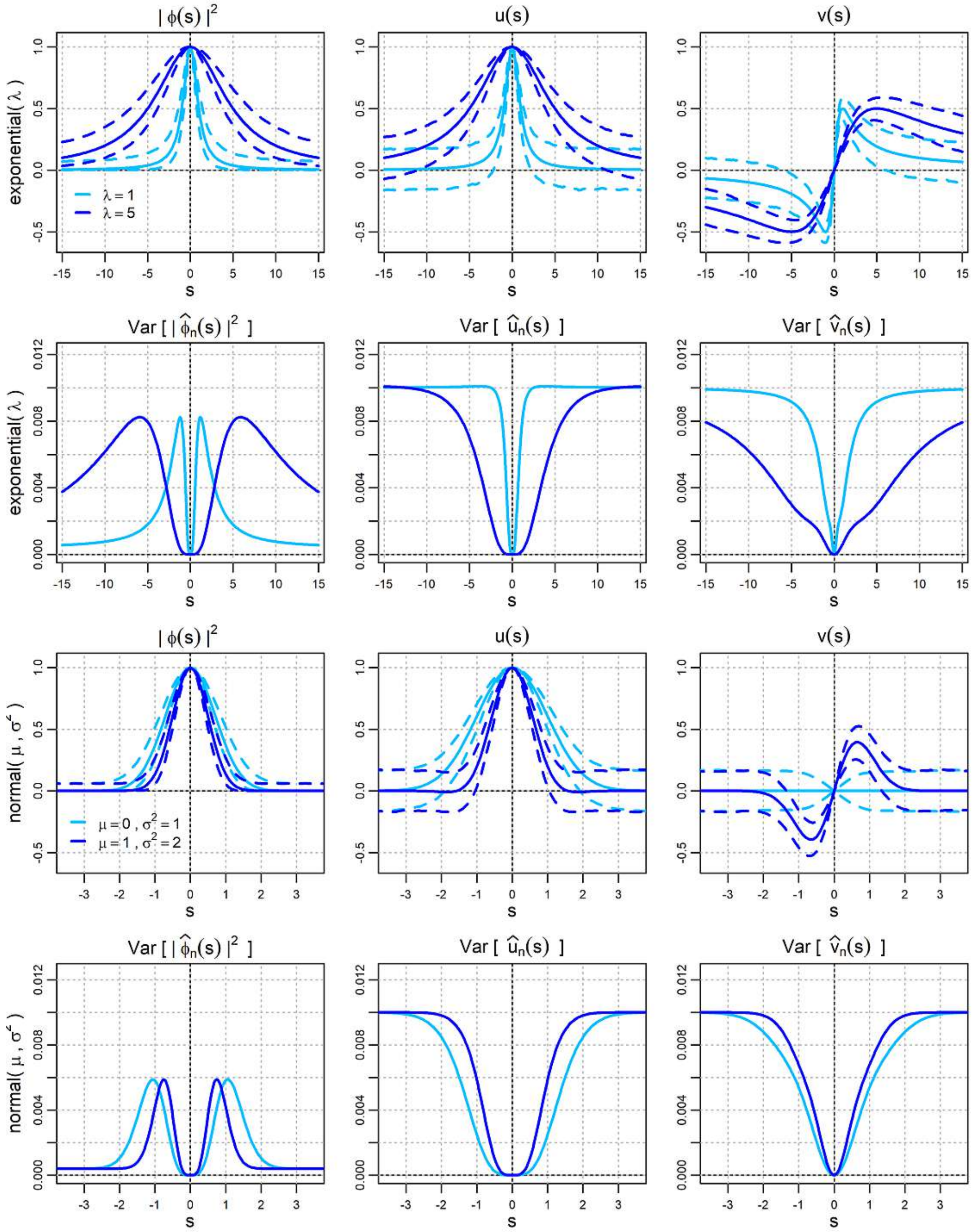


Figure 2.4: Variances and 95% confidence bands of $|\widehat{\phi}_n(s)|^2$, $\widehat{u}_n(s)$ and $\widehat{v}_n(s)$ (dashed) as estimators of $|\phi(s)|^2$, $u(s)$ and $v(s)$, for the normal and exponential distributions. Sample size $n = 50$, 3000 replications.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

For the mean and variance functionals (corresponding to sample mean and sample variance) one easily finds that $\mathrm{IF}[x;\mathbb{E}[X]] = x - \mathbb{E}[X]$ and $\mathrm{IF}[x;\mathbb{V}[X]] = (x-\mathbb{E}[X])^2 - \mathbb{V}[X]$. These IFs are unbounded in x, which means that the effect of the contamination at x may be arbitrarily large. Functionals (estimators) with unbounded IFs are termed as *unrobust*. For a robust functional one requires bounded IF, ideally decaying to zero as $x \to \pm\infty$.

IFs have two important properties. Firstly,

$$\mathbb{E}_F[\mathrm{IF}[X;T(F)]] = \int \mathrm{IF}[x;T(F)]\mathrm{d}F(x) = 0\,. \tag{2.41}$$

Secondly, if $T(F)$ is Fréchet differentiable (see Jurečková et al. (2019) for details) and if $\mathbb{V}_F[\mathrm{IF}(X,T(F))] < \infty$, we have

$$\sqrt{n}\Big(T(\widehat{F}_n) - T(F)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \mathbb{V}_F[\mathrm{IF}[X;T(F)]]\Big). \tag{2.42}$$

For a vector of p such functionals $\boldsymbol{T}(F) = [T_1(F),...,T_p(F)]^\top$ one defines the *joint influence function*, $\mathrm{IF}[x;\boldsymbol{T}(F)] = [\mathrm{IF}[x;T_1(F)],...,\mathrm{IF}[x;T_p(F)]]^\top$, for which we have

$$\sqrt{n}\Big(\boldsymbol{T}(\widehat{F}_n) - \boldsymbol{T}(F)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}_p\Big(\mathbf{0}\,, \boldsymbol{\Sigma}\Big), \tag{2.43}$$

where the asymptotic covariance matrix is given by

$$\boldsymbol{\Sigma} = \int \mathrm{IF}[x;\boldsymbol{T}(F)]\,\mathrm{IF}[x;\boldsymbol{T}(F)]^\top \mathrm{d}F(x). \tag{2.44}$$

Thus, in addition to diagnosing robustness, the IF also provides an alternative method for deriving the asymptotic variance or covariance matrix of statistics associated with the underlying functional.

2.3.1.1 Empirical influence

One may also define the empirical version of IF based on a realized sample $(x_1,...,x_n)$ with $\widehat{F}_n(x)$. For this purpose, consider a contaminated the sample $(x_1,...,x_n,x_\mathrm{c})$ where x_c a contaminating observation, and denote its distribution function by $\widehat{F}^\mathrm{c}_{n+1}(x)$. Then, the *empirical influence* (EI) associated with $T(F)$ is

$$\mathrm{EI}[x_\mathrm{c};T(F)] = T(\widehat{F}^\mathrm{c}_{n+1}) - T(\widehat{F}_n). \tag{2.45}$$

It can be shown that

$$\mathrm{IF}[x;T(F)] = \lim_{n\to\infty} (n+1)\,\mathrm{EI}[x;T(F)]\,. \tag{2.46}$$

With large sample size n, EI allows for Monte Carlo simulation of IF. This can be beneficial when the derivation of IF is difficult or requires validation; see Nasser and Alam (2006).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

2.3.1.2 Functions of functionals

In robustness analysis, we often seek the IF of a function applied to a functional with known IF. That is, we seek $\mathrm{IF}[x;Y]$, where $Y = g(T(F))$ and $\mathrm{IF}[x;T]$ is known. If $g(\cdot)$ is differentiable, the chain rule for derivatives applies, and

$$\mathrm{IF}[x;Y] = \frac{\mathrm{d}g(T(F))}{\mathrm{d}T(F)} \cdot \mathrm{IF}[x;T(F)]\,. \tag{2.47}$$

If Y is a differentiable function of p functionals, $Y = g(T_1(F), ..., T_p(F))$, described by a joint influence function $[\mathrm{IF}[x;T_1(F)], ..., \mathrm{IF}[x;T_p(F)]]^\top$, then the multivariate version of chain rule yields

$$\mathrm{IF}[x;Y] = \sum_{k=1}^{p} \frac{\partial g(T_1(F), ..., T_p(F))}{\partial T_k(F)} \cdot \mathrm{IF}[x;T_k(F)]\,, \tag{2.48}$$

see Campbell (1993). This formula will be extensively used to compute IFs of new estimators constructed in this thesis.

2.3.2 Influence functions for transforms

Robustness and influence functions of integral transforms and their functionals have been studied in several works, including Luong and Thompson (1987), Campbell (1993) and Meintanis and Donatos (1996). For an integral transform $\mathcal{T}(s) = \mathbb{E}_F[K(s,x)]$, it follows easily from (2.40) that

$$\mathrm{IF}[x;\mathcal{T}(s)] = K(s,x) - \mathcal{T}(s)\,. \tag{2.49}$$

Hence, boundedness of the transform (in s) and the kernel (in s and x) ensures a bounded IF of the transform. For the transforms used in this thesis, we have then

$$\mathrm{IF}[x;L(s)] = e^{-sx} - L(s)\,, \tag{2.50}$$

$$\mathrm{IF}[x;u(s)] = \cos(sx) - u(s)\,, \tag{2.51}$$

$$\mathrm{IF}[x;v(s)] = \sin(sx) - v(s)\,. \tag{2.52}$$

Applying (2.48) to (2.8), we also obtain

$$\mathrm{IF}[x;|\phi(s)|^2] = 2[u(s)\cos(sx) + v(s)\sin(sx)] - 2|\phi(s)|^2\,. \tag{2.53}$$

Let us analyze the properties of these IFs. Regarding the IF of LT in (2.50), two cases arise. For distributions on $\mathbb{R}_+$ or $\mathbb{R}_+^0$ (e.g. Pareto or gamma), we clearly have $|\mathrm{IF}[x;L(s)]| \leq 1$, $\lim_{x\to 0}\mathrm{IF}[x;L(s)] = 1 - L(s)$ and $\lim_{x\to\infty}\mathrm{IF}[x;L(s)] = -L(s)$ for all $s \geq 0$. Therefore, estimation of $L(s)$ using $\widehat{L}_n(s)$ becomes increasingly robust as $s \to \infty$. However, for two-sided distributions (e.g. normal, logistic), where $L(s)$ is unbounded, the estimation is completely unrobust for any $s \neq 0$. In contrast, the IFs of the CF components (2.51), (2.52) and (2.53) are bounded for all $s, x \in \mathbb{R}$ for any distribution and exhibit oscillations. Moreover, for continuous distributions, we find that for fixed x

$$\lim_{|s|\to 0}\mathrm{IF}[x;|\phi(s)|^2] = 0, \qquad \text{and} \qquad \lim_{|s|\to\infty}\mathrm{IF}[x;|\phi(s)|^2] = 0\,. \tag{2.54}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

2.4 Derivatives of integral transforms

Since the estimators proposed in this thesis rely on differential equations of transforms, we will also overview the properties of the derivatives of transforms and their sample counterparts. For generality, we consider the m-th derivative of $\mathcal{T}(s)$, namely

$$\mathcal{T}^{(m)}(s) = \int_{-\infty}^{\infty} K^{(m)}(s,x)\mathrm{d}F(x)\,. \tag{2.55}$$

In particular, the m-th derivatives of LT, CF and its real and imaginary part are

$$L^{(m)}(s) \;=\; \int_{-\infty}^{\infty}(-x)^m e^{-sx}\mathrm{d}F(x) = \mathbb{E}[(-X)^m e^{-sX}]\,, \tag{2.56}$$

$$\phi^{(m)}(s) \;=\; \int_{-\infty}^{\infty}(ix)^m e^{isx}\mathrm{d}F(x) = \mathbb{E}[(iX)^m e^{isX}]\,, \tag{2.57}$$

$$u^{(m)}(s) \;=\; \int_{-\infty}^{\infty} x^m \cos(sx+\tfrac{1}{2}m\pi)\mathrm{d}F(x) = \mathbb{E}[X^m\cos(sX+\tfrac{1}{2}m\pi)]\,, \tag{2.58}$$

$$v^{(m)}(s) \;=\; \int_{-\infty}^{\infty} x^m \sin(sx+\tfrac{1}{2}m\pi)\mathrm{d}F(x) = \mathbb{E}[X^m\sin(sX+\tfrac{1}{2}m\pi)]\,, \tag{2.59}$$

and

$$|\phi^{(m)}(s)|^2 \;=\; [u^{(m)}(s)]^2 + [v^{(m)}(s)]^2\,. \tag{2.60}$$

The interval of convergence of $\mathcal{T}^{(m)}(s)$ may not be the same as that of $\mathcal{T}(s)$. In case of LT and CF, these intervals may differ by the inclusion of $s=0$, depending on existence of $\mathbb{E}[X^m]$. For example, consider $L(s)$ of the Pareto(α,β) distribution (Table 2.2) with $\alpha = 1.5$, so that $\mathbb{E}[X]$ exists but $\mathbb{E}[X^2]$ does not. Then, $L(s)$ and $L'(s)$ exists on $\mathcal{D}=[0,\infty)$ but $L^{(m)}(s)$ exists only on $(0,\infty)$ for $m\geq 2$.

What distinguishes the m-th derivatives from their original LT or CF is the boundedness, which depends on the value and existence of the m-th moment.

2.4.1 Derivatives of empirical transforms

The empirical counterparts of (2.56)-(2.60), associated with sample $X_1,...,X_n$, are

$$\widehat{L}_n^{(m)}(s) \;=\; \frac{1}{n}\sum_{j=1}^{n}(-X_j)^m e^{-sX_j}\,, \tag{2.61}$$

$$\widehat{\phi}_n^{(m)}(s) \;=\; \frac{1}{n}\sum_{j=1}^{n}(iX_j)^m e^{isX_j}\,, \tag{2.62}$$

$$\widehat{u}_n^{(m)}(s) \;=\; \frac{1}{n}\sum_{j=1}^{n}X_j^m\cos(sX_j+\tfrac{1}{2}m\pi)\,, \tag{2.63}$$

$$\widehat{v}_n^{(m)}(s) \;=\; \frac{1}{n}\sum_{j=1}^{n}X_j^m\sin(sX_j+\tfrac{1}{2}m\pi)\,, \tag{2.64}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

and

$$|\widehat{\phi}_n^{(m)}(s)|^2 = [\widehat{u}_n^{(m)}(s)]^2 + [\widehat{v}_n^{(m)}(s)]^2 = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(X_jX_k)^m \cos(s(X_j - X_k)) . \quad (2.65)$$

We have of course $\mathbb{E}[\widehat{L}_n^{(m)}(s)] = L(s)$, $\mathbb{E}[\widehat{u}_n^{(m)}(s)] = u(s)$, $\mathbb{E}[\widehat{v}_n^{(m)}(s)] = v(s)$ and, by steps similar to those for (2.33), we find that

$$\mathbb{E}[|\widehat{\phi}_n^{(m)}(s)|^2] = |\phi^{(m)}(s)|^2 + \frac{\mathbb{E}[X^{2m}] - |\phi^{(m)}(s)|^2}{n} . \quad (2.66)$$

Using (2.17), we also obtain

$$n\mathbb{V}[\widehat{L}_n^{(m)}(s)] = L^{(2m)}(2s) - [L^{(m)}(s)]^2 , \quad (2.67)$$

$$n\mathbb{V}[\widehat{u}_n^{(m)}(s)] = \tfrac{1}{2}\mathbb{E}[X^{2m}] + \tfrac{1}{2}u^{(2m)}(2s) - [u^{(m)}(s)]^2 , \quad (2.68)$$

$$n\mathbb{V}[\widehat{v}_n^{(m)}(s)] = \tfrac{1}{2}\mathbb{E}[X^{2m}] - \tfrac{1}{2}u^{(2m)}(2s) - [v^{(m)}(s)]^2 . \quad (2.69)$$

Notice that the behavior of $\widehat{L}_n^{(m)}(s)$ is desirable: it has no bias and finite variance for all s where $L^{(2m)}(2s)$ is finite. In contrast, the bias of $|\widehat{\phi}_n^{(m)}(s)|^2$, as well as the variances of $\widehat{u}_n^{(m)}(s)$ and $\widehat{v}_n^{(m)}(s)$, depends on the value (and finiteness) of $\mathbb{E}[X^{2m}]$. Assuming finiteness of the above variances, we deduce from (2.18) that

$$\sqrt{n}\Big(\widehat{L}_n^{(m)}(s) - L(s)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \mathbb{V}[\widehat{L}_n^{(m)}(s)]\Big) , \quad (2.70)$$

$$\sqrt{n}\Big(\widehat{u}_n^{(m)}(s) - u(s)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \mathbb{V}[\widehat{u}_n^{(m)}(s)]\Big) , \quad (2.71)$$

$$\sqrt{n}\Big(\widehat{v}_n^{(m)}(s) - v(s)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \mathbb{V}[\widehat{v}_n^{(m)}(s)]\Big) . \quad (2.72)$$

The derivation of the finite-sample variance of $|\widehat{\phi}_n^{(m)}(s)|^2$ is very tedious and does not produce a useful expression. However, the asymptotic variance can be obtained by considering the corresponding U-statistic. By applying (A.9) and (A.10) from Appendix A, we obtain after a lengthy yet straightforward derivation that

$$\sqrt{n}\Big(|\widehat{\phi}_n^{(m)}(s)|^2 - |\phi^{(m)}(s)|^2\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, 2\sigma_1^2\Big), \quad (2.73)$$

where

$$\sigma_1^2 = \mathbb{E}[X^{2m}]|\phi^{(m)}(s)|^2 - |\phi^{(m)}(s)|^4 + u^{(2m)}(2s)\Big([u^{(m)}(s)]^2 - [v^{(m)}(s)]^2\Big) + v^{(2m)}(2s)u^{(m)}(s)v^{(m)}(s). \quad (2.74)$$

We conclude, therefore, that the derivatives of ECF are statistically applicable only to distributions with finite $2m$-th moment.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

2.4.2 Influence functions of the derivatives

Finally, we shall compute IFs of the derivatives of transforms. By (2.49), we have in general

$$\mathrm{IF}[x;\mathcal{T}^{(m)}(s)] = K^{(m)}(s,x) - \mathcal{T}^{(m)}(s)\,, \tag{2.75}$$

and in particular

$$\begin{aligned}
\mathrm{IF}[x;L^{(m)}(s)] &= (-x)^m e^{-sx} - L^{(m)}(s)\,, &(2.76)\\
\mathrm{IF}[x;u^{(m)}(s)] &= x^m\cos(sx+\tfrac{m\pi}{2}) - u^{(m)}(s)\,, &(2.77)\\
\mathrm{IF}[x;v^{(m)}(s)] &= x^m\sin(sx+\tfrac{m\pi}{2}) - v^{(m)}(s)\,. &(2.78)
\end{aligned}$$

By the chain rule (2.48), we also obtain

$$\mathrm{IF}[x;|\phi^{(m)}(s)|^2] = 2x^m[u^{(m)}(s)\cos(sx+\tfrac{m\pi}{2}) + v^{(m)}(s)\sin(sx+\tfrac{m\pi}{2})] - 2|\phi^{(m)}(s)|^2. \tag{2.79}$$

Notice that the IFs of the transforms' derivatives are unbounded both in x and in s. However, for suitably large s their absolute value is smaller than that of IFs (2.50)-(2.53) of the original transforms for $x\to\infty$ and/or $x\to 0$ (that is, for x representing the two types of outlying contamination). It is so because for any $m\geq 1$, $L^{(m)}(s)$ and $|\phi^{(m)}(s)|^2$ decay faster than $L(s)$ and $|\phi(s)|^2$ as $s\to\infty$ (for distributions on $\mathbb{R}_+$ and continuous distributions on $\mathbb{R}_+$ or $\mathbb{R}$, respectively). This issue is illustrated in Fig. 2.5, where we plotted $\mathrm{IF}[x;L^{(m)}(s)]$ and $\mathrm{IF}[x;|\phi^{(m)}(s)|^2]$ for $m=0,1,2$, setting either $s=0.5$ or $s=3$. The underlying distribution is exponential(λ) with $\lambda=1$.

In summary, when employing the derivatives of transforms for statistical purposes we can maintain robustness. This, in turn, extends to the robustness of resulting estimators. The expressions (2.76)-(2.79) will be used in Chapter 4 to derive more intricate IFs for our new estimators.

2.5 Summary

This chapter gave a comparative review of LTs and CFs along with their sample counterparts. We highlighted distinctions in the transform properties in connection with types of continuous distributions (one- versus two-sided). Boundedness emerged as a fundamental property, ensuring both bounded IF and bounded variance of the empirical transform. In this context, CF is a more universal tool than LT which is not unbounded for two-sided distributions. On the other hand, considering the analytical difficulties associated with the complex-valued CF and the oscillation of its components, LT appears to be more useful for one-sided (life-time) models. The LT holds an additional advantage in terms of derivatives: the derivatives of ELT are unbiased and consistently exhibit finite variance for some $s\in\mathcal{D}$. In contrast, the bias and variance of the squared modulus of the m-th derivative of CF depend on the $2m$-th moment, and, thus, may be infinite.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

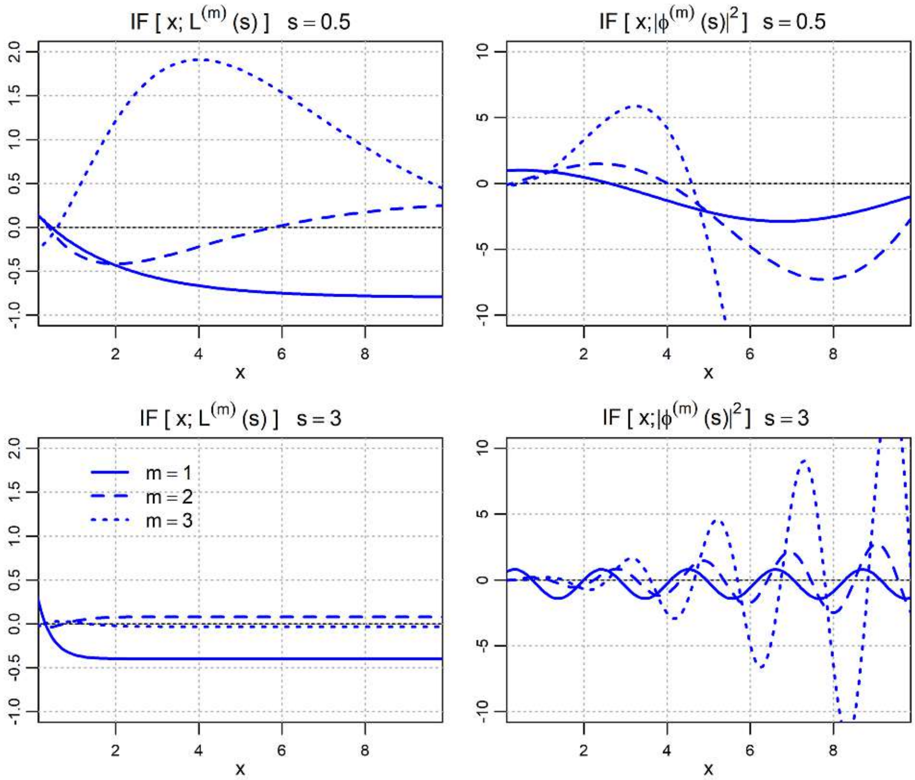


Figure 2.5: IFs of $L^{(m)}(s)$ and $|\phi^{(m)}(s)|^2$, with $s=0.5$ or $s=3$ for the exponential(1) distribution; $m = 0, 1, 2$. For fixed contamination x, the robustness of the transforms' derivatives decreases with the derivative order m.

With the introduction of the derivatives of LTs and CFs, the inherent boundedness of transforms is completely lost. Nevertheless, as demonstrated in the last section, the IF of any order derivative has lower magnitude compared to the IF of the transform itself as $s \to \infty$ (for LT) or as $|s| \to \infty$ (for the squared modulus of CF).

The reviewed properties determine the construction of the estimators in next chapters and help explain their performance.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

CHAPTER 3

Differential equations satisfied by transforms

This chapter explores methods of obtaining the differential equations of LTs and CFs of probability distributions and their usability for statistical inference.

3.1 Direct differentiation

We begin by noting that for transforms with closed-form expressions, differential equation can be derived by direct differentiation. This method applies as long as the result of differentiation contains the transform expression as a factor or summand. Let us consider some examples.

3.1.1 Examples

3.1.1.1 Gamma distribution

The LT of the gamma distribution is given by

$$L(s) = \left(\frac{\beta}{\beta + s}\right)^{\alpha}, \qquad s > -\beta. \tag{3.1}$$

By taking derivative of both sides w.r.t to s, we obtain

$$L'(s) = -\alpha\left(\frac{\beta}{\beta + s}\right)^{\alpha-1}\frac{\beta}{(\beta + s)^2}\,, \tag{3.2}$$

which implies

$$(\beta + s)L'(s) + \alpha L(s) = 0\,. \tag{3.3}$$

Repeated differentiation leads to higher order equations. In particular, equations involving only the k-th and $(k-1)$-th derivatives can be written in general as

$$(\beta + s)L^{(k)}(s) - (\alpha + k - 1)L^{(k-1)}(s) = 0\,, \tag{3.4}$$

whereas those consisting only of the k-th derivative and the transform itself are

$$(\beta + s)^k L^{(k)}(s) + \prod_{j=0}^{k-1}(\alpha + j)L(s) = 0\,. \tag{3.5}$$

3.1.1.2 Normal distribution

To illustrate equations satisfied by CFs, we consider that of the normal distribution

$$\phi(s) = \exp\{i\mu s - \sigma^2 s^2/2\}\,, \qquad s \in \mathbb{R}. \tag{3.6}$$

Taking derivative of both sides leads to

$$\phi'(s) - (i\mu - \sigma^2 s)\phi(s) = 0, \tag{3.7}$$

and, by doing it k times, we find that

$$\phi^{(k)}(s) - (i\mu - \sigma^2 s)\phi^{(k-1)}(s) + (k-1)\sigma^2\phi^{(k-2)}(s) = 0. \tag{3.8}$$

Generalizing similar to (3.5) in this case is challenging.

3.1.1.3 Rayleigh distribution

The final example concerns a transform expressed in terms of a special function but with a manageable derivative. In Meintanis and Iliopoulos (2003b), we can find that LT of the Rayleigh(ϑ) distribution is

$$L(s) = 1 - s\vartheta\sqrt{\frac{\pi}{2}}\exp\left(\frac{\vartheta^2 s^2}{2}\right)\operatorname{erfc}\left(\frac{\vartheta s\sqrt{2}}{2}\right), \qquad s \in \mathbb{R}. \tag{3.9}$$

Given the derivative of the complementary error function, $\operatorname{erfc}'(s) = -2\exp\{-s^2\}/\sqrt{\pi}$, it is not difficult to show that

$$sL'(s) - (1 + s^2\vartheta^2)L(s) + 1 = 0. \tag{3.10}$$

For the CF, $\phi(s) = L(-is)$, we obtain in a similar manner

$$s\phi'(s) - (1 - s^2\vartheta^2)\phi(s) + 1 = 0. \tag{3.11}$$

The first order ODEs which can be obtained throught direct differentiation of the transforms listed in Tables 2.1-2.2 and 2.3 are compiled in Table 3.1 on p. 45.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

3.2 Alternative method of deriving the differential equations

This section explores distributions with LTs and CFs lacking closed-form expressions, making direct computation of their derivatives impractical or impossible. Examples falling into this category occurred in Table 2.2. Fortunately, in quite many such cases, the desired equations can be found by an 'indirect' differentiation approach recently proposed by Gaunt (2021) (who presented it for CF of the Student-t distribution). We will refer to this approach as the *alternative method*. Its somewhat generalized description is provided below.

The method starts by first identifying the differential equation satisfied by the density $f(x)$ in question, and, it applies when the equation has coefficients being polynomials in x. The process involves the following steps:

1. Differentiate the density $f(x) = f(x|\boldsymbol{\theta})$ w.r.t. x and obtain a differential equation.
2. Rewrite the equation in the following form:

$$\sum_{k=0}^{m} w_k(\boldsymbol{\theta}) x^k f(x) \ + \ \sum_{k=0}^{r} \xi_k(\boldsymbol{\theta}) x^k f'(x) = 0\,, \tag{3.12}$$

   where $w_0(\boldsymbol{\theta}), ..., w_m(\boldsymbol{\theta})$ and $\xi_0(\boldsymbol{\theta}), ..., \xi_r(\boldsymbol{\theta})$ are some functions of the parameter.

3. Compute the LT or CF of both sides.

   In case of LT, this computation leads to

$$\sum_{k=0}^{m} w_k(\boldsymbol{\theta}) \int\limits_{\substack{0\\(-\infty)}}^{\infty} x^k f(x) e^{-sx} \mathrm{d}x \ + \ \sum_{k=0}^{r} \xi_k(\boldsymbol{\theta}) \int\limits_{\substack{0\\(-\infty)}}^{\infty} x^k f'(x) e^{-sx} \mathrm{d}x = 0\,, \tag{3.13}$$

   whereas in the case of CF it gives

$$\sum_{k=0}^{m} w_k(\boldsymbol{\theta}) \int\limits_{\substack{0\\(-\infty)}}^{\infty} x^k f(x) e^{isx} \mathrm{d}x \ + \ \sum_{k=0}^{r} \xi_k(\boldsymbol{\theta}) \int\limits_{\substack{0\\(-\infty)}}^{\infty} x^k f'(x) e^{isx} \mathrm{d}x = 0\,, \tag{3.14}$$

   (where the lower limits of integration are set to 0 for one-sided distributions and $-\infty$ for the two-sided ones).

   Now, computing the first integrals in (3.13) and (3.14) is straightforward using

$$\int\limits_{\substack{0\\(-\infty)}}^{\infty} x^k f(x) e^{-sx} \mathrm{d}x = (-1)^k L^{(k)}(s)\,, \tag{3.15}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

and
$$\int_{\substack{0\\(-\infty)}}^{\infty} x^k f(x) e^{isx} \mathrm{d}x = (-i)^k \phi^{(k)}(s)\,, \tag{3.16}$$

the well-known properties of the Laplace and Fourier transforms; see e.g. Poularikas (2010) and recall secs. 2.1.1.4 and 2.1.3.3.

The second integrals in (3.13) and (3.14) can be computed by parts, as follows:

$$\int_{\substack{0\\(-\infty)}}^{\infty} x^k f'(x) e^{-sx} \mathrm{d}x \;=\; \underbrace{\Big[x^k f(x) e^{-sx}\Big]_{\substack{0\\(-\infty)}}^{\infty}}_{A_{\mathrm{LT}}} - k\int_{\substack{0\\(-\infty)}}^{\infty} x^{k-1} f(x) e^{-sx} \mathrm{d}x \,+\, s\int_{\substack{0\\(-\infty)}}^{\infty} x^k f(x) e^{-sx} \mathrm{d}x,$$

and

$$\int_{\substack{0\\(-\infty)}}^{\infty} x^k f'(x) e^{isx} \mathrm{d}x \;=\; \underbrace{\Big[x^k f(x) e^{isx}\Big]_{\substack{0\\(-\infty)}}^{\infty}}_{A_{\mathrm{CF}}} - k\int_{\substack{0\\(-\infty)}}^{\infty} x^{k-1} f(x) e^{isx} \mathrm{d}x \,-\, is\int_{\substack{0\\(-\infty)}}^{\infty} x^k f(x) e^{isx} \mathrm{d}x.$$

The derivations will be successful if the terms A_{LT} or A_{CF} in the above two lines are finite. This condition may impose some restrictions on the parameter space, or the proof of validity of the final equation may necessitate additional arguments for parameter values where $A_{\mathrm{LT}} = \infty$ and/or $A_{\mathrm{CF}} = \infty$). However, it is noteworthy that, for most distributions studied in this thesis, the terms A_{LT} or A_{CF} not only turned out to be finite but they equal zero over the entire parameter space. This is demonstrated by examples given in the next section.

4. If A_{LT} and/or A_{CF} are indeed zero, we can leverage (3.15) and (3.16), along with the derivative properties of both transforms, to obtain

$$\int_{\substack{0\\(-\infty)}}^{\infty} x^k f'(x) e^{-sx} \mathrm{d}x = \begin{cases} sL(s) - f(0^+), & k=0 \text{ and support of } f(x) \text{ is } \mathbb{R}_+ \text{ or } \mathbb{R}_+^0 \\ sL(s), & k=0 \text{ and support of } f(x) \text{ is } \mathbb{R}\,, \\ (-1)^k[sL^{(k)}(s) + kL^{(k-1)}(s)], & k=1,2,\ldots \quad \text{in any case}\,; \end{cases} \tag{3.17}$$

and

$$\int_{\substack{0\\(-\infty)}}^{\infty} x^k f'(x) e^{isx} \mathrm{d}x = \begin{cases} -is\phi(s) - f(0^+), & k=0 \text{ and support of } f(x) \text{ is } \mathbb{R}_+ \text{ or } \mathbb{R}_+^0 \\ -is\phi(s), & k=0 \text{ and support of } f(x) \text{ is } \mathbb{R}\,, \\ (-1)^k i^{k-1}[s\phi^{(k)}(s) + k\phi^{(k-1)}(s)], & k=1,2,\ldots \quad \text{in any case}\,. \end{cases} \tag{3.18}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

5. Substituting (3.15) and (3.17) into (3.13) or (3.16) and (3.18) into (3.14), yields the desired differential equations satisfied by either LT or CF.

If the differential equation satisfied by $f(x)$ is not polynomial in x, unfortunately, the method cannot be applied, at least in the present version. This limitation arises, for instance, in the case of the Weibull(c, λ) distribution with a non-integer c or for the log-normal distribution. The following examples illustrate situations where the method can be effectively employed.

3.2.1 Examples

3.2.1.1 Pareto-Lomax distribution

The Pareto(α,β) distribution of the Lomax type has the density

$$f(x) = \frac{\alpha\beta^{\alpha}}{(x+\beta)^{\alpha+1}}\,, \quad x \geq 0\,, \;\; \alpha, \beta > 0\,. \tag{3.19}$$

Differentiating it w.r.t x we get

$$f'(x) = -f(x)\left(\frac{\alpha+1}{x+\beta}\right), \tag{3.20}$$

which leads to the following differential equation

$$(\alpha+1)f(x) + \beta f'(x) + xf'(x) = 0\,, \tag{3.21}$$

whose coefficients are polynomials in x. Comparing the structure of this equation with (3.12), we observe that here $m = 0$, $w_0(\boldsymbol{\theta}) = (\alpha+1)$, and $r = 1$, $\xi_0(\boldsymbol{\theta}) = \beta$, $\xi_1(\boldsymbol{\theta}) = 1$. Now, computing LT of both sides yields

$$(\alpha+1)\int_0^{\infty} f(x)e^{-sx}\mathrm{d}x + \beta\int_0^{\infty} f'(x)e^{-sx}\mathrm{d}x + \int_0^{\infty} xf'(x)e^{-sx}\mathrm{d}x = 0\,. \tag{3.22}$$

We utilize (3.15) to compute the first integral and (3.17) to compute the second and third ones. Since the term of the type A_{LT} is zero, we can apply (3.17), and we have $\int_0^{\infty} f'(x)e^{-sx}\mathrm{d}x = sL(s) - f(0)$ and $\int_0^{\infty} xf'(x)e^{-sx}\mathrm{d}x = -L(s) - sL'(s)$; we also have $f(0) = \alpha/\beta$. By combining these results and rearranging, we arrive at

$$(\alpha+\beta s)L(s) - sL'(s) - \alpha = 0\,. \tag{3.23}$$

The differential equation satisfied by CF is found in a similar manner. First, by taking the CF of both sides of (3.21), we obtain

$$(\alpha+1)\int_0^{\infty} f(x)e^{isx}\mathrm{d}x + \beta\int_0^{\infty} f'(x)e^{isx}\mathrm{d}x + \int_0^{\infty} xf'(x)e^{isx}\mathrm{d}x = 0\,. \tag{3.24}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Given that the terms of the type A_{CF} are zero, we can use (3.18), where we have $\int_0^\infty f'(x)e^{isx}\mathrm{d}x = -is\phi(s) - f(0)$ and $\int_0^\infty xf'(x)e^{-sx}\mathrm{d}x = -\phi(s) - s\phi'(s)$. Rearranging the terms gives

$$(\alpha - i\beta s)\phi(s) - s\phi'(s) - \alpha = 0\,. \tag{3.25}$$

3.2.1.2 Generalized inverse Gaussian distribution

The density of this three-parameter distribution is

$$f(x) = \frac{\mu^p x^{p-1}}{2\mathrm{K}_p(\lambda/\mu)} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x,\,\lambda,\,\mu\,>\,0, \quad p \in \mathbb{R}, \tag{3.26}$$

where $\mathrm{K}_p(\cdot)$ stands for the modified Bessel function of 2nd kind. Note that, usually, authors work with a different parametrization; see Jørgensen (1982) or sec. 9.3 in Johnson et al. (1994) for details. Putting $p = -1/2$ yields the classical inverse Gaussian law, as presented in Table 2.1. Upon differentiating both sides of (3.26), we obtain

$$\lambda f(x) + 2(p-1)xf(x) - (\lambda/\mu^2)x^2 f(x) - 2x^2 f'(x) = 0\,. \tag{3.27}$$

Computing the LT of both sides, we apply (3.15) with $k = 0, 1, 2$, as well as (3.17) with $k = 2$ (allowed as the term of type A_{LT} is zero). This yields

$$(2s{+}\lambda/\mu^2)L''(s) + 2(p+1)L'(s) - \lambda L(s) = 0. \tag{3.28}$$

Similarly, taking CF of both sides of (3.27) and using (3.16) along with (3.18) for the same k-values, leads to

$$(2is{+}\lambda/\mu^2)\phi''(s) + 2i(p+1)\phi'(s){-}\lambda\phi(s) = 0. \tag{3.29}$$

3.2.1.3 Beta distributions of 1st and 2nd kind

By differentiating the density of the beta distribution of the 1st kind, namely

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\,, \quad 0 \le x \le 1\,,\ \alpha,\beta > 0\,, \tag{3.30}$$

we find that it satisfies

$$(1-\alpha)f(x) + (\alpha+\beta-2)xf(x) + xf'(x) - x^2 f'(x) = 0\,. \tag{3.31}$$

After taking LT or CF of both sides, all the integrals can be shown to be finite (even for $0 < \beta < 1$). Following the previously established steps, we obtain

$$sL''(s) + (\alpha+\beta+s)L'(s) + \alpha L(s) = 0\,, \tag{3.32}$$

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

and

$$s\phi''(s) + (\alpha + \beta - is)\phi'(s) - i\alpha\phi(s) = 0\,. \tag{3.33}$$

Since the density of $X \sim$ beta(α,β) satisfies a differential equation with coefficients that are polynomials in x, the same should hold for the distribution of $X/(1-X)$, which is the beta distribution of the 2nd kind. Its density is expressed as

$$f(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{\mathrm{B}(\alpha,\beta)}\,, \quad x \geq 0\,,\ \alpha,\beta > 0\,, \tag{3.34}$$

and, indeed, by differentiation w.r.t x, we find that

$$(1-\alpha)f(x) + (1+\beta)xf(x) + xf'(x) + x^2 f'(x) = 0\,. \tag{3.35}$$

The established steps lead to

$$sL''(s) + (1 - s - \beta)L'(s)\ -\ \alpha L(s) = 0\,, \tag{3.36}$$

and

$$s\phi''(s) + (1 + is - \beta)\phi'(s) + i\alpha\phi(s) = 0\,. \tag{3.37}$$

Table 3.2 presents differential equations of transforms derived by the alternative method for a group of standard distributions. The table also includes equations for the Student-t, Maxwell, Rayleigh, inverted Rayleigh, inverted Maxwell and inverted gamma distributions. Derivations thereof are similar to those given in the above examples. We notice that if a transform of distribution of X can be handled by the method, then, in principle, the transform related to the inverted variable $1/X$ can be handled too. The derivation of the equation for CF of the Student-t(ν), which requires additional arguments for degrees of freedom $\nu = 1$ and 2, is given in the already cited work of Gaunt (2021). All the presented equations have been validated by Monte Carlo simulations.

3.3 Remarks on the differential equations

As seen in the above examples, the differential equations satisfied by LTs and CFs are linear ordinary differential equations (ODEs) of order $k \geq 1$. Their scope of validity is identical with $\mathcal{D}$ (the interval of convergence of the particular transform) if $\mathbb{E}[X^k]$ is finite, or the scope equals $\mathcal{D}\backslash\{0\}$ if this moment does not exist. Inclusion of $s = 0$ may depend, in some cases, on the parameter value (e.g., the Pareto case).

All the equations derived by direct differentiation (Table 3.1) are generically of first order. In contrast, those produced by the alternative method (Table 3.2) vary in order, with each determined by the highest power of x present in the formula of the underlying

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

density. All the obtained equations are homogenous, except for the uniform, Pareto and Rayleigh cases, where free terms (independent of s) are involved[1].

Finally, note that the coefficients of the equations may be linear or non-linear in $\boldsymbol{\theta}$ and s. As we will see, linearity in $\boldsymbol{\theta}$ guarantees closed-form expressions for the estimators to be defined in the next chapter. In turn, the GoF statistics based on these equations (which do not require a solution for $\boldsymbol{\theta}$ but a substitution of its independent estimate) may admit closed forms without linearity in $\boldsymbol{\theta}$.

3.4 Conditions for consistency in statistical applications

Consistency of statistical procedures based on the presented differential equations will depend on uniqueness of their solutions. However, it is crucial to carefully differentiate this requirement with respect to GoF testing and parameter estimation.

3.4.1 Consistency of testing

To ensure the consistency of a GoF test (carried out with an independent consistent estimator $\widehat{\boldsymbol{\theta}}_n$, such as ML or MM), we should require that the transform of the tested distribution is the *only* solution to the particular equation within the class of all probability transforms of that type (LT/CF), subject to specific initial conditions. Among the LT-based equations in Tables 3.1 and 3.2, such uniqueness is guaranteed for all the distributions with finite moments of the appropriate order. This assurance stems from the well-known Existence and Uniqueness Theorem for linear ODEs with continuous coefficients; see, for instance, Farlow (1994) p. 189. In particular, the initial conditions for the k-th order equations of that type are as follows:

$$\begin{aligned} L(0) &= 1, \\ L^{(1)}(0) &= -\mathbb{E}[X], \\ &\vdots \\ L^{(k-1)}(0) &= (-1)^{k-1}\mathbb{E}[X^{k-1}]. \end{aligned} \tag{3.38}$$

As the theory for the solvability and uniqueness of solutions to linear ODEs holds consistently whether the functions and coefficients are real or complex, the above considerations also extend to CF-based equations. Proving the uniqueness for both the LT and CF-based equations when (some) moments of the distributions are infinite is a more intricate task.

This uniqueness in most general sense turns the corresponding GoF test into an *omnibus test*, meaning it is consistent against all distributional alternatives. Consistency of a test in a smaller class of distributions requires, understandably, more refined analysis.

[1] The case of the Rayleigh distribution was the only one where we managed to obtain differential equations using both methods. Specifically, the second-order equations obtained via the alternative method (see Table 3.2) follow from differentiation of the first-order ones, obtained by the direct differentiation approach in (3.11).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Table 3.1: Differential equations of transforms obtained by direct differentiation.

distribution	differential equation
normal(μ, σ^2) $x, \mu \in \mathbb{R},\ \sigma^2 > 0$	$L'(s) + (\mu - \sigma^2 s)L(s) = 0, \quad s \in \mathbb{R}$ $\phi'(s) - (i\mu - \sigma^2 s)\phi(s) = 0, \quad s \in \mathbb{R}$
exponential(λ) $x \geq 0,\ \lambda > 0$	$(\lambda + s)L'(s) + L(s) = 0, \quad s \in (-1/\lambda, \infty)$ $(i\lambda + s)\phi'(s) + \phi(s) = 0, \quad s \in \mathbb{R}$
gamma(α, β) $x > 0,\ \alpha, \beta > 0$	$(\beta + s)L'(s) + \alpha L(s) = 0, \quad s \in (-\beta, \infty)$ $(i\beta + s)\phi'(s) + \alpha\phi(s) = 0, \quad s \in \mathbb{R}$
inv.Gauss.(λ, μ) $x, \lambda, \mu > 0$	$\sqrt{1 + 2\mu^2 s/\lambda}\, L'(s) + \mu L(s) = 0, \quad s \in [-\lambda/2\mu, \infty)$ $\sqrt{1 - 2\mu^2 is/\lambda}\, \phi'(s) + i\mu\phi(s) = 0, \quad s \in \mathbb{R}$
Rayleigh(ϑ) $x \geq 0,\ \vartheta > 0$	$sL'(s) - (1 + s^2\vartheta^2)L(s) + 1 = 0, \quad s \in \mathbb{R}$ $s\phi'(s) - (1 - s^2\vartheta^2)\phi(s) + 1 = 0 \quad s \in \mathbb{R}$
Laplace(μ, β) $x, \mu \in \mathbb{R},\ \beta > 0$	$(1 - \beta^2 s^2)L'(s) + [\mu(1 - \beta^2 s^2) - 2\beta^2 s]L(s) = 0, \quad s \in (-\frac{1}{\beta}, \frac{1}{\beta})$ $(1 + \beta^2 s^2)\phi'(s) - [i\mu(1 + \beta^2 s^2) - 2\beta^2 s]\phi(s) = 0, \quad s \in \mathbb{R}$
logistic(μ, β) $x, \mu \in \mathbb{R},\ \beta > 0$	$L'(s) + [\mu + \cot(\pi\beta s)\pi\beta - 1/s]L(s) = 0, \quad s \in (-1/\beta, 1/\beta)$ $\phi'(s) - [i\mu - \coth(\pi\beta s)\pi\beta + 1/s]\phi(s) = 0, \quad s \in \mathbb{R}$
uniform(a, b) $x \in (a,b),\ a,b \in \mathbb{R},\ b > a$	$sL'(s) + L(s) - (be^{-sb} - ae^{-sb})/(b - a) = 0, \quad s \in \mathbb{R}$ $s\phi'(s) + \phi(s) - (be^{isb} - ae^{isa})/(b - a) = 0, \quad s \in \mathbb{R}$
Cauchy(μ, γ) $x, \mu \in \mathbb{R},\ \beta > 0$	$\phi'(s) - (i\mu + \beta)\phi(s) = 0 \quad s < 0$ $\phi'(s) - (i\mu - \beta)\phi(s) = 0, \quad s > 0$
stable(μ,γ,α) $x, \mu \in \mathbb{R},\ \gamma > 0,\ \alpha \in (0,2]$	$\phi'(s) - (i\mu - \gamma\alpha s^{\alpha-1})\phi(s) = 0, \quad s < 0$ $\phi'(s) - (i\mu + \gamma\alpha s^{\alpha-1})\phi(s) = 0, \quad s > 0$
Poisson(λ) $x = 0,1,2,\ldots,\ \lambda > 0$	$\phi'(s) - i\lambda e^{is}\phi(s) = 0, \quad s \in \mathbb{R}$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Table 3.2: Differential equations of transforms obtained by the alternative method.

distribution	differential equation
Pareto(α,β) $x\geq 0,\ \alpha,\beta>0$	$sL'(s)-(\alpha+\beta s)L(s)+\alpha=0, \quad s>0$ * $s\phi'(s)-(\alpha-i\beta s)\phi(s)+\alpha=0, \quad s\in\mathbb{R}\backslash\{0\}$ *
Beta(α,β) $0<x<1,\ \alpha,\beta>0$	$sL''(s)+(\alpha+\beta+s)L'(s)+\alpha L(s)=0, \quad s\in\mathbb{R}$ $s\phi''(s)+(\alpha+\beta-is)\phi'(s)-i\alpha\phi(s)=0, \quad s\in\mathbb{R}$
Beta-2(α,β) $x\geq 0,\ \alpha,\beta>0$	$sL''(s)+(1-s-\beta)L'(s)-\alpha L(s)=0, \quad s>0$ * $s\phi''(s)+(1+is-\beta)\phi'(s)+i\alpha\phi(s)=0, \quad \mathbb{R}\backslash\{0\}$ *
inv.Gauss.(λ,μ) $x,\lambda,\mu>0$	$(2s+\lambda/\mu^2)L''(s)+L'(s)-\lambda L(s)=0, \quad s\in[-\frac{\lambda}{2\mu},\infty)$ $(2is-\lambda/\mu^2)\phi''(s)+i\phi'(s)-\lambda\phi(s)=0, \quad s\in\mathbb{R}$
GIG(λ,μ,p) $x,\lambda,\mu>0,\ p\in\mathbb{R}$	$(2s+\lambda/\mu^2)L''(s)+2(p+1)L'(s)-\lambda L(s)=0, \quad s\in[-\frac{\lambda}{2\mu},\infty)$ $(2is+\lambda/\mu^2)\phi''(s)+2i(p+1)\phi'(s)-\lambda\phi(s)=0, \quad s\in\mathbb{R}$
Rayleigh(ϑ) $x\geq 0,\ \vartheta>0$	$L''(s)-\vartheta^2 sL'(s)-2\vartheta^2 L(s)=0, \quad s\in\mathbb{R}$ $\phi''(s)+\vartheta^2 s\phi'(s)+2\vartheta^2\phi(s)=0, \quad s\in\mathbb{R}$
Maxwell(σ) $x,\sigma>0$	$3\sigma^2 L(s)+\sigma^2 sL'(s)-L''(s)=0, \quad s\in\mathbb{R}$ $3\sigma^2\phi(s)+\sigma^2 s\phi'(s)+\phi''(s)=0, \quad s\in\mathbb{R}$
inv.Rayleigh(ϑ) $x,\vartheta>0$	$s\vartheta^2 L'''(s)+L(s)=0, \quad s>0$ *
inv.Maxwell(σ) $x,\sigma>0$	$sL'''(s)-L''(s)-(1/\sigma^2)L(s)=0, \quad s>0$ *
inv. gamma(α,β) $x>0,\ \alpha,\beta>0$	$sL''(s)-(\alpha-1)L'(s)-\beta L(s)=0, \quad s>0$ *
Student-t(ν) $x\in\mathbb{R},\nu>0$	$s\phi''(s)-(\nu-1)\phi'(s)-\nu s\phi(s)=0, \quad s\in\mathbb{R}\backslash\{0\}$ *

* In these cases, the domain of the differential equation may include $s=0$, provided the moment of the order equal to the order the equation exists.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

3.4.2 Uniqueness within the model and estimation consistency

Consistency of estimators based on the differential equations relies on slightly different reasoning. Namely, when constructing such an estimator (in the way outlined in the introduction and further developed in the upcoming Chapter 4), we need to verify that our equation is satisfied by the transform of the assumed distribution for only *a single value* of the parameter, *regardless of the initial conditions*. It is not necessary to prove that the transform of no other distribution satisfies this equation, as the estimation is conducted solely within the presumed distributional family $F_{\boldsymbol{\theta}}(x)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

This narrower sense of uniqueness certainly holds for all the two-term homogenous equations of the form

$$h(\boldsymbol{\theta}, s)L^{(r)}(s) + L^{(m)}(s) = 0\,. \tag{3.39}$$

To see this, let us consider the same equation also satisfied for some $\boldsymbol{\theta}_* \neq \boldsymbol{\theta}$, which is

$$h(\boldsymbol{\theta}_*, s)L^{(r)}(s) + L^{(m)}(s) = 0\,, \tag{3.40}$$

and subtract it from the above one. We obtain

$$\Big[h(\boldsymbol{\theta}, s) - h(\boldsymbol{\theta}_*, s)\Big]L^{(r)}(s) = 0\,, \tag{3.41}$$

which implies either that $\boldsymbol{\theta}_* = \boldsymbol{\theta}$ (contradiction) or that $L^{(r)}(s) = 0$, which is impossible for any probability LT.

However, for equations involving three or more derivatives, the uniqueness may not hold in general. As a counter-example, consider the following (artificially constructed) second-order equation with scalar parameter $\boldsymbol{\theta} = \lambda > 0$,

$$(2\lambda + s)(\lambda + s)L''(s) + 2(3\lambda + 2s)L'(s) + L(s) = 0\,. \tag{3.42}$$

We can easily check that it is satisfied both by $L(s) = \lambda/(\lambda + s)$ and $L(s) = 2\lambda/(2\lambda + s)$, representing the LTs of the exponential distribution with means $1/\lambda$ and $1/2\lambda$, respectively. Of course, by specifying concrete initial conditions, such as $L(0) = 1$ and $L'(0) = -\mathbb{E}[X] = -1/\lambda$ or $L(0) = 1$ or $L'(0) = -\mathbb{E}[X] = -1/2\lambda$, the solution becomes unique. Note, however, that in a real-world situation, an estimator obtained from the empirical version of (3.42), does not 'know' this condition. In result, some realizations in the sample $X_1, ..., X_n$ will pull the estimator towards λ, while others pull it towards 2λ, yielding an in-between limiting value as $n \to \infty$ and, hence, leading to an inconsistent behavior.

Therefore, the condition discussed here, which we term *parameter-uniqueness within the model*, must be individually examined for each distribution and specific equation intended for constructing estimators. Let us now examine two additional examples from Table 3.2 in this context.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

The three-term equation satisfied by the Rayleigh(ϑ) LT is

$$L''(s) - \vartheta^2 sL'(s) - 2\vartheta^2 L(s) = 0\,. \tag{3.43}$$

Let us assume that it is satisfied by a Rayleigh LT with a different $\vartheta_* \neq \vartheta$, so that

$$L''(s) - \vartheta_*^2 sL'(s) - 2\vartheta_*^2 L(s) = 0\,. \tag{3.44}$$

Subtracting (3.44) from (3.43), we get

$$(\vartheta^2 - \vartheta_*^2)[sL'(s) + 2L(s)] = 0\,. \tag{3.45}$$

This shows that either $\vartheta = \vartheta_*$, as we wish, or that

$$sL'(s) + 2L(s) = 0\,. \tag{3.46}$$

The last line, however, implies that $L(s) = ks^{-2}$, for some constant k, which is impossible because for any probability distribution $L(0) = 1$. This verifies the parameter uniqueness within the Rayleigh model regarding the Eq. (3.43). By the same argument the uniqueness holds for the corresponding third-order ODE, namely

$$L'''(s) - \vartheta^2 sL''(s) - 3\vartheta^2 L'(s) = 0\,. \tag{3.47}$$

Both equations will be considered in sec. 4.5.3 for the construction of estimators of ϑ.

As the second example, consider the Pareto(α, β) LT satisfying the non-homogenous equation

$$sL'(s) - (\alpha + \beta s)L(s) + \alpha = 0. \tag{3.48}$$

Assume that the equation is also satisfied by the Pareto LT with another pair of parameters, α_* and β_*, namely

$$sL'(s) - (\alpha_* + \beta_* s)L(s) + \alpha_* = 0. \tag{3.49}$$

The difference of the two gives

$$(\alpha_* - \alpha)L(s) + (\beta_* - \beta)sL(s) = \alpha_* - \alpha, \tag{3.50}$$

which holds when $\alpha_* = \alpha$ and $\beta_* = \beta$. If $\alpha_* \neq \alpha$ and $\beta_* \neq \beta$, then

$$L(s) = \frac{\alpha_* - \alpha}{\alpha_* - \alpha + (\beta_* - \beta)s}, \tag{3.51}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

which is not the LT of the Pareto but LT of the exponential distribution, and, in addition, it satisfies neither (3.48) nor (3.49). Hence, the parameter uniqueness is verified in the Pareto case as well.

Verification of the CF-based versions of the above equations follows analogously.

3.5 Differential equations for non-standard distributions

The integral transforms discussed so far were associated with the standard probability distributions like normal, exponential, Pareto, inverse Gaussian. However, as explained in the introduction, the crucial role of transforms in probability is to cope with non-standard distributions for which densities and, hence, likelihood functions are either exceptionally complex or non-derivable. Many such distributions, encountered especially in fields of risk, insurance, queueing or inventory management, involve summation, compounding, and mixing of standard random variables. We now explore several examples to signal the broader potential of the transform methods, with a particular emphasis on those utilizing differential equations.

3.5.1 Lagged-normal distribution

Only a few probability distributions are closed under variable summation, implying that $X_1 + ... + X_k$ shares the same distribution as the summands X_j's, with possibly altered parameters. This property holds, for example, in cases of the Poisson and normal variables, as well as the gamma variables with the same scale parameter. In most other cases, however, the analytical form of the sum density, resulting from multiple convolutions of the summands' densities, cannot be explicitly computed. The same challenge applies to sums of variables distributed heterogeneously.

In connection with the former scenario, consider the sum

$$Z = X + Y,$$

where $X \sim \text{normal}(\mu,\, \sigma^2)$ with density $f_X(x|\mu,\sigma^2)$, and $Y \sim \text{gamma}(\alpha,\, \beta)$ with density $f_Y(y|\alpha,\beta)$. The distribution of Z is called the *lagged-normal distribution*. If X and Y are independent, the density of Z can be expressed by a convolution integral,

$$\begin{aligned} f_Z(z|\mu,\sigma^2,\alpha,\beta) &= \int_0^\infty f_X(z-y|\mu,\sigma^2) f_Y(y|\alpha,\beta)\mathrm{d}y \\ &= \frac{\beta^\alpha}{\sqrt{2\pi}\sigma\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} \exp\left\{\frac{-[(z-y-\mu)^2 + 2\beta\sigma^2 y]}{2\sigma^2}\right\}\mathrm{d}y\,, \quad (3.52) \end{aligned}$$

which does not admit a tractable expression.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

On the other hand, according to the convolution property (see e.g. Feller (1971)), the LT (or CF) of Z is simply the product of LTs (or CFs) of X and Y,

$$L_Z(s|\mu,\sigma^2,\alpha,\beta) = L(s) = \exp\{-\mu s + \sigma^2 s^2/2\}\left(\frac{\beta}{\beta+s}\right)^{\alpha}, \qquad s > -\beta. \quad (3.53)$$

While the expression is explicit, statistical inference based on it is challenging from a computational viewpoint. Working with the associated differential equation, however, is more convenient. By taking derivatives of both sides of (3.53), we arrive at

$$L'(s) + [\mu - \sigma^2 s + \alpha/(\beta+s)]L(s) = 0\,. \quad (3.54)$$

This ODE is of the form we obtained for a series of standard distributions, namely $L'(s) + h(\boldsymbol{\theta}, s)L(s) = 0$, with $\boldsymbol{\theta} = (\sigma^2, \mu, \alpha, \beta)^\top$, and may be utilized to derive estimators of the parameters from a sample $\{Z_1, ..., Z_n\}$. The estimators can be of the $\mathcal{L}^2$ type, as defined in (1.10), or of the transform-matching type defined in (1.11). However, as the coefficient in front of $L(s)$ is not linear in the elements of $\boldsymbol{\theta}$, the estimators will not be explicit. Rather, to obtain explicit estimates one must use a set of differential equations satisfied by $\log L(s)$ (or, in the CF-based setting, by $\log \phi(s)$); see ch. 4 of Campbell (1992) for full derivation.

3.5.2 Normal variance-gamma distribution

Many non-standard distributions are derived through randomization (mixing) of their parameters. In this process, a parameter is treated as a random variable, and the resultant density is obtained by integrating the original one over that of the parameter. The usual goal of mixing is to enhance the model's flexibility.

An important example arises with the normal(μ,σ^2) distribution in which the variance σ^2 is treated as a gamma(α,β) variable, forming the so-called *normal variance-gamma mixture*. Such distribution features longer tails than a regular normal model, providing an alternative to the Student t-distribution while maintaining finiteness of all moments and closure under summation. Applications include finance, where it is employed to model log returns of stock indices; see, e.g., Fergusson and Platen (2006).

The density of the mixture can be written as

$$f(x|\mu,\alpha,\beta) = \frac{\beta^\alpha}{\sqrt{2\pi}\Gamma(\alpha)}\int_0^\infty \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}(\sigma^2)^{\alpha-3/2}\exp\{-\beta\sigma^2\}\mathrm{d}\sigma^2\,, \quad (3.55)$$

which, similarly to the convolution integral (3.52), does not yield a closed-form expression.

Fortunately, it is far easier to perform the mixing in transform domain. By integrating the normal LT, $L(s) = \exp\{-\mu s + \sigma^2 s^2/2\}$, with respect to the distribution of σ^2, we

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

obtain

$$L(s|\mu,\alpha,\beta) = L(s) \;=\; \exp\{-\mu s\}\int_0^\infty \exp\{\sigma^2 s^2/2\}\frac{\beta^{\alpha}}{\Gamma(\alpha)}(\sigma^2)^{\alpha-1}\exp\{-\beta\sigma^2\}\mathrm{d}\sigma^2. \quad (3.56)$$

The remaining integral can be interpreted as the LT of the gamma density with transform variable $-s^2/2$, so that we can write

$$L(s) \;=\; \exp\{-\mu s\}\left(\frac{\beta}{\beta - s^2/2}\right)^{\alpha}, \qquad |s| < \sqrt{2\beta}\,. \quad (3.57)$$

By taking derivative w.r.t. s, we obtain the differential equation

$$(\beta - s^2/2)L'(s) + [\mu(\beta - s^2/2) - \alpha s]L(s) = 0\,, \quad (3.58)$$

which is more computationally useful for inference than the transform (3.57) itself, being highly non-linear as a function of the parameters.

The corresponding CF and its differential equation are obtained in a similar vein. We have

$$\phi_{\mu,\alpha,\beta}(s) \;=\; \phi(s) = \exp\{i\mu s\}\left(\frac{\beta}{\beta + s^2/2}\right)^{\alpha}, \quad (3.59)$$

and

$$(\beta + s^2/2)\phi'(s) - [i\mu(\beta + s^2/2) - \alpha s]\phi(s) = 0\,, \quad (3.60)$$

$s \in \mathbb{R}$. If we fix β, then μ and α can be computed explicitly in order to derive explicit estimators. This example will be continued in sec. 4.5.6.

3.5.3 Compound distributions

Compound distributions refer to summation of a random number of i.i.d. random variables:

$$Z = X_1 + X_2 + ... + X_K.$$

In such models, X_j's $\sim f_X(x|\boldsymbol{\theta})$ are typically positive and continuous whereas K is non-negative and discrete, with probability mass function $p(k)$. Compounding with K being Poisson, binomial or negative binomial is popular in insurance-related modeling to characterize claims aggregated over periods of time. Often, we observe only realizations of Z and, perhaps, frequencies of K, but have no access to sizes of particular X_j's. We need, however, to infer about the distribution of X given a sample of Z, $(Z_1, ..., Z_n)$.

Deriving the distribution of Z is virtually impossible. Its density is a (possibly infinite) weighted sum of multiple self-convolutions of the density of X, namely

$$f_Z(z|\boldsymbol{\theta}) = \sum_{k=0}^{\infty} p(k)[f_X(x|\boldsymbol{\theta})]^{k\circledast}\,, \quad (3.61)$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

whereby $[f_X(x|\boldsymbol{\theta})]^{0\circledast} \equiv 0$. Fortunately, formulation of the problem in terms of transforms is straightforward. By taking the LT of both sides of (3.61), we obtain

$$L_Z(s|\boldsymbol{\theta}) = \sum_{k=0}^{\infty} p(k)[L_X(s|\boldsymbol{\theta})]^k \,. \tag{3.62}$$

We recognize then that

$$L_Z(s|\boldsymbol{\theta}) = G(L_X(s|\boldsymbol{\theta}))\,, \tag{3.63}$$

where G denotes the probability generating function (PGF) of K. Depending on the structure and parametrization of G and L, we can use (3.63) in various ways - working with it directly or deriving the corresponding differential equations.

The latter approach brings the advantage of finding relations between the moments of X, K, and Z, and thus deriving the MM estimators defined in (1.3). For example, if $\boldsymbol{\theta} = (\theta_1, \theta_2)$, we compute the first and second derivatives of (3.63),

$$L'_Z(s|\boldsymbol{\theta}) = G'(L_X(s|\boldsymbol{\theta}))L'_X(s|\boldsymbol{\theta})\,, \tag{3.64}$$

and

$$L''_Z(s|\boldsymbol{\theta}) = G''(L_X(s|\boldsymbol{\theta}))[L'_X(s|\boldsymbol{\theta})]^2 + G'(L_X(s|\boldsymbol{\theta}))L''_X(s|\boldsymbol{\theta})\,, \tag{3.65}$$

and set $s = 0$. By the known transform properties, $L_X(0) = 1$, $L_X^{(\ell)}(0) = (-1)^\ell \mathbb{E}[X^\ell]$, $\ell = 1, 2, \ldots$, $G'(1) = \mathbb{E}[K]$, and $G''(1) = \mathbb{V}[K]$, we obtain the relations:

$$\mathbb{E}[Z] = \mathbb{E}[K]\mathbb{E}_{\boldsymbol{\theta}}[X], \tag{3.66}$$

$$\mathbb{E}[Z^2] = \mathbb{V}[K]\mathbb{E}^2_{\boldsymbol{\theta}}[X] + \mathbb{E}[K]\mathbb{E}_{\boldsymbol{\theta}}[X^2]\,. \tag{3.67}$$

Now, replacing the means and variances of Z and K by their sample counterparts, yields the following system:

$$\begin{cases} \overline{Z}_n = \overline{K}_n\, \mathbb{E}_{\boldsymbol{\theta}}[X]\,, \\ \widehat{\mathbb{V}}_n[Z] = \widehat{\mathbb{V}}_n[K](\mathbb{E}_{\boldsymbol{\theta}}[X])^2 + \overline{K}_n\, \mathbb{E}_{\boldsymbol{\theta}}[X] - (\overline{Z}_n)^2\,, \end{cases}$$

a solution to which for $\boldsymbol{\theta}$ yields the MM estimator, $\widehat{\boldsymbol{\theta}}_n^{\mathrm{MM}} = (\widehat{\theta}_{1,n}^{\mathrm{MM}}, \widehat{\theta}_{2,n}^{\mathrm{MM}})^\top$.

To perform the transform-based inference in this model, we first need to specify the (typically parametric) distribution of K. As mentioned above, a popular and analytically convenient choice is with $K \sim \mathrm{Poisson}(\lambda)$, so that

$$L_Z(s|\boldsymbol{\theta}, \lambda) = \exp\{-\lambda[1 - L_X(s|\boldsymbol{\theta})]\}. \tag{3.68}$$

The differential equation has then the general form

$$L'_Z(s|\boldsymbol{\theta}, \lambda) = \lambda L_Z(s|\boldsymbol{\theta}, \lambda)L'_X(s|\boldsymbol{\theta}). \tag{3.69}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Given a sample of Z, we now want to estimate or test the hypothesized distribution of X. While explicit estimation of $\boldsymbol{\theta}$ may not be feasible for many standard choices of X, the goodness-of-fit test statistic is computable explicitly quite often. For example, to test whether $X \sim \text{Exponential}(\beta)$, meaning that $L'_X(s|\boldsymbol{\theta}) = -\beta/(\beta+s)^2$, we can use the empirical version of the equation

$$(\beta + s)^2 L'_Z(s|\boldsymbol{\theta}, \lambda) + \lambda\beta L_Z(s|\boldsymbol{\theta}, \lambda) = 0, \tag{3.70}$$

in which λ, β and the transforms are replaced by their consistent estimators, $\widehat{\lambda}_n$, $\widehat{\beta}_n$, $\widehat{L}_n(s)$ and $\widehat{L}'_n(s)$. Specifically, the resulting weighted $\mathcal{L}^2$ test statistic of the type (1.16), based on $(Z_1, ..., Z_n)$, will be

$$T_n = n\int_0^\infty \left[(\overline{K}_n/\overline{Z}_n + s)^2 \widehat{L}'_n(s) + \overline{K}_n^2/\overline{Z}_n \widehat{L}_n(s)\right]^2 \mathrm{d}W(s)\,, \tag{3.71}$$

with $\widehat{L}_n(s) = \frac{1}{n}\sum_{j=1}^n e^{-sZ_j}$ and $\widehat{L}'_n(s) = -\frac{1}{n}\sum_{j=1}^n Z_j e^{-sZ_j}$, $\overline{K}_n = \widehat{\lambda}_n$ and $\overline{K}_n/\overline{Z}_n = \widehat{\beta}_n$.

The reader also is referred to the recent work by Goffard et al. (2022), who derive more complicated but computationally feasible test statistics based on differential equations like (3.64). In their more general assumptions, K belongs to the entire *Katz family*[2], while X follows the gamma or inverse Gaussian distribution.

[2]The Katz family is characterized by the fact that its PGF $G(s)$ satisfies $(1-\beta s)G'(s) - \alpha G(s) = 0$, $\alpha > 0$, $\beta < 1$, subject to $G(1) = 1$. The family encompasses the binomial, negative binomial, and Poisson distributions as special cases; see Johnson et al. (1993), Fang (2003) and sec. 6.1.3.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

CHAPTER 4

Estimators based on differential equations

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

As already explained, the parameter estimation in our approach is done by replacing the model transforms, namely $L_{\boldsymbol{\theta}}(s)$ or $\phi_{\boldsymbol{\theta}}(s)$, and their derivatives present in a differential equation by the empirical counterparts, namely $\widehat{L}_n(s), \widehat{L}_n'(s), \ldots$ or $\widehat{\phi}_n(s), \widehat{\phi}_n'(s), \ldots$ and so on. The estimator of $\boldsymbol{\theta}$ is then defined as the minimizer of certain norm of this empirical version of the equation over the parameter space. In this thesis, we exclusively focus on estimators defined by weighted $\mathcal{L}^2$ norms. Other choices, such as (weighted) $\mathcal{L}^q$ or infimum norms, are possible as well. However, opting for these alternatives would not lead to explicit estimators, and exploring their properties, which is already complicated in the $\mathcal{L}^2$ setting, would become intractable.

4.1 Transform-based $\mathcal{L}^2$ estimators

Given the variety of forms of the differential equations derived in the previous chapter, we confine our attention to the following estimators. Among the LT-based estimators, these will be

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m} = \operatorname*{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int_0^\infty \left[h(\boldsymbol{\theta},s)\widehat{L}_n^{(r)}(s) + \widehat{L}_n^{(m)}(s)\right]^2 \mathrm{d}W(s)\,, \tag{4.1}$$

and

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m,\ell} = \operatorname*{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int_0^\infty \left[h(\boldsymbol{\theta},s)\widehat{L}_n^{(r)}(s) + g(\boldsymbol{\theta},s)\widehat{L}_n^{(m)}(s) + \widehat{L}_n^{(\ell)}(s)\right]^2 \mathrm{d}W(s)\,, \tag{4.2}$$

which are suitable for one-sided (life-time) distributions.

Among the CF-based estimators, we will analyze the estimators of the form

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,r,m} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \int_{-\infty}^{\infty} \left| h(\boldsymbol{\theta}, s)\widehat{\phi}_n^{(r)}(s) + \widehat{\phi}_n^{(m)}(s) \right|^2 \mathrm{d}W(s)\,, \tag{4.3}$$

which are suitable for both one- and two-sided distributions.

In the above definitions, r, m and ℓ denote the orders of the derivatives. The coefficients $h(\boldsymbol{\theta}, s)$ and $g(\boldsymbol{\theta}, s)$ are real or complex functions, found by appropriate rewriting of the particular equation in use. With (4.1)-(4.3), we can provide estimators for nearly all distributions listed in Tables 3.1 and 3.2. Due to space limitations, we neither consider the estimator of the type (4.2) based on CF nor the LT- and CF-based estimators employing non-homogeneous equations (as in the uniform and Pareto examples). Nonetheless, the theory presented below easily extends to those cases.

4.1.1 Choice of the weight function

The role of the user-defined weight function $W(s)$ in (4.1)-(4.3) is threefold:

(i) to render the integrals finite;

(ii) to control the properties of the estimators;

(iii) to provide closed-form expressions for the $\mathcal{L}^2$ measure (and eventually for the estimators).

The following sections, 4.2 and 4.3, present theorems on asymptotic normality and influence functions of (4.1)-(4.3). For generality, these theorems do not specify the form of $W(s)$, assuming only integrability of the weighted expression. In further sections, however, we will apply the estimators to concrete distributions using specific differentiable weight functions, $\mathrm{d}W(s) = w(s)\mathrm{d}s$. In particular, for the LT-based estimators (4.1) and (4.2), we will choose $W(s)$ such that

$$w(s) = s^{\gamma} \exp(-as)\,, \qquad s > 0,\ a > 0,\ \gamma = 0, 1, 2, \dots\,. \tag{4.4}$$

For the CF-based estimator (4.3) in turn, it will be

$$w(s) = s^{\gamma} \exp(-as^2)\,, \qquad s \in \mathbb{R},\ a > 0, \gamma = 0, 2, 4, \dots\,, \tag{4.5}$$

which is always an even function.

Both choices will fulfill the requirements (i) and (iii) in the distributional cases we aim to consider. Additionally, both are also suitable concerning (ii), as they can *simultaneously* control efficiency and robustness to outliers. Note that if our sole objective were maximum efficiency, then pure exponential and Gaussian functions, namely $\exp(-as)$ and

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

$\exp(-as^2)$, would suffice. In fact, these functions are sufficient for the LT- and CF-based GoF tests employing the 'traditional' $\mathcal{L}^2$ distances such as

$$T_n \;=\; n\int_0^\infty \Big[L(s|\widehat{\boldsymbol{\theta}}_n) - \widehat{L}_n(s)\Big]^2 \mathrm{d}W(s)\,, \tag{4.6}$$

and

$$T_n \;=\; n\int_{-\infty}^\infty \Big|\phi(s|\widehat{\boldsymbol{\theta}}_n) - \widehat{\phi}_n(s)\Big|^2 \mathrm{d}W(s)\,, \tag{4.7}$$

used by many authors, e.g., Henze and Klar (2002), Epps (2005), Henze et al. (2012) or Henze and Koch (2020). In these tests, the analyst's goal is to upweight the most informative region of the transform domain, the neighborhood of $s = 0$. However, for conducting robust estimation of parameters using the derivatives of transforms, we need the opposite action - the ability to downweight the vicinity of $s = 0$, where the empirical derivatives have the highest absolute IFs and tend to be unrobust; recall the findings in sec. 2.4.2 and refer to Fig. 2.5. Therefore, we need weight functions such that $\lim_{s\to 0} w(s) = 0$, and the user-controlled factor s^γ in (4.4) and (4.5) serves this purpose.

4.1.2 Estimators with stepwise weight function

Before delving into the general theory, let us discuss the important case when $W(s)$ is a stepwise function, mapping $\mathbb{R}$ or $\mathbb{R}_+$ to [0,1]. As mentioned in the introduction, such a weight function is generally not favored from a computational perspective because it does not lead to explicit estimators. However, with $k=p=\dim(\boldsymbol{\theta})$ and equal-size steps of $W(s)$, the resulting estimator can be an explicit solver of a system of p equations. For example, for the estimator (4.1) with such a stepwise $W(s)$ the system is

$$\begin{cases} h(\boldsymbol{\theta}, s_1)\widehat{L}_n^{(r)}(s_1) + \widehat{L}_n^{(m)}(s_1) = 0\,, \\ \vdots \\ h(\boldsymbol{\theta}, s_p)\widehat{L}_n^{(r)}(s_p) + \widehat{L}_n^{(m)}(s_p) = 0\,, \end{cases} \tag{4.8}$$

and has an explicit solution if $h(\boldsymbol{\theta}, s)$ is linear in elements of $\boldsymbol{\theta}$ (or in invertible functions thereof), expressed in terms of $\widehat{L}_n^{(r)}(s_j)$ and $\widehat{L}_n^{(m)}(s_j)$, $j = 1, ..., p$.

We denote such an estimator as $\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}(s_1, ..., s_p)$ and refer to it as the *transform-matching estimator*. Its properties, like variance and robustness, will depend on the choice of $s_1, ..., s_p$. To simplify this choice and control the properties through a single tuning variable s, a *limit form* can be derived,

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}(s) \;=\; \lim_{s_1\to,...,\to s_p\to s} \widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}(s_1, ..., s_p)\,. \tag{4.9}$$

However, the routine application of d'Hospital's rule, necessary to compute the limit, typically results in an increase in the orders of transform derivatives in the final estimator formulas (here m and r).

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The same considerations hold in the CF setting, whereby the estimating equations are obtained after decomposition of CF into the real and imaginary part and/or modulus.

The transform-matching estimators are special cases of (4.1)-(4.3) and, clearly, do not require separate asymptotic and robustness theory. Either way, when estimators of this kind are explicit, their asymptotic normal distribution can be found by the standard Delta Method, while their IFs are derivable by the chain rule (2.48). Several examples will be studied in secs. 4.5.2, 4.5.3 and 4.5.5, for the exponential(λ), Rayleigh(ϑ) and Pareto(α,β) distributions (that is with parameter dimension $p = 1$ or 2).

4.2 Asymptotic normality of the estimators

In this section, we establish the asymptotic normality of the $\mathcal{L}^2$ estimators (4.1), (4.2) and (4.3). This is covered by the Theorems 1, 2 and 3, respectively. The proof of Theorem 2 is similar to that of Theorem 1 and can be omitted. The proofs exploit the theory of V- and U-statistics summarized in Appendix A to which we refer at relevant points.

The proofs make use of the following general assumptions:

I. The random sample $\{X_1, ..., X_n\}$ is i.i.d.

II. The underlying differential equations

$$h(\boldsymbol{\theta}, s)L_{\boldsymbol{\theta}}^{(r)}(s) + L_{\boldsymbol{\theta}}^{(m)}(s) = 0\,,$$
$$h(\boldsymbol{\theta}, s)L_{\boldsymbol{\theta}}^{(r)}(s) + g(\boldsymbol{\theta}, s)L_{\boldsymbol{\theta}}^{(m)}(s) + L_{\boldsymbol{\theta}}^{(\ell)}(s) = 0,$$
$$h(\boldsymbol{\theta}, s)\phi_{\boldsymbol{\theta}}^{(r)}(s) + \phi_{\boldsymbol{\theta}}^{(m)}(s) = 0,$$

are uniquely solved by $L_{\boldsymbol{\theta}}(s)$ or $\phi_{\boldsymbol{\theta}}(s)$ in the sense of sec. 3.4.2.

III. The functions $h(\boldsymbol{\theta}, s)$ and $g(\boldsymbol{\theta}, s)$ and their derivatives with respect to $\boldsymbol{\theta}$ up to the second order are square integrable w.r.t. $W(s)$ for all $\boldsymbol{\theta}$.

IV. The moment $\mathbb{E}[X^{2k}]$ where $k = \max(r, m)$ or $k = \max(r, m, \ell)$ is finite.

Remark. For the Laplace transform-based estimators this moment condition may be relaxed in special cases. Namely, if $W(s)$ and $h(\boldsymbol{\theta}, s)$ and $g(\boldsymbol{\theta}, s)$ avoid 0 like this: assume $\mathrm{d}W(s) = w(s)\mathrm{d}s$, and, for $s \to 0$

$$w(s) = O(|s|^{\gamma}),$$

$$h(\boldsymbol{\theta}, s), \frac{\partial}{\partial\theta_i}h(\boldsymbol{\theta}, s), \frac{\partial^2}{\partial\theta_i\partial\theta_j}h(\boldsymbol{\theta}, s) = O(|s|^{\delta}),$$

$$g(\boldsymbol{\theta}, s), \frac{\partial}{\partial\theta_i}g(\boldsymbol{\theta}, s), \frac{\partial^2}{\partial\theta_i\partial\theta_j}g(\boldsymbol{\theta}, s) = O(|s|^{\eta}),$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

$i, j = 1, ..., p$, with $\gamma, \delta, \eta \geq 0$. Then we only need that $\mathbb{E}(|X|^k)$ is finite, where k is given by $k = \max(2r - 2\delta - \gamma - 1, 2m - \gamma - 1, 0)$, resp. $k = \max(2r - 2\delta - \gamma - 1, 2m - 2\eta - \gamma - 1, 2\ell - \gamma - 1, 0)$.

Therefore, for $h(\boldsymbol{\theta}, s)$ and $g(\boldsymbol{\theta}, s)$ being polynomials in s (which is the case for most of the LT-based differential equations we use), and for the special form of the weight we proposed, namely $w(s) = s^\gamma \exp(-as)$, $\gamma = 0, 1, ...$, the assumption may be relaxed.

Theorem 1. *Let $\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}$ be the weighted $\mathcal{L}^2$ estimator of the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ of a one-sided distribution, based on linear ODE involving the r-th and m-th derivative of its LT ($r, m \geq 0$, $r \neq m$):*

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \int_0^\infty \left[h(\boldsymbol{\theta}, s)\widehat{L}_n^{(r)}(s) + \widehat{L}_n^{(m)}(s)\right]^2 \mathrm{d}W(s)\,. \tag{4.10}$$

Under the assumptions I-IV, the estimator is strongly consistent and converges in distribution to a p-variate normal random vector, that is

$$\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m} - \boldsymbol{\theta}\Big) \xrightarrow{\mathrm{d}} \mathcal{N}_p\Big(\mathbf{0}, \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\Big)\,, \tag{4.11}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the $p \times p$ symmetric matrix whose (i,j)-th element is

$$\begin{aligned}\sigma_{i,j}^2(\boldsymbol{\theta}) = 4\int_0^\infty\!\!\int_0^\infty \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}\frac{\partial h(\boldsymbol{\theta}, t)}{\partial\theta_j}\Big\{&\Big[2h(\boldsymbol{\theta}, s)L_{\boldsymbol{\theta}}^{(r)}(s) + L_{\boldsymbol{\theta}}^{(m)}(s)\Big]\Big[2h(\boldsymbol{\theta}, t)L_{\boldsymbol{\theta}}^{(r)}(t) + L_{\boldsymbol{\theta}}^{(m)}(t)\Big]L_{\boldsymbol{\theta}}^{(2r)}(s+t)\\ &+ \Big[2h(\boldsymbol{\theta}, s)L_{\boldsymbol{\theta}}^{(r)}(s) + L_{\boldsymbol{\theta}}^{(m)}(s)\Big]L_{\boldsymbol{\theta}}^{(r)}(t)L_{\boldsymbol{\theta}}^{(r+m)}(s+t)\\ &+ \Big[2h(\boldsymbol{\theta}, t)L_{\boldsymbol{\theta}}^{(r)}(t) + L_{\boldsymbol{\theta}}^{(m)}(t)\Big]L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(r+m)}(s+t)\\ &+ L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(r)}(t)L_{\boldsymbol{\theta}}^{(r+m)}(s+t)\Big\}\mathrm{d}W(s)\mathrm{d}W(t)\,,\end{aligned} \tag{4.12}$$

and $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is the $p \times p$ symmetric matrix whose (i,j)-th element is expressed by

$$\omega_{i,j}(\boldsymbol{\theta}) = 2\int_0^\infty\Big\{\frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i\partial\theta_j}\Big[h(\boldsymbol{\theta}, s)[L_{\boldsymbol{\theta}}^{(r)}(s)]^2 + L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)\Big] + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}\frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_j}[L_{\boldsymbol{\theta}}^{(r)}(s)]^2\Big\}\mathrm{d}W(s)\,. \tag{4.13}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Proof. First, note that the estimator $\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}$ can be written as the p-dimensional minimizer of a degree-2 V-statistic, namely

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m} = \operatorname*{argmin}_{\boldsymbol{\theta}} \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\psi_{\boldsymbol{\theta}}(X_j, X_k),$$

whose kernel is given by

$$\psi_{\boldsymbol{\theta}}(X_j, X_k) = \int_0^\infty \Big(h(\boldsymbol{\theta}, s)(\text{-}X_j)^r e^{-sX_j} + (\text{-}X_j)^m e^{-sX_j}\Big)\Big(h(\boldsymbol{\theta}, s)(\text{-}X_k)^r e^{-sX_k} + (\text{-}X_k)^m e^{-sX_k}\Big)\mathrm{d}W(s).$$

Let us denote this V-statistic by

$$J_n(\boldsymbol{\theta}) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\psi_{\boldsymbol{\theta}}(X_j, X_k),$$

and by $J(\boldsymbol{\theta})$ denote the underlying expectation functional, that is $J(\boldsymbol{\theta}) = \mathbb{E}_F[\psi_{\boldsymbol{\theta}}(X_1, X_2)]$. When differentiation under integration sign is permitted, consider the following gradient

$$\nabla J_n(\boldsymbol{\theta}) = \left(\frac{\partial J_n(\boldsymbol{\theta})}{\partial\theta_1}, \ldots, \frac{\partial J_n(\boldsymbol{\theta})}{\partial\theta_p}\right)^\top,$$

with i-th element

$$\begin{aligned}
\frac{\partial J_n(\boldsymbol{\theta})}{\partial\theta_i} &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial\psi_{\boldsymbol{\theta}}(X_j, X_k)}{\partial\theta_i} \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\int_0^\infty\Big\{2\frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}h(\boldsymbol{\theta}, s)(X_jX_k)^r e^{-s(X_j+X_k)} \qquad (4.14)\\
&+ \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}(\text{-}X_j)^r(\text{-}X_k)^m e^{-s(X_j+X_k)} + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}(\text{-}X_j)^m(\text{-}X_k)^r e^{-s(X_j+X_k)}\Big\}\mathrm{d}W(s).
\end{aligned}$$

Note that each element of $\nabla J_n(\boldsymbol{\theta})$ is also a V-statistic because $\frac{\partial\psi_{\boldsymbol{\theta}}(X_j,X_k)}{\partial\theta_i}$ is symmetric in X_j and X_k. The value of the estimator is the solver of the set of p equations

$$\nabla J_n(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}) = \mathbf{0}. \qquad (4.15)$$

Consider also the Hessian of $J_n(\boldsymbol{\theta})$, that is

$$\nabla^2 J_n(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2 J_n(\boldsymbol{\theta})}{\partial\theta_1^2} & \cdots & \frac{\partial^2 J_n(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J_n(\boldsymbol{\theta})}{\partial\theta_p\partial\theta_1} & \cdots & \frac{\partial^2 J_n(\boldsymbol{\theta})}{\partial\theta_p^2} \end{pmatrix}.$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The asymptotic normality of the estimator is shown by expanding $\nabla J_n(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\, r,m})$ into the first order multivariate Taylor expansion about the true value of the vector parameter $\boldsymbol{\theta}$, namely

$$\nabla J_n(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\, r,m}) \;=\; \nabla J_n(\boldsymbol{\theta}) + \nabla^2 J_n(\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\, r,m} - \boldsymbol{\theta}\,) + \dots\, .$$

Because of (4.15), the right side vanishes and we can rewrite the expansion as

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\, r,m} - \boldsymbol{\theta}\,) \;\approx\; \sqrt{n}\Big[\nabla^2 J_n(\boldsymbol{\theta})\Big]^{-1} \nabla J_n(\boldsymbol{\theta}) + \dots\, . \tag{4.16}$$

We first need to establish the asymptotic normality of $\nabla J_n(\boldsymbol{\theta})$. We know that each element of $\nabla J_n(\boldsymbol{\theta})$ is a degree-2 V-statistic with finite second moment, and, as explained in Appendix A, each of them alone can be proved to be asymptotically normal based on the asymptotic normality of the corresponding U-statistic; see (A.9) in Appendix A. The joint asymptotic normality follows by applying the Crámer-Wold theorem; see e.g. Staudte and Sheather (1990) p. 292. The entries of the asymptotic covariance matrix equal the asymptotic covariances of the corresponding U-statistics, see (A.14) in Appendix A. That is, we have

$$\sqrt{n}(\nabla J_n(\boldsymbol{\theta}) - \nabla J(\boldsymbol{\theta})) \xrightarrow{\mathrm{d}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))\,, \tag{4.17}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the $p \times p$ symmetric covariance matrix whose (i,j)-th element equals

$$\sigma^2_{i,j}(\boldsymbol{\theta}) = 4\mathbb{C}\mathrm{ov}\Bigg[\,\mathbb{E}\Bigg[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_i}\Bigg]_{x_1 = X_1}, \mathbb{E}\Bigg[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_j}\Bigg]_{x_1 = X_1}\Bigg]\,. \tag{4.18}$$

Given (4.17), to show the multivariate normality of the product $[\nabla^2 J_n(\boldsymbol{\theta})]^{-1} \nabla J_n(\boldsymbol{\theta})$ in (4.16), we apply the multivariate version of the Slutsky theorem (see e.g. Hunter (2014), sec. 2.3.3). This gives

$$\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\, r,m} - \boldsymbol{\theta}\Big) \;\xrightarrow{\mathrm{d}}\; \mathcal{N}_p\Big(\mathbf{0}, \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\Big)\,, \tag{4.19}$$

where the (i,j)-th element of the $p \times p$ matrix $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is expressed by

$$\omega_{i,j}(\boldsymbol{\theta}) = \mathbb{E}\Bigg[\frac{\partial^2 \psi_{\boldsymbol{\theta}}(X_1, X_2)}{\partial \theta_i \partial \theta_j}\Bigg],$$

and proves the main statement of the theorem.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

What remains is to derive the entries of the matrices $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}(\boldsymbol{\theta})$. First, consider the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Note that, for all $i, j = 1, ..., p$,

$$\mathbb{E}\Big[\mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_i}\Big]_{x_1=X_1}\Big] \;=\; \mathbb{E}\Big[\mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_j}\Big]_{x_1=X_1}\Big] = 0\,.$$

Therefore, the computation of the covariance term (4.18) boils down to finding the cross expectation

$$\sigma^2_{i,j}(\boldsymbol{\theta}) = 4\,\mathbb{E}\Big[\;\mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_i}\Big]_{x_1=X_1} \cdot \mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_j}\Big]_{x_1=X_1}\;\Big].$$

For this purpose, let us compute

$$\begin{aligned}
\mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_i}\Big]_{x_1=X_1} &= \Big[\int_0^\infty \frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_i}\mathrm{d}F(x)\Big]_{x_1=X_1} \\
&= \int_0^\infty \Big\{ \int_0^\infty \Big\{ 2\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i} h(\boldsymbol{\theta}, s)(x_1 X_2)^r e^{-s(x_1+X_2)} \\
&\qquad + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}(\text{-}x_1)^r(\text{-}X_2)^m e^{-s(x_1+X_2)} \\
&\qquad + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}(\text{-}x_1)^m(\text{-}X_2)^r e^{-s(x_1+X_2)}\Big\}\mathrm{d}F(x)\Big\}\mathrm{d}W(s)\Big|_{x_1=X_1}.
\end{aligned}$$

Recognizing that $\int_0^\infty (\text{-}x)^r e^{-sx}\mathrm{d}F(x) = L^{(r)}_{\boldsymbol{\theta}}(s)$ and rearranging, we obtain

$$\mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial \theta_i}\Big]_{x_1=X_1} = \int_0^\infty \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}\Big\{\Big[2h(\boldsymbol{\theta}, s)L^{(r)}_{\boldsymbol{\theta}}(s) + L^{(m)}_{\boldsymbol{\theta}}(s)\Big](\text{-}X_1)^r e^{-sX_1} + L^{(r)}_{\boldsymbol{\theta}}(s)(\text{-}X_1)^m e^{-sX_1}\Big\}\mathrm{d}W(s).$$

With this result, the (i, j)-th element of the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, as defined by (4.18), is the product of two integrals which can be expressed as the following double integral:

$$\begin{aligned}
\sigma^2_{i,j}(\boldsymbol{\theta}) = 4\mathbb{E}\Big[\int_0^\infty\!\!\int_0^\infty \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}\Big\{\Big[2h(\boldsymbol{\theta}, s)L^{(r)}_{\boldsymbol{\theta}}(s) + L^{(m)}_{\boldsymbol{\theta}}(s)\Big](\text{-}X_1)^r e^{-sX_1} + L^{(r)}_{\boldsymbol{\theta}}(s)(\text{-}X_1)^m e^{-sX_1}\Big\} \\
\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}\Big\{\Big[2h(\boldsymbol{\theta}, t)L^{(r)}_{\boldsymbol{\theta}}(t) + L^{(m)}_{\boldsymbol{\theta}}(t)\Big](\text{-}X_1)^r e^{-tX_1} + L^{(r)}_{\boldsymbol{\theta}}(t)(\text{-}X_1)^m e^{-tX_1}\Big\}\mathrm{d}W(s)\mathrm{d}W(t)\Big].
\end{aligned}$$

Taking the expectation in the way shown just above yields (4.12) in the theorem.

Computation of the matrix $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is easier due to independence of X_1 and X_2 in the kernel of the U-statistic. Its (i,j)-element is

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

$$\omega_{i,j}(\boldsymbol{\theta}) = \mathbb{E}\Big[\frac{\partial^2 \psi_{\boldsymbol{\theta}}(X_1, X_2)}{\partial\theta_i \partial\theta_j}\Big] = \int_0^\infty \frac{\partial^2 \psi_{\boldsymbol{\theta}}(X_1, X_2)}{\partial\theta_i \partial\theta_j}\,\mathrm{d}F(x)$$

$$= \int_0^\infty \Big\{ \int_0^\infty \Big\{ 2\Big[\frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i\theta_j} h(\boldsymbol{\theta}, s) + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}\frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_j}\Big](X_1X_2)^r e^{-s(X_1+X_2)}$$

$$+ \frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i\partial\theta_j}(\text{-}X_1)^r(\text{-}X_2)^m e^{-s(X_1+X_2)} + \frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i\partial\theta_j}(\text{-}X_1)^m(\text{-}X_2)^r e^{-s(X_1+X_2)}\Big\}\mathrm{d}F(x)\Big\}\mathrm{d}W(s)$$

$$= 2\int_0^\infty \Big\{ \frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i\partial\theta_j}\Big[h(\boldsymbol{\theta}, s)[L_{\boldsymbol{\theta}}^{(r)}(s)]^2 + L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)\Big] + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}\frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_j}[L_{\boldsymbol{\theta}}^{(r)}(s)]^2\Big\}\mathrm{d}W(s),$$

which is (4.13) of the theorem. □

Theorem 2. *Let $\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m,\ell}$ be the weighted $\mathcal{L}^2$ estimator of the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ of a one-sided distribution, based on ODE involving the r-th, m-th and ℓ-th derivative of LT $(r, m, \ell \geq 0, \quad r \neq m \neq \ell)$:*

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m,\ell} = \operatorname*{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int_0^\infty \Big[h(\boldsymbol{\theta}, s)\widehat{L}_n^{(r)}(s) + g(\boldsymbol{\theta}, s)\widehat{L}_n^{(m)}(s) + \widehat{L}_n^{(\ell)}(s)\Big]^2 \mathrm{d}W(s)\,. \quad (4.20)$$

Under the assumptions I-IV, the estimator converges in distribution to a p-variate normal random vector, that is

$$\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m,\ell} - \boldsymbol{\theta}\Big) \xrightarrow{\mathrm{d}} \mathcal{N}_p\Big(\mathbf{0}, \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\Big)\,, \quad (4.21)$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the $p \times p$ symmetric matrix whose (i,j)-th element is

$$\sigma_{i,j}^2(\boldsymbol{\theta}) = 4\int_0^\infty\int_0^\infty \Big\{A_i(s)A_j(t)L_{\boldsymbol{\theta}}^{(2r)}(s+t) + B_i(s)B_j(t)L_{\boldsymbol{\theta}}^{(2m)}(s+t) + C_i(s)C_j(t)L_{\boldsymbol{\theta}}^{(2\ell)}(s+t)$$

$$+\Big[A_i(s)B_j(t)+B_i(s)A_j(t)\Big]L_{\boldsymbol{\theta}}^{(r+m)}(s+t) + \Big[A_i(s)C_j(t)+C_i(s)A_j(t)\Big]L_{\boldsymbol{\theta}}^{(r+\ell)}(s+t)$$

$$+\Big[B_i(s)C_j(t)+C_i(s)B_j(t)\Big]L_{\boldsymbol{\theta}}^{(m+\ell)}(s+t)\Big\}\mathrm{d}W(s)\mathrm{d}W(t)\,, \quad (4.22)$$

with

$$A_i(s) = \frac{\partial h^2(\boldsymbol{\theta}, s)}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(r)}(s) + \frac{\partial\{h(\boldsymbol{\theta}, s)g(\boldsymbol{\theta}, s)\}}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(m)}(s) + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(\ell)}(s),$$

$$B_i(s) = \frac{\partial g^2(\boldsymbol{\theta}, s)}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(m)}(s) + \frac{\partial\{h(\boldsymbol{\theta}, s)g(\boldsymbol{\theta}, s)\}}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(r)}(s) + \frac{\partial g(\boldsymbol{\theta}, s)}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(\ell)}(s),$$

$$C_i(s) = \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(r)}(s) + \frac{\partial g(\boldsymbol{\theta}, s)}{\partial\theta_i}L_{\boldsymbol{\theta}}^{(m)}(s)\,,$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

and $\mathbf{\Omega}(\boldsymbol{\theta})$ is the $p \times p$ symmetric matrix whose (i,j)-th element is expressed by

$$\omega_{i,j}(\boldsymbol{\theta}) = \int_0^{\infty}\Big\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\Big[h(\boldsymbol{\theta},s)L_{\boldsymbol{\theta}}^{(r)}(s)+g(\boldsymbol{\theta},s)L_{\boldsymbol{\theta}}^{(m)}(s)\Big]^2 + \\ + 2\Big(\frac{\partial^2 h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(\ell)}(s)+\frac{\partial^2 g(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}L_{\boldsymbol{\theta}}^{(m)}(s)L_{\boldsymbol{\theta}}^{(\ell)}(s)\Big)\Big\}\mathrm{d}W(s) \tag{4.23}$$

Proof. The proof is similar that of Theorem 1 but involves more complicated algebra. □

Theorem 3. *Let $\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,r,m}$ be the weighted $\mathcal{L}^2$ estimator of the parameter $\boldsymbol{\theta} \in \mathbf{\Theta} \subseteq \mathbb{R}^p$, based on differential equation involving the r-th and m-th ($r,m \geq 0$, $r \neq m$) derivative of CF $\phi_{\boldsymbol{\theta}}(s)$ of a distribution on $\mathbb{R}$, $\mathbb{R}_+$ or $\mathbb{R}_+^0$:*

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,r,m} = \underset{\boldsymbol{\theta}\in\mathbf{\Theta}}{\operatorname{argmin}} \int_{-\infty}^{\infty}\Big|h(\boldsymbol{\theta},s)\widehat{\phi}_n^{(r)}(s)+\widehat{\phi}_n^{(m)}(s)\Big|^2\mathrm{d}W(s)\,. \tag{4.24}$$

Under the assumptions I-IV, the estimator converges in distribution to a p-variate normal random vector, that is

$$\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,r,m}-\boldsymbol{\theta}\Big) \xrightarrow{\mathrm{d}} \mathcal{N}_p\Big(\mathbf{0},4\mathbf{\Omega}^{-1}(\boldsymbol{\theta})\mathbf{\Sigma}(\boldsymbol{\theta})\mathbf{\Omega}^{-1}(\boldsymbol{\theta})\Big)\,, \tag{4.25}$$

where $\mathbf{\Sigma}(\boldsymbol{\theta})$ is the $p \times p$ symmetric covariance matrix of the random vector

$$\sigma_j(\boldsymbol{\theta},X) = \int_{-\infty}^{\infty}\Big\{\Big(\frac{\partial|h(\boldsymbol{\theta},s)|^2}{\partial\theta_j}u_{\boldsymbol{\theta}}^{(r)}(s)+\frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}u_{\boldsymbol{\theta}}^{(m)}(s)+\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}v_{\boldsymbol{\theta}}^{(m)}(s)\Big)X^r\cos(sX+\tfrac{r\pi}{2}) \\ +\Big(\frac{\partial|h(\boldsymbol{\theta},s)|^2}{\partial\theta_j}v_{\boldsymbol{\theta}}^{(r)}(s)+\frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}v_{\boldsymbol{\theta}}^{(m)}(s)-\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}u_{\boldsymbol{\theta}}^{(m)}(s)\Big)X^r\sin(sX+\tfrac{r\pi}{2}) \\ +\Big(\frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}u_{\boldsymbol{\theta}}^{(r)}(s)-\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}v_{\boldsymbol{\theta}}^{(r)}(s)\Big)X^m\cos(sX+\tfrac{m\pi}{2}) \\ +\Big(\frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}v_{\boldsymbol{\theta}}^{(r)}(s)+\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}u_{\boldsymbol{\theta}}^{(r)}(s)\Big)X^m\sin(sX+\tfrac{m\pi}{2})\Big\}\mathrm{d}W(s),\ j=1,...,p\,, \tag{4.26}$$

and $\mathbf{\Omega}(\boldsymbol{\theta})$ is the $p \times p$ symmetric matrix whose (i,j)-th element is expressed by

$$\omega_{i,j}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty}\Big\{\frac{\partial^2|h(\boldsymbol{\theta},s)|^2}{\partial\theta_i\partial\theta_j}|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2+2\frac{\partial^2\mathrm{Re}h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}\Big[u_{\boldsymbol{\theta}}^{(r)}(s)u_{\boldsymbol{\theta}}^{(m)}(s)-v_{\boldsymbol{\theta}}^{(r)}(s)v_{\boldsymbol{\theta}}^{(m)}(s)\Big] \\ +2\frac{\partial^2\mathrm{Im}h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}\Big[u_{\boldsymbol{\theta}}^{(r)}(s)v_{\boldsymbol{\theta}}^{(m)}(s)-u_{\boldsymbol{\theta}}^{(m)}(s)v_{\boldsymbol{\theta}}^{(r)}(s)\Big]\Big\}\mathrm{d}W(s). \tag{4.27}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Proof. The proof utilizes the same idea as used for the LT-based estimators. Namely, we take the first order multivariate Taylor expansion of the gradient of the V-statistic $J_n(\boldsymbol{\theta})$, associated with the estimator, about the true value of the parameter. This yields

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\, r,m} - \boldsymbol{\theta}) \;\approx\; \sqrt{n}\Big[\nabla^2 J_n(\boldsymbol{\theta})\Big]^{-1} \nabla J_n(\boldsymbol{\theta}) + \dots ,$$

where $\nabla J_n(\boldsymbol{\theta})$ is the gradient and $\nabla^2 J_n(\boldsymbol{\theta})$ is the Hessian of that V-statistic.

In analogy to the proof of Theorem 1, the multivariate Slutsky theorem, given the asymptotic multivariate normality of $\nabla J_n(\boldsymbol{\theta})$, asserts that

$$\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\, r,m} - \boldsymbol{\theta}\Big) \;\xrightarrow{\mathrm{d}}\; \mathcal{N}_p\Big(\mathbf{0}, 4\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\Big),$$

where both $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}(\boldsymbol{\theta})$ are $p \times p$ matrices to be determined.

For this purpose, we write the estimator as

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\, r,m} &= \operatorname*{argmin}_{\boldsymbol{\theta}} \int_{-\infty}^{\infty} \Big\{ \Big(\widehat{u}_n^{(r)}(s)\mathrm{Re}\, h(\boldsymbol{\theta},s) - \widehat{v}_n^{(r)}(s)\mathrm{Im}\, h(\boldsymbol{\theta},s) + \widehat{u}_n^{(m)}(s)\Big)^2 \\
&\qquad + \Big(\widehat{v}_n^{(r)}(s)\mathrm{Re}\, h(\boldsymbol{\theta},s) + \widehat{u}_n^{(r)}(s)\mathrm{Im}\, h(\boldsymbol{\theta},s) + \widehat{v}_n^{(m)}(s)\Big)^2 \Big\} \mathrm{d}W(s) \\
&= \operatorname*{argmin}_{\boldsymbol{\theta}} J_n(\boldsymbol{\theta}) \\
&= \operatorname*{argmin}_{\boldsymbol{\theta}} \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \psi_{\boldsymbol{\theta}}(X_j, X_k).
\end{aligned}$$

After multiplications and cancellations in the integrand, the kernel of the associated V-statistic becomes

$$\begin{aligned}
&\psi_{\boldsymbol{\theta}}(X_j, X_k) = \int_{-\infty}^{\infty} \Big\{ \\
&X_j^r X_k^r \cos(sX_j + \tfrac{r\pi}{2}) \cos(sX_k + \tfrac{r\pi}{2}) \mathrm{Re}^2 h(\boldsymbol{\theta},s) \;+\; X_j^r X_k^m \cos(sX_j + \tfrac{r\pi}{2}) \cos(sX_k + \tfrac{m\pi}{2}) \mathrm{Re} h(\boldsymbol{\theta},s) \\
&X_j^r X_k^r \sin(sX_j + \tfrac{r\pi}{2}) \sin(sX_k + \tfrac{r\pi}{2}) \mathrm{Im}^2 h(\boldsymbol{\theta},s) \;-\; X_j^r X_k^m \sin(sX_j + \tfrac{r\pi}{2}) \cos(sX_k + \tfrac{m\pi}{2}) \mathrm{Im} h(\boldsymbol{\theta},s) \\
&X_j^m X_k^r \cos(sX_j + \tfrac{m\pi}{2}) \cos(sX_k + \tfrac{r\pi}{2}) \mathrm{Re} h(\boldsymbol{\theta},s) \;-\; X_j^m X_k^r \cos(sX_j + \tfrac{m\pi}{2}) \sin(sX_k + \tfrac{r\pi}{2}) \mathrm{Im} h(\boldsymbol{\theta},s) \\
&X_j^r X_k^r \sin(sX_j + \tfrac{r\pi}{2}) \sin(sX_k + \tfrac{r\pi}{2}) \mathrm{Re}^2 h(\boldsymbol{\theta},s) \;+\; X_j^r X_k^m \sin(sX_j + \tfrac{r\pi}{2}) \sin(sX_k + \tfrac{m\pi}{2}) \mathrm{Re} h(\boldsymbol{\theta},s) \\
&X_j^r X_k^r \cos(sX_j + \tfrac{r\pi}{2}) \cos(sX_k + \tfrac{r\pi}{2}) \mathrm{Im} h^2(\boldsymbol{\theta},s) \;+\; X_j^r X_k^m \cos(sX_j + \tfrac{r\pi}{2}) \sin(sX_k + \tfrac{m\pi}{2}) \mathrm{Im} h(\boldsymbol{\theta},s) \\
&X_j^m X_k^r \sin(sX_j + \tfrac{m\pi}{2}) \sin(sX_k + \tfrac{r\pi}{2}) \mathrm{Re} h(\boldsymbol{\theta},s) \;+\; X_j^m X_k^r \sin(sX_j + \tfrac{m\pi}{2}) \cos(sX_k + \tfrac{r\pi}{2}) \mathrm{Im} h(\boldsymbol{\theta},s) \\
&X_j^m X_k^m \cos(sX_j + \tfrac{m\pi}{2}) \cos(sX_k + \tfrac{m\pi}{2}) \;+\; X_j^m X_k^m \sin(sX_j + \tfrac{m\pi}{2}) \sin(sX_k + \tfrac{m\pi}{2}) \\
&\qquad\qquad \Big\} \mathrm{d}W(s).
\end{aligned} \tag{4.28}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Now, the i,j-th element of the matrix $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is obtained by taking

$$\omega_{i,j}(\boldsymbol{\theta}) = \mathbb{E}\Big[\frac{\partial^2 \psi_{\boldsymbol{\theta}}(X_1, X_2)}{\partial\theta_i \partial\theta_j}\Big].$$

By recognizing that $\mathbb{E}[X^m \cos(sX_j + \frac{m\pi}{2})] = u_{\boldsymbol{\theta}}^{(m)}(s)$ and $\mathbb{E}[X^m \sin(sX_j + \frac{m\pi}{2})] = v_{\boldsymbol{\theta}}^{(m)}(s)$, we obtain (4.27). In turn, the elements of the matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are the covariances of the vector

$$\sigma_i(\boldsymbol{\theta}, X) = \mathbb{E}\Big[\frac{\partial \psi_{\boldsymbol{\theta}}(x_1, X_2)}{\partial\theta_i}\Big]_{x_1 = X}, \quad i = 1, .., p,$$

and this leads to the expression (4.26).

4.3 Influence functions of the estimators

In this section, we derive IFs of the estimators (4.1), (4.2) and (4.3).

Theorem 4. *Let $\widehat{\boldsymbol{\theta}}_n^{\,\mathrm{LT}\, r,m}$ be the weighted $\mathcal{L}^2$ estimator of the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ defined in (4.1). Under the assumptions I-IV, the joint influence function of the corresponding functional is*

$$\mathrm{IF}(x; \boldsymbol{\theta}^{\,\mathrm{LT}\, r,m}) \; = \; \boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}(x; \boldsymbol{\theta})\,,$$

where $\boldsymbol{K}(\boldsymbol{\theta})$ is the $p \times p$ symmetric matrix whose (i,j)-th element is

$$\kappa_{i,j}(\boldsymbol{\theta}) = \int_0^\infty \Big\{ \Big[\frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i \partial\theta_j} h(\boldsymbol{\theta}, s) + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_i} \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_j} \Big] [L_{\boldsymbol{\theta}}^{(r)}(s)]^2 + \frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial\theta_i \partial\theta_j} L_{\boldsymbol{\theta}}^{(r)}(s) L_{\boldsymbol{\theta}}^{(m)}(s) \Big\} \mathrm{d}W(s),$$

and $\boldsymbol{\nu}(x; \boldsymbol{\theta})$ is the $p \times 1$ vector whose j-th element is

$$\nu_j(x; \boldsymbol{\theta}) \; = \int_0^\infty \Big\{ 2 \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_j} h(\boldsymbol{\theta}, s) L_{\boldsymbol{\theta}}^{(r)}(s) \mathrm{IF}[x; L_{\boldsymbol{\theta}}^{(r)}(s)] + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial\theta_j} \mathrm{IF}[x; L_{\boldsymbol{\theta}}^{(r)}(s) L_{\boldsymbol{\theta}}^{(m)}(s)] \Big\} \mathrm{d}W(s),$$

with

$$\begin{aligned} \mathrm{IF}[x; L_{\boldsymbol{\theta}}^{(r)}(s)] \; &= \; (-x)^r \exp(-sx) - L_{\boldsymbol{\theta}}^{(r)}(s)\,, \\ \mathrm{IF}[x; L_{\boldsymbol{\theta}}^{(r)}(s) L_{\boldsymbol{\theta}}^{(m)}(s)] \; &= \; L_{\boldsymbol{\theta}}^{(m)}(s)(-x)^r \exp(-sx) + L_{\boldsymbol{\theta}}^{(r)}(s)(-x)^m \exp(-sx) - 2L_{\boldsymbol{\theta}}^{(r)}(s) L_{\boldsymbol{\theta}}^{(m)}(s). \end{aligned}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Proof. We know that the estimator solves the set of equations

$$0 = \frac{\partial}{\partial \theta_j} \int_0^\infty \left[h(\boldsymbol{\theta}, s)\widehat{L}_n^{(r)}(s) + \widehat{L}_n^{(m)}(s)\right]^2 \mathrm{d}W(s), \qquad j = 1, ..., p.$$

It is defined implicitly, and, thus, its IF must be obtained by implicit differentiation. Assuming that differentiation under integration sign is justified, we have

$$\begin{aligned} 0 &= \int_0^\infty 2\left[h(\boldsymbol{\theta}, s)\widehat{L}_n^{(r)}(s) + \widehat{L}_n^{(m)}(s)\right]\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\widehat{L}_n^{(r)}(s)\mathrm{d}W(s) \\ &= \int_0^\infty \left[\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j} h(\boldsymbol{\theta}, s)[\widehat{L}_n^{(r)}(s)]^2 + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\widehat{L}_n^{(r)}(s)\widehat{L}_n^{(m)}(s)\right]\mathrm{d}W(s)\,, \qquad j = 1, ..., p. \end{aligned}$$

Replacing the empirical quantities by their corresponding functionals, we compute the IF of both sides applying the product rule for derivatives. Namely, for $j = 1, ..., p$, we obtain

$$\begin{aligned} 0 = \int_0^\infty \Big\{ &\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j} h(\boldsymbol{\theta}, s)\mathrm{IF}\left[x; [L^{(r)}(s)]^2\right] + [L^{(r)}(s)]^2 \mathrm{IF}\left[x; \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j} h(\boldsymbol{\theta}, s)\right] \\ &+ \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\mathrm{IF}\left[x; L^{(r)}(s)L^{(m)}(s)\right] + L^{(r)}(s)L^{(m)}(s)\mathrm{IF}\left[x; \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\right]\Big\}\mathrm{d}W(s)\,. \end{aligned} \tag{4.29}$$

Two of the IFs in the expression above can be computed using the chain rule. We obtain

$$\begin{aligned} \mathrm{IF}\left[x; \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j} h(\boldsymbol{\theta}, s)\right] &= \sum_{i=1}^p \frac{\partial}{\partial \theta_i}\left(\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j} h(\boldsymbol{\theta}, s)\right)\mathrm{IF}[x; \theta_i] \\ &= \sum_{i=1}^p \left[\frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial \theta_i \theta_j} h(\boldsymbol{\theta}, s) + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\right]\mathrm{IF}[x; \theta_i], \\ \mathrm{IF}\left[x; \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\right] &= \sum_{i=1}^p \frac{\partial}{\partial \theta_i}\left(\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\right)\mathrm{IF}[x; \theta_i] \;=\; \sum_{i=1}^p \frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial \theta_i \partial \theta_j}\mathrm{IF}[x; \theta_i]\,. \end{aligned}$$

Putting both results back into (4.29) and rearranging, we obtain

$$\begin{aligned} &\sum_{i=1}^p \mathrm{IF}[x; \theta_i] \int_0^\infty \Big\{\Big[\frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial \theta_i \partial \theta_j} h(\boldsymbol{\theta}, s) + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_i}\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\Big][L_{\boldsymbol{\theta}}^{(r)}(s)]^2 + \frac{\partial^2 h(\boldsymbol{\theta}, s)}{\partial \theta_i \partial \theta_j} L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\mathrm{d}W(s) \\ &\quad = \int_0^\infty \Big\{2\frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j} h(\boldsymbol{\theta}, s)L_{\boldsymbol{\theta}}^{(r)}(s)\mathrm{IF}[x; L_{\boldsymbol{\theta}}^{(r)}(s)] + \frac{\partial h(\boldsymbol{\theta}, s)}{\partial \theta_j}\mathrm{IF}[x; L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)]\Big\}\mathrm{d}W(s)\,, \end{aligned}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

which is nothing but $\sum_{i=1}^{p} \kappa_{i,j}(\boldsymbol{\theta})\text{IF}[x;\theta_i] = \nu_j(\boldsymbol{\theta}),\ j = 1,...,p$ (4.29), and finishes the derivation. □

Theorem 5. *Let* $\widehat{\boldsymbol{\theta}}_n^{\text{LT}\,r,m,\ell}$ *be the weighted* $\mathcal{L}^2$ *estimator of the parameter* $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ *defined in (4.2). Under the assumptions I-IV, the joint influence function of the corresponding functional is*

$$\text{IF}(x;\boldsymbol{\theta}^{\text{LT}\,r,m,\ell}) = \boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}(x;\boldsymbol{\theta}),$$

where $\boldsymbol{K}(\boldsymbol{\theta})$ *is the* $p \times p$ *symmetric matrix whose (i,j)-th element is*

$$\kappa_{i,j}(\boldsymbol{\theta}) = \int_0^\infty \Big\{\frac{1}{2}\frac{\partial^2 h^2(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}[L_{\boldsymbol{\theta}}^{(r)}(s)]^2 + \frac{1}{2}\frac{\partial^2 g^2(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}[L_{\boldsymbol{\theta}}^{(m)}(s)]^2 + \frac{\partial^2\{h(\boldsymbol{\theta},s)g(\boldsymbol{\theta},s)\}}{\partial\theta_i\partial\theta_j}L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)$$
$$+ \frac{\partial^2 h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(\ell)}(s) + \frac{\partial^2 g(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}L_{\boldsymbol{\theta}}^{(m)}(s)L_{\boldsymbol{\theta}}^{(\ell)}(s)\Big\}\text{d}W(s),$$

and $\boldsymbol{\nu}(x;\boldsymbol{\theta})$ *is the* $p \times 1$ *vector whose* j*-th element is*

$$\nu_j(x;\boldsymbol{\theta}) = \int_0^\infty \Big\{\frac{1}{2}\frac{\partial h^2(\boldsymbol{\theta},s)}{\partial\theta_j}\text{IF}\Big[x;[L_{\boldsymbol{\theta}}^{(r)}(s)]^2\Big] + \frac{1}{2}\frac{\partial g^2(\boldsymbol{\theta},s)}{\partial\theta_j}\text{IF}\Big[x;[L_{\boldsymbol{\theta}}^{(m)}(s)]^2\Big]$$
$$+ \frac{\partial\{h(\boldsymbol{\theta},s)g(\boldsymbol{\theta},s)\}}{\partial\theta_j}\text{IF}\Big[x;L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)\Big]$$
$$+ \frac{\partial h(\boldsymbol{\theta},s)}{\partial\theta_j}\text{IF}\Big[x;L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(\ell)}(s)\Big] + \frac{\partial g(\boldsymbol{\theta},s)}{\partial\theta_j}\text{IF}\Big[x;L_{\boldsymbol{\theta}}^{(m)}(s)L_{\boldsymbol{\theta}}^{(\ell)}(s)\Big]\Big\}\text{d}W(s),$$

with

$$\text{IF}\Big[x;[L_{\boldsymbol{\theta}}^{(r)}(s)]^2\Big] = 2L_{\boldsymbol{\theta}}^{(r)}(s)\Big[(-x)^r\exp(-sx) - L_{\boldsymbol{\theta}}^{(r)}(s)\Big],$$
$$\text{IF}\Big[x;L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s)\Big] = L_{\boldsymbol{\theta}}^{(m)}(s)(-x)^r\exp(-sx) + L_{\boldsymbol{\theta}}^{(r)}(s)(-x)^m\exp(-sx) - 2L_{\boldsymbol{\theta}}^{(r)}(s)L_{\boldsymbol{\theta}}^{(m)}(s).$$

Proof. The proof parallels that of Theorem 4 but with more intricate algebra. □

Theorem 6. *Let* $\widehat{\boldsymbol{\theta}}_n^{\text{CF}\,r,m}$ *be the weighted* $\mathcal{L}^2$ *estimator of the parameter* $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ *defined in (4.3). Under the assumptions I-IV, the joint influence function of the corresponding functional is*

$$\text{IF}(x;\boldsymbol{\theta}^{\text{CF}\,r,m}) = \boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}(x;\boldsymbol{\theta}),$$

where $\boldsymbol{K}(\boldsymbol{\theta})$ *is the* $p \times p$ *symmetric matrix whose (i,j)-th element is expressed by*

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

$$\kappa_{i,j}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty}\Big\{\frac{1}{2}\frac{\partial^2|h(\boldsymbol{\theta},s)|^2}{\partial\theta_i\partial\theta_j}|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2 + \frac{\partial^2\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}\mathrm{Re}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\} + \frac{\partial^2\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}\mathrm{Im}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big\}\mathrm{d}W(s),$$

and $\boldsymbol{\nu}(x;\boldsymbol{\theta})$ *is the* $p\times 1$ *vector whose* j*-th element is*

$$\nu_j(x;\boldsymbol{\theta}) = \int_{-\infty}^{\infty}\Big\{\frac{1}{2}\frac{\partial|h(\boldsymbol{\theta},s)|^2}{\partial\theta_j}\mathrm{IF}\Big[x;|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2\Big] + \frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}\mathrm{IF}\Big[x;\mathrm{Re}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big] + \frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}\mathrm{IF}\Big[x;\mathrm{Im}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big]\Big\}\mathrm{d}W(s),$$

with

$$\mathrm{IF}\Big[x;|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2\Big] = 2x^r[u_{\boldsymbol{\theta}}^{(r)}(s)\cos(sx+\tfrac{r\pi}{2}) + v_{\boldsymbol{\theta}}^{(r)}(s)\sin(sx+\tfrac{r\pi}{2})] - 2|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2, \quad (4.30)$$

$$\mathrm{IF}\Big[x;\mathrm{Re}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big] = u_{\boldsymbol{\theta}}^{(m)}(s)x^r\cos(sx+\tfrac{r\pi}{2}) + u_{\boldsymbol{\theta}}^{(r)}(s)x^m\cos(sx+\tfrac{m\pi}{2}) - 2u_{\boldsymbol{\theta}}^{(m)}(s)u_{\boldsymbol{\theta}}^{(r)}(s) + v_{\boldsymbol{\theta}}^{(m)}(s)x^r\sin(sx+\tfrac{r\pi}{2}) + v_{\boldsymbol{\theta}}^{(r)}(s)x^m\sin(sx+\tfrac{m\pi}{2}) - 2v_{\boldsymbol{\theta}}^{(m)}(s)v_{\boldsymbol{\theta}}^{(r)}(s) \quad (4.31)$$

$$\mathrm{IF}\Big[x;\mathrm{Im}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big] = v_{\boldsymbol{\theta}}^{(m)}(s)x^r\cos(sx+\tfrac{r\pi}{2}) + u_{\boldsymbol{\theta}}^{(r)}(s)x^m\sin(sx+\tfrac{m\pi}{2}) - 2u_{\boldsymbol{\theta}}^{(r)}(s)v_{\boldsymbol{\theta}}^{(m)}(s) - v_{\boldsymbol{\theta}}^{(r)}(s)x^m\cos(sx+\tfrac{m\pi}{2}) - u_{\boldsymbol{\theta}}^{(m)}(s)x^r\sin(sx+\tfrac{r\pi}{2}) + 2u_{\boldsymbol{\theta}}^{(m)}(s)v_{\boldsymbol{\theta}}^{(r)}(s) \quad (4.32)$$

Proof. The estimator can be written as

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,r,m} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}\int_{-\infty}^{\infty}\Big\{\Big[\widehat{u}_n^{(r)}(s)\mathrm{Re}\,h(\boldsymbol{\theta},s) - \widehat{v}_n^{(r)}(s)\mathrm{Im}\,h(\boldsymbol{\theta},s) + \widehat{u}_n^{(m)}(s)\Big]^2 + \Big[\widehat{v}_n^{(r)}(s)\mathrm{Re}\,h(\boldsymbol{\theta},s) + \widehat{u}_n^{(r)}(s)\mathrm{Im}\,h(\boldsymbol{\theta},s) + \widehat{v}_n^{(m)}(s)\Big]^2\Big\}\mathrm{d}W(s),$$

and, hence, it is the solution to the following set of p estimating equations:

$$0 = \int_{-\infty}^{\infty}\Big\{\Big[\widehat{u}_n^{(r)}(s)\mathrm{Re}\,h(\boldsymbol{\theta},s) - \widehat{v}_n^{(r)}(s)\mathrm{Im}\,h(\boldsymbol{\theta},s) + \widehat{u}_n^{(m)}(s)\Big]\Big[\widehat{u}_n^{(r)}(s)\frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j} - \widehat{v}_n^{(r)}(s)\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}\Big] + \Big[\widehat{v}_n^{(r)}(s)\mathrm{Re}\,h(\boldsymbol{\theta},s) + \widehat{u}_n^{(r)}(s)\mathrm{Im}\,h(\boldsymbol{\theta},s) + \widehat{v}_n^{(m)}(s)\Big]\Big[\widehat{v}_n^{(r)}(s)\frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j} + \widehat{u}_n^{(r)}(s)\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}\Big]\Big\}\mathrm{d}W(s),$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

for $j = 1, ..., p$. Multiplying, canceling some terms, and rearranging the remaining ones yield

$$0 = \int_{-\infty}^{\infty} \Big\{ \Big[\frac{\partial \mathrm{Re}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{Re}\, h(\boldsymbol{\theta}, s) + \frac{\partial \mathrm{Im}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{Im}\, h(\boldsymbol{\theta}, s) \Big] \Big[[\widehat{u}_n^{(r)}(s)]^2 + [\widehat{v}_n^{(r)}(s)]^2 \Big] + \frac{\partial \mathrm{Re}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \Big[\widehat{u}_n^{(r)}(s) \widehat{u}_n^{(m)}(s) + \widehat{v}_n^{(r)}(s) \widehat{v}_n^{(m)}(s) \Big] + \frac{\partial \mathrm{Im}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \Big[\widehat{u}_n^{(r)}(s) \widehat{v}_n^{(m)}(s) - \widehat{v}_n^{(r)}(s) \widehat{u}_n^{(m)}(s) \Big] \Big\} \mathrm{d}W(s)\,, \quad j = 1, ..., p. \tag{4.33}$$

Note that the following equalities arise from complex number algebra:

$$\frac{\partial \mathrm{Re}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{Re}\, h(\boldsymbol{\theta}, s) + \frac{\partial \mathrm{Im}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{Im}\, h(\boldsymbol{\theta}, s) = \frac{1}{2} \frac{\partial \Big([\mathrm{Re}\, h(\boldsymbol{\theta}, s)]^2 + [\mathrm{Im}\, h(\boldsymbol{\theta}, s)]^2 \Big)}{\partial \theta_j} = \frac{1}{2} \frac{\partial |h(\boldsymbol{\theta}, s)|^2}{\partial \theta_j},$$

$$[\widehat{u}_n^{(r)}(s)]^2 + [\widehat{v}_n^{(r)}(s)]^2 = |\widehat{\phi}_n^{(r)}(s)|^2,$$

$$\widehat{u}_n^{(r)}(s) \widehat{u}_n^{(m)}(s) + \widehat{v}_n^{(r)}(s) \widehat{v}_n^{(m)}(s) = \mathrm{Re}\Big\{ \overline{\widehat{\phi}_n^{(r)}(s)} \widehat{\phi}_n^{(m)}(s) \Big\},$$

$$\widehat{u}_n^{(r)}(s) \widehat{v}_n^{(m)}(s) - \widehat{v}_n^{(r)}(s) \widehat{u}_n^{(m)}(s) = \mathrm{Im}\Big\{ \overline{\widehat{\phi}_n^{(r)}(s)} \widehat{\phi}_n^{(m)}(s) \Big\}\,,$$

and, thus, the set of p estimating equations (4.33) can be more compactly represented as

$$0 = \int_{-\infty}^{\infty} \Big\{ \frac{1}{2} \frac{\partial |h(\boldsymbol{\theta}, s)|^2}{\partial \theta_j} |\widehat{\phi}_n^{(r)}(s)|^2 + \frac{\partial \mathrm{Re}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{Re}\Big\{ \overline{\widehat{\phi}_n^{(r)}(s)} \widehat{\phi}_n^{(m)}(s) \Big\} + \frac{\partial \mathrm{Im}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{Im}\Big\{ \overline{\widehat{\phi}_n^{(r)}(s)} \widehat{\phi}_n^{(m)}(s) \Big\} \Big\} \mathrm{d}W(s), \qquad j = 1, ..., p.$$

As in the proof of Theorem 4, the implicit definition of the estimator requires obtaining the IF through implicit differentiation. Substituting the empirical quantities with the corresponding functionals, and computing the IF of both sides via the chain rule (2.48), we get

$$0 = \int_{-\infty}^{\infty} \Big\{ \frac{1}{2} \sum_{i=1}^{p} \frac{\partial^2 |h(\boldsymbol{\theta}, s)|^2}{\partial \theta_i \partial \theta_j} \mathrm{IF}[x; \theta_i] |\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2 + \frac{1}{2} \frac{\partial |h(\boldsymbol{\theta}, s)|^2}{\partial \theta_j} \mathrm{IF}\Big[x; |\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2\Big]$$
$$+ \sum_{i=1}^{p} \frac{\partial^2 \mathrm{Re}\, h(\boldsymbol{\theta}, s)}{\partial \theta_i \partial \theta_j} \mathrm{IF}[x; \theta_i] \mathrm{Re}\Big\{ \overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)} \phi_{\boldsymbol{\theta}}^{(m)}(s) \Big\} + \frac{\partial \mathrm{Re}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{IF}\Big[x; \mathrm{Re}\Big\{ \overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)} \phi_{\boldsymbol{\theta}}^{(m)}(s) \Big\}\Big]$$
$$+ \sum_{i=1}^{p} \frac{\partial^2 \mathrm{Im}\, h(\boldsymbol{\theta}, s)}{\partial \theta_i \partial \theta_j} \mathrm{IF}[x; \theta_i] \mathrm{Im}\Big\{ \overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)} \phi_{\boldsymbol{\theta}}^{(m)}(s) \Big\} + \frac{\partial \mathrm{Im}\, h(\boldsymbol{\theta}, s)}{\partial \theta_j} \mathrm{IF}\Big[x; \mathrm{Im}\Big\{ \overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)} \phi_{\boldsymbol{\theta}}^{(m)}(s) \Big\}\Big] \Big\} \mathrm{d}W(s).$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Rearranging the above gives

$$\sum_{i=1}^{p}\mathrm{IF}[x;\theta_i]\int_{-\infty}^{\infty}\Big\{\frac{1}{2}\frac{\partial^2|h(\boldsymbol{\theta},s)|^2}{\partial\theta_i\partial\theta_j}|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2 \;+\; \frac{\partial^2\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}\mathrm{Re}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}$$
$$+\;\frac{\partial^2\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_i\partial\theta_j}\mathrm{Im}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big\}\mathrm{d}W(s)$$
$$=\int_{-\infty}^{\infty}\Big\{\frac{1}{2}\frac{\partial|h(\boldsymbol{\theta},s)|^2}{\partial\theta_j}\mathrm{IF}\Big[x;|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2\Big] \;+\; \frac{\partial\mathrm{Re}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}\mathrm{IF}\Big[x;\mathrm{Re}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big]$$
$$+\;\frac{\partial\mathrm{Im}\,h(\boldsymbol{\theta},s)}{\partial\theta_j}\mathrm{IF}\Big[x;\mathrm{Im}\Big\{\overline{\phi_{\boldsymbol{\theta}}^{(r)}(s)}\phi_{\boldsymbol{\theta}}^{(m)}(s)\Big\}\Big]\Big\}\mathrm{d}W(s)\,,$$

which is equivalent to $\sum_{i=1}^{p}\kappa_{i,j}(\boldsymbol{\theta})\mathrm{IF}[x;\theta_i]=\nu_j(\boldsymbol{\theta}),\;\; j=1,...,p$ expressed by (4.30). The expressions for the component IFs (4.30), (4.31) and (4.32) follow from the chain rule (2.48). □

4.4 Practical implications of the theory

Theorems 1-6 reveal that the general expressions for the asymptotic covariance matrices and influence functions of the proposed estimators are intricate, requiring laborious computations of single or double integrals. However, such complexity is not unique to our estimators; many minimum distance estimators based on integrated distances and divergences exhibit similar intricacies, as can be seen in relevant theorems in Basu et al. (2011) or Duchesne et al. (1997). Fortunately, as shown in the upcoming sections, these expression may be simplified or even lead to closed forms. For instance, in the case of the CF based estimators for the normal distribution equipped with a Gaussian weight function.

It is crucial to emphasize yet another difficulty: the integrals involve the transforms themselves and their higher-order derivatives. In the worst case, as discussed already at several points, we may not have an analytical formula for the transform (LT or CF). This necessitates the use of a special function (when available in the computational package) or numerical computation. Of course, the derivatives can be then obtained via recursion derived from the differential equation in use. This will be demonstrated in sec. 4.5.3 devoted to estimation of the Rayleigh distribution.

But let us consider the implications for practical usage of the proposed estimators. In statistical practice, the asymptotic variance (or covariance matrix) is needed to determine the approximated confidence interval (or p-dimensional confidence elipsoid) of the estimate for a realized sample $(x_1, x_2, ..., x_n)$; see Casella and Berger (2002). The above mentioned computational costs may be indeed prohibitive in applications where estimator is employed frequently (e.g. many samples are estimated sequentially, possibly in an automated manner). On the other hand, according to the current computer-based trends, we also can bootstrap the confidence intervals; see Efron and Hastie (2016). Note that

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

bootstrapping is even more appropriate in smaller samples, for which the asymptotic normality may not hold. Fortunately, explicit formulas for our estimators facilitate the use of bootstrapping, distinguishing them from various minimum distance estimators based on distribution or density function that lack explicit representations. Examples of the latter can be found in Basu et al. (2011). In a simulation study by Hołyński and Haneczok (2018) on various minimum distance estimators from that monograph, it was observed that numerical problems often arise when optimizing the non-linear distances, especially in non-typical settings of the distribution parameters or in the presence of large outliers. This makes these estimators challenging to simulate and resample.

The complexity of IFs is less problematic in practical terms. IFs are theoretical measures, typically evaluated for illustrative purposes. Therefore, the potential computational burden associated with them does not significantly impact the usability of the estimators, regardless of the application context or sample sizes.

4.5 Estimation in specific distributions

In the six following sections, we construct and thoroughly examine the proposed estimators for five standard distributions and one non-standard distribution, namely

1. normal(μ,σ^2), using CF;
2. exponential(λ), using CF and LT;
3. Rayleigh(ϑ^2), using LT;
4. gamma(α,β), using LT;
5. Pareto(α,β), using LT;
6. normal variance-gamma(μ,α,β) using CF.

We apply either the weight function (4.4) or (4.5), controlled by shape and scale parameters (γ and a), or the step-wise weight function, which results in estimators controlled by a single transform variable s (the transform-matching estimators).

For each distribution, we

- compute ARE w.r.t. the ML estimator in function of a and γ or s;
- analyze the robustness-efficiency trade-off;
- propose data-driven selection of the parameters a or s;
- conduct simulations to compare the proposed estimators with several standard robust and non-robust estimators, including the ML, MM, median-based and selected minimum distance estimators.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

For the normal and gamma distributions, we also study the effect of model misspecification.

4.5.1 Normal distribution

The normal distribution has the following density and CF:

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \tag{4.34}$$

$$\phi_{\boldsymbol{\theta}}(s) = \exp\{i\mu s - \sigma^2 s^2/2\}, \tag{4.35}$$

x, μ, $s \in \mathbb{R}$, $\sigma^2 > 0$. The unknown parameter is $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$. The CF satisfies the first order ODE,

$$(s\sigma^2 - i\mu)\phi_{\boldsymbol{\theta}}(s) + \phi'_{\boldsymbol{\theta}}(s) = 0, \qquad s \in \mathbb{R}. \tag{4.36}$$

To ensure the consistency of estimation, we must verify the parameter-uniqueness in the sense of sec. 3.4.2. For there could be a normal CF with different values of parameters, μ_* and σ_*^2, satisfying (4.36) as well. To exclude this possibility, we suppose that indeed $(s\sigma_*^2 - i\mu_*)\phi_{\boldsymbol{\theta}}(s) + \phi'_{\boldsymbol{\theta}}(s) = 0$. Subtracting this from (4.36), we obtain $[s(\sigma^2 - \sigma_*^2) - i(\mu - \mu_*)]\phi_{\boldsymbol{\theta}}(s) = 0$, meaning that either $\mu = \mu_*$ and $\sigma^2 = \sigma_*^2$ (which contradicts the assumption) or $\phi_{\boldsymbol{\theta}}(s) = 0$ (which is impossible).

4.5.1.1 Considered $\mathcal{L}^2$ estimators

We consider the estimator of the type (4.3) employing $W(s) = w(s)\mathrm{d}s$, $w(s) = s^\gamma e^{-as^2}$, $a > 0$, that is

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,0,1} = \left(\widehat{\mu}_n^{\mathrm{CF}\,0,1}, \widehat{\sigma_n^2}^{\mathrm{CF}\,0,1}\right)^\top = \underset{\substack{\mu \in (-\infty,\infty) \\ \sigma^2 > 0}}{\operatorname{argmin}} \int_{-\infty}^{\infty} \left|(s\sigma^2 - i\mu)\widehat{\phi}_n(s) + \widehat{\phi}'_n(s)\right|^2 s^\gamma e^{-as^2} \mathrm{d}s, \tag{4.37}$$

where $\gamma = 0$ or $\gamma = 2$.

The estimator admits the following closed-form expressions:

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

$$\widehat{\mu}_n^{\,\mathrm{CF}\,0,1} = \begin{cases} \dfrac{\sum\limits_{j,k=1}^{n}(X_j+X_k)\exp\Big(-(X_j-X_k)^2/4a\Big)}{2\sum\limits_{j,k=1}^{n}\exp\Big(-(X_j-X_k)^2/4a\Big)} & \gamma=0 \\[2ex] \dfrac{\sum\limits_{j,k=1}^{n}(X_j+X_k)\exp\Big(-(X_j-X_k)^2/4a\Big)\Big[(X_j-X_k)^2/a-2\Big]}{2\sum\limits_{j,k=1}^{n}\exp\Big(-(X_j-X_k)^2/4a\Big)\Big[(X_j-X_k)^2/a-2\Big]} & \gamma=2\,, \end{cases} \tag{4.38}$$

$$\widehat{\sigma_n^2}^{\,\mathrm{CF}\,0,1} = \begin{cases} \dfrac{-\sum\limits_{j,k=1}^{n}(X_j-X_k)^2\exp\Big(-(X_j-X_k)^2/4a\Big)}{\sum\limits_{j,k=1}^{n}\Big[(X_j-X_k)^2/a-2\Big]\exp\Big(-(X_j-X_k)^2/4a)\Big)} & \gamma=0 \\[2ex] \dfrac{-\sum\limits_{j,k=1}^{n}(X_j-X_k)^2\Big[(X_j-X_k)^2-6a\Big]\exp\Big(-(X_j-X_k)^2/4a\Big)}{4a\sum\limits_{j,k=1}^{n}\Big[(X_j-X_k)^4/4a^2-3(X_j-X_k)^2/a+3\Big]\exp\Big(-(X_j-X_k)^2/4a\Big)} & \gamma=2\,. \end{cases} \tag{4.39}$$

The derivation is sketched in Appendix B.3.

4.5.1.2 Asymptotic normality and influence functions

Asymptotic normality of the aforementioned estimator is asserted by Theorem 3, from which the covariance matrix can be computed. However, it is somewhat more straightforward to derive this matrix from the corresponding IF using the relation (2.44). Therefore, based on Theorem 6, we first present and prove the following theorem.

Theorem 7. *The functional corresponding to the $\mathcal{L}^2$-type estimator $\widehat{\boldsymbol{\theta}}_n^{\,\mathrm{CF}\,0,1} = (\widehat{\mu}_n^{\,\mathrm{CF}\,0,1}, \widehat{\sigma_n^2}^{\,\mathrm{CF}\,0,1})^\top$ defined by (4.37) has the following joint IF for $\gamma=0$*

$$\begin{pmatrix} \mathrm{IF}\Big[x;\mu^{\,\mathrm{CF}\,0,1}\Big] \\ \mathrm{IF}\Big[x;\sigma^{2\,\mathrm{CF}\,0,1}\Big] \end{pmatrix} = \begin{pmatrix} (x-\mu)\exp\Big(\dfrac{-(x-\mu)^2}{4(a+\sigma^2/2)}\Big)\Big(\dfrac{a+\sigma^2}{a+\sigma^2/2}\Big)^{3/2} \\ \Big((x-\mu)^2\dfrac{a+\sigma^2}{a+\sigma^2/2}-\sigma^2\Big)\exp\Big(\dfrac{-(x-\mu)^2}{4(a+\sigma^2/2)}\Big)\Big(\dfrac{a+\sigma^2}{a+\sigma^2/2}\Big)^{3/2} \end{pmatrix},$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

whereas for $\gamma = 2$ this IF is

$$\begin{pmatrix} \mathrm{IF}\left[x; \mu^{\mathrm{CF}\,0,1}\right] \\ \mathrm{IF}\left[x; \sigma^{2\,\mathrm{CF}\,0,1}\right] \end{pmatrix} = \begin{pmatrix} (x-\mu)\Big((x-\mu)^2 \dfrac{2(a+\sigma^2)}{(2a+\sigma^2)^2} - \dfrac{2(a+2\sigma^2)}{2a+\sigma^2}\Big)\exp\Big(\frac{-(x-\mu)^2}{4(a+\sigma^2/2)}\Big)\Big(\frac{a+\sigma^2}{a+\sigma^2/2}\Big)^{3/2} \\ \Big(\dfrac{-(x-\mu)^4(a+\sigma^2)}{6(a+\sigma^2/2)^2} + \dfrac{(x-\mu)^2(a+3\sigma^2)}{a+\sigma^2/2} - \sigma^2\Big)\exp\Big(\frac{-(x-\mu)^2}{4(a+\sigma^2/2)}\Big)\Big(\frac{a+\sigma^2}{a+\sigma^2/2}\Big)^{5/2} \end{pmatrix}.$$

Proof. (Derivation of IF). We use Theorem 6 putting $r = 0$ and $m = 1$, which gives $|\phi_{\boldsymbol{\theta}}^{(r)}(s)|^2 = |\phi_{\boldsymbol{\theta}}(s)|^2 = e^{-s^2\sigma^2}$, and

$$h(\boldsymbol{\theta}, s) = s\sigma^2 - i\mu, \qquad \operatorname{Re} h(\boldsymbol{\theta}, s) = s\sigma^2,$$
$$|h(\boldsymbol{\theta}, s)|^2 = s^2\sigma^4 + \mu^2, \qquad \operatorname{Im} h(\boldsymbol{\theta}, s) = -\mu.$$

The partial derivatives required in the theorem are:

$$\frac{\partial\,\mathrm{Re} h(\boldsymbol{\theta},s)}{\partial\mu} = 0, \qquad \frac{\partial\,\mathrm{Re} h(\boldsymbol{\theta},s)}{\partial\sigma^2} = s, \qquad \frac{\partial^2\,\mathrm{Re} h(\boldsymbol{\theta},s)}{\partial\mu\partial\sigma^2} = 0,$$
$$\frac{\partial\,\mathrm{Im} h(\boldsymbol{\theta},s)}{\partial\mu} = -1, \qquad \frac{\partial\,\mathrm{Im}\, h(\boldsymbol{\theta},s)}{\partial\sigma^2} = 0, \qquad \frac{\partial^2\,\mathrm{Im}\, h(\boldsymbol{\theta},s)}{\partial\mu\partial\sigma^2} = 0,$$
$$\frac{\partial\,|h(\boldsymbol{\theta},s)|^2}{\partial\mu} = 2\mu, \qquad \frac{\partial\,|h(\boldsymbol{\theta},s)|^2}{\partial\sigma^2} = 2s^2\sigma^2, \qquad \frac{\partial^2\,|h(\boldsymbol{\theta},s)|^2}{\partial\mu\partial\sigma^2} = 0,$$
$$\frac{\partial^2\,\mathrm{Re} h(\boldsymbol{\theta},s)}{\partial\mu^2} = 0, \qquad \frac{\partial^2\,\mathrm{Re} h(\boldsymbol{\theta},s)}{\partial(\sigma^2)^2} = 0, \qquad \frac{\partial^2\,|h(\boldsymbol{\theta},s)|^2}{\partial\mu^2} = 2,$$
$$\frac{\partial^2\,|h(\boldsymbol{\theta},s)|^2}{\partial(\sigma^2)^2} = 2s^2, \qquad \frac{\partial^2\,\mathrm{Im} h(\boldsymbol{\theta},s)}{\partial\mu^2} = 0, \qquad \frac{\partial^2\,\mathrm{Im} h(\boldsymbol{\theta},s)}{\partial(\sigma^2)^2} = 0.$$

Since most of the above terms are zero, the integral formulas in Theorem 6 simplify very much. In particular, the entries of the 2×2 matrix $\boldsymbol{K}(\boldsymbol{\theta})$ are $\kappa_{1,2}(\boldsymbol{\theta}) = \kappa_{2,1}(\boldsymbol{\theta}) = 0$ and

$$\kappa_{1,1}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} s^\gamma e^{-(a+\sigma^2)s^2} \mathrm{d}s, \qquad \kappa_{2,2}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} s^{\gamma+2} e^{-(a+\sigma^2)s^2} \mathrm{d}s.$$

Using the integrals (C.2), (C.5) and (C.7) from Appendix C, we find that

$$\boldsymbol{K}(\boldsymbol{\theta}) = \begin{pmatrix} \sqrt{\dfrac{\pi}{a+\sigma^2}} & 0 \\ 0 & \dfrac{1}{2}\sqrt{\dfrac{\pi}{(a+\sigma^2)^3}} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{K}(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{1}{2}\sqrt{\dfrac{\pi}{(a+\sigma^2)^3}} & 0 \\ 0 & \dfrac{3}{4}\sqrt{\dfrac{\pi}{(a+\sigma^2)^5}} \end{pmatrix}$$

for $\gamma = 0$ and $\gamma = 2$, respectively.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

As $\boldsymbol{K}(\boldsymbol{\theta})$ is diagonal in both cases, the sought expression for the IF vector is

$$\begin{pmatrix} \mathrm{IF}\Big[x;\mu^{\,\mathrm{CF}\,0,1}\Big] \\ \mathrm{IF}\Big[x;\sigma^{2\,\mathrm{CF}\,0,1}\Big] \end{pmatrix} = \boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{\nu}(x;\boldsymbol{\theta}) = \begin{pmatrix} \kappa_{1,1}^{-1}(\boldsymbol{\theta})\nu_1(x;\boldsymbol{\theta}) \\ \kappa_{2,2}^{-1}(\boldsymbol{\theta})\nu_2(x;\boldsymbol{\theta}) \end{pmatrix},$$

Computation of the terms expressing the shape of IFs, namely $\nu_1(x;\boldsymbol{\theta})$ and $\nu_2(x;\boldsymbol{\theta})$, is more complicated. The relevant formulas in Theorem 6 turn into

$$\begin{aligned} \nu_1(x;\boldsymbol{\theta}) \;=\; \int_{-\infty}^{\infty} \Big\{ & 2\mu\Big(u_{\boldsymbol{\theta}}(s)\cos(sx) + v_{\boldsymbol{\theta}}(s)\sin(sx) - e^{-s^2\sigma^2}\Big) \\ & -v'_{\boldsymbol{\theta}}(s)\cos(sx) - u_{\boldsymbol{\theta}}(s)x\cos(sx) + 2u_{\boldsymbol{\theta}}(s)v'_{\boldsymbol{\theta}}(s) \\ & +v_{\boldsymbol{\theta}}(s)x\sin(sx) + u'_{\boldsymbol{\theta}}(s)\sin(sx) - 2u'_{\boldsymbol{\theta}}(s)v_{\boldsymbol{\theta}}(s)\Big\} s^{\gamma} e^{-as^2}\mathrm{d}s\,, \end{aligned}$$

and

$$\begin{aligned} \nu_2(x;\boldsymbol{\theta}) \;=\; \int_{-\infty}^{\infty} \Big\{ & 2\sigma^2 s^2\Big(u_{\boldsymbol{\theta}}(s)\cos(sx) + v_{\boldsymbol{\theta}}(s)\sin(sx) - e^{-s^2\sigma^2}\Big) \\ & +s\Big(u'_{\boldsymbol{\theta}}(s)\cos(sx) - u_{\boldsymbol{\theta}}(s)x\sin(sx) - 2u'_{\boldsymbol{\theta}}(s)u_{\boldsymbol{\theta}}(s) \\ & +v'_{\boldsymbol{\theta}}(s)\sin(sx) + v_{\boldsymbol{\theta}}(s)x\cos(sx) - 2v'_{\boldsymbol{\theta}}(s)v_{\boldsymbol{\theta}}(s)\Big)\Big\} s^{\gamma} e^{-as^2}\mathrm{d}s\,, \end{aligned}$$

where $u_{\boldsymbol{\theta}}(s) = e^{-s^2\sigma^2/2}\cos(s\mu)$ and $v_{\boldsymbol{\theta}}(s) = e^{-s^2\sigma^2/2}\sin(s\mu)$. By applying the 'product-to-sum' trigonometric identities and integrating using (C.9), (C.10), (C.11), and (C.12), we obtain the desired expressions. The validity of the IFs was confirmed through simulations. □

Notice that the shape of the IFs of both estimators depends only on σ^2 but not on the location μ, as required for estimators in a location-scale family. In Fig. 4.1, the IFs are plotted for $a = 0.3, 1$ and 3, in an exemplary setting with $\mu = 5$ and $\sigma^2 = 2$. They are compared with the IFs of the unrobust ML estimators, namely the sample mean $\widehat{\mu}_n^{\mathrm{ML}} = \overline{X}_n$ and sample variance $\widehat{\sigma^2}_n^{\mathrm{ML}} = S_n^2$. These simple and well-known IFs are

$$\begin{aligned} \mathrm{IF}[x;\mu^{\,\mathrm{ML}}] \;&=\; x-\mu \\ \mathrm{IF}[x;\sigma^{2\,\mathrm{ML}}] \;&=\; (x-\mu)^2 - \sigma^2\,. \end{aligned} \tag{4.40}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

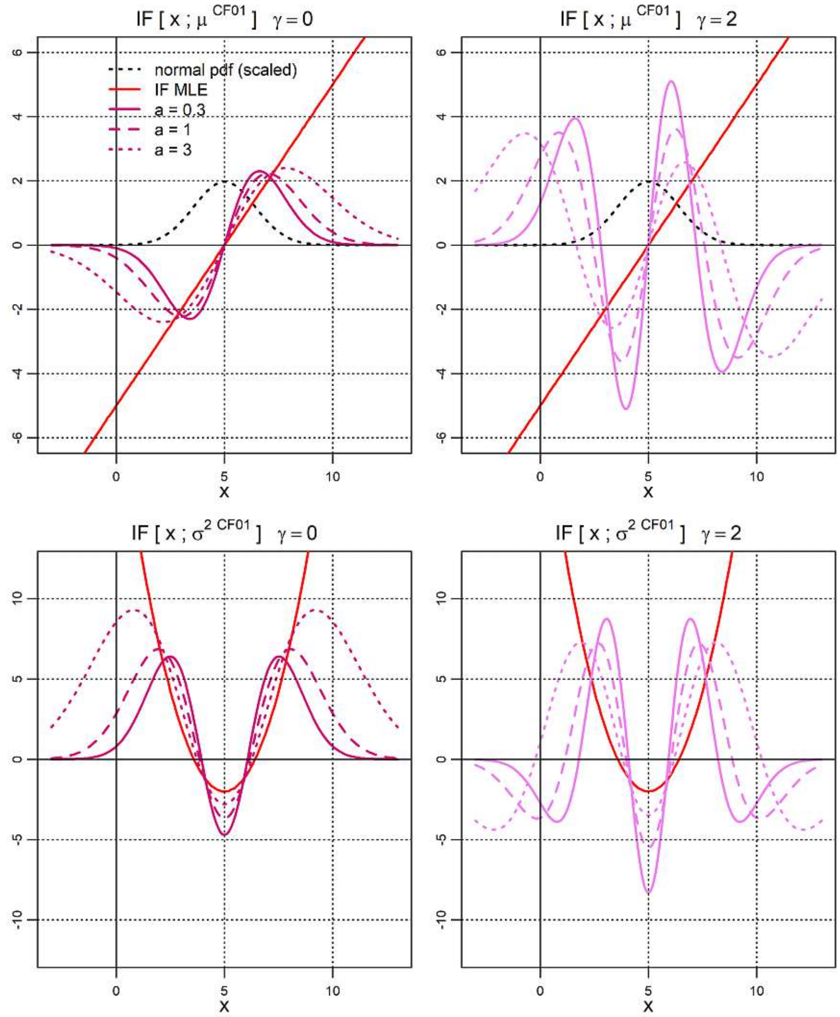


Figure 4.1: Estimation in the normal(μ,σ^2) distribution for $\mu = 5$ and $\sigma^2 = 2$: IFs of the estimators of mean and variance from Theorem 7 compared with IFs of the ML estimators.

Observe that our estimators have three desired properties of robust estimators, see Hampel et al. (1986), namely:

1. their IFs are bounded, i.e. the gross error sensitivity of the estimators expressed generally by $\sup_x \text{IF}[x; T(F)]$ is finite;
2. their IFs are smooth, meaning that the estimators are protected against rounding errors (local shift sensitivity);
3. their IFs redescend to zero as $x \to \infty$, meaning that influence of more and more extreme outliers diminishes to zero.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

When it comes to the third property, the robustness expressed in terms of IF is stronger for $\gamma = 0$ than for $\gamma = 2$. Therefore, the former setting is preferable in applications. Regardless, as $a \to 0$, the rate at which the IF descends to zero increases. Therefore, users can adjust the robustness level by manipulating the parameter a.

We now turn to the asymptotic distribution and efficiency of the estimator. Its asymptotic normality is asserted by the following theorem.

Theorem 8. *The $\mathcal{L}^2$ estimator given by (4.37) converges in distribution to the bivariate normal random variable, that is*

$$\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,0,1} - \boldsymbol{\theta}\Big) \stackrel{\mathrm{d}}{\longrightarrow} \mathcal{N}_2\Big(\mathbf{0}, \boldsymbol{\Sigma}(\sigma^2)\Big),$$

where the asymptotic covariance matrix for $\gamma = 0$ is

$$\boldsymbol{\Sigma}(\sigma^2) = \begin{pmatrix} 8\sigma^2\left(\dfrac{a+\sigma^2}{2a+\sigma^2}\right)^3\left(\dfrac{2a+\sigma^2}{2a+3\sigma^2}\right)^{3/2} & 0 \\ 0 & \dfrac{\sigma^4(9\sigma^4+16a\sigma^2+8a^2)(2a+\sigma^2)^{1/2}(a+\sigma^2)^3}{(2a+3\sigma^2)^{5/2}(a+\sigma^2/2)^3} \end{pmatrix},$$

whereas for $\gamma = 2$ it is

$$\boldsymbol{\Sigma}(\sigma^2) = \begin{pmatrix} \dfrac{32\sigma^2(15\sigma^8+36a\sigma^6+34a^2\sigma^4+16a^3\sigma^2+4a^4)}{(a+\sigma^2)^{-3}[(2a+\sigma^2)(2a+3\sigma^2)]^{7/2}} & 0 \\ 0 & \dfrac{64\sigma^4(97\sigma^8+284a\sigma^6+334a^2\sigma^4+192a^3\sigma^2+48a^4)}{3(a+\sigma^2)^{-5}[(2a+\sigma^2)(2a+3\sigma^2)]^{9/2}} \end{pmatrix}.$$

Proof. (Derivation of the covariance matrix). We use (2.44), which in the present example is

$$\boldsymbol{\Sigma}(\sigma^2) = \int_{-\infty}^{\infty} \mathrm{IF}\Big[x; \boldsymbol{\theta}^{\mathrm{CF}\,0,1}\Big]\mathrm{IF}\Big[x; \boldsymbol{\theta}^{\mathrm{CF}\,0,1}\Big]^\top \mathrm{d}F_{\boldsymbol{\theta}}(x). \tag{4.41}$$

Note that upon change of variables $y = x - \mu$, the cross-terms of the sought matrices for $\gamma = 0$ and $\gamma = 2$ comprise only the Gaussian-type integrals

$$A\int_{-\infty}^{\infty} y^n \exp(-By^2)\mathrm{d}y, \qquad n \text{ odd}, \tag{4.42}$$

(where A and B are some constants dependent on a, σ^2 and π). All these integrals equal zero due to oddity of integrands. Therefore, $\Sigma_{1,2}(\sigma^2) = \Sigma_{2,1}(\sigma^2) = 0$ for $\gamma = 0, 2$. On the other hand, integrals of the same type occurring in diagonal variance terms, $\Sigma_{1,1}(\sigma^2)$ and $\Sigma_{2,2}(\sigma^2)$, have all even integrands, and can be computed using

$$A\int_{-\infty}^{\infty} y^n \exp(-By^2)\mathrm{d}y = \frac{A}{2} B^{-(n+1)/2}\,\Gamma\left(\frac{n+1}{2}\right), \tag{4.43}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

with $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(3/2) = \sqrt{\pi}/2$, $\Gamma(5/2) = 3\sqrt{\pi}/4$, ... □

As Theorem 8 guarantees asymptotic normality with zero correlation, the estimators are also asymptotically independent.

We now examine the limiting forms of the covariance matrices as the parameter of the weight function $a \to \infty$ and $a \to 0$. For $\gamma = 0$, we obtain

$$\lim_{a\to\infty} \mathbf{\Sigma}(\sigma^2) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}, \qquad \lim_{a\to 0} \mathbf{\Sigma}(\sigma^2) = \begin{pmatrix} \frac{8\sqrt{3}}{9}\sigma^2 & 0 \\ 0 & \frac{8\sqrt{27}}{9}\sigma^4 \end{pmatrix},$$

whereas for $\gamma = 2$,

$$\lim_{a\to\infty} \mathbf{\Sigma}(\sigma^2) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}, \qquad \lim_{a\to 0} \mathbf{\Sigma}(\sigma^2) = \begin{pmatrix} \frac{160\sqrt{27}}{81}\sigma^2 & 0 \\ 0 & \frac{6208\sqrt{3}}{729}\sigma^4 \end{pmatrix}.$$

Note that as $a \to \infty$, the matrices tend to that of the ML estimator. This implies that for both $\gamma = 0$ and $\gamma = 2$, the estimators attain full asymptotic efficiency in this limit. On the other hand, as $a \to 0$, ARE w.r.t. ML estimator (defined as the ratio of their limiting variances) drops to a constant non-zero value. In particular, for $\gamma = 0$, we obtain

$$\lim_{a\to 0} \mathrm{ARE}\Big[\widehat{\mu}_n^{\mathrm{CF}\,0,1}, \widehat{\mu}_n^{\mathrm{ML}}\Big] = \frac{3\sqrt{3}}{8} \approx 65\%\,, \qquad \lim_{a\to 0} \mathrm{ARE}\Big[\widehat{\sigma}_n^{2\,\mathrm{CF}\,0,1}, \widehat{\sigma}_n^{2\,\mathrm{ML}}\Big] = \frac{\sqrt{27}}{12} \approx 43\%\,,$$

whereas for $\gamma = 2$,

$$\lim_{a\to 0} \mathrm{ARE}\Big[\widehat{\mu}_n^{\mathrm{CF}\,0,1}, \widehat{\mu}_n^{\mathrm{ML}}\Big] = \frac{3\sqrt{27}}{160} \approx 10\%\,, \qquad \lim_{a\to 0} \mathrm{ARE}\Big[\widehat{\sigma}_n^{2\,\mathrm{CF}\,0,1}, \widehat{\sigma}_n^{2\,\mathrm{ML}}\Big] = \frac{243\sqrt{3}}{3104} \approx 13\%\,.$$

We also derive the joint ARE, which for $\gamma = 0$ is

$$\mathrm{ARE}\Big[\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,0,1}, \widehat{\boldsymbol{\theta}}_n^{\mathrm{ML}}\Big] = \frac{\det[\mathbf{\Sigma}_{\mathrm{ML}}(\sigma^2)]}{\det[\mathbf{\Sigma}(\sigma^2)]} = \frac{(2a+3\sigma^2)^4(a+\sigma^2/2)^3(2a+\sigma^2)}{4(9\sigma^4+16a\sigma^2+8a^4)(a+\sigma^2)^6},$$

whereas for $\gamma = 2$

$$\frac{3[(2a+\sigma^2)(2a+3\sigma^2)/(a+\sigma^2)]^8/1024}{(15\sigma^8+36a\sigma^6+34a^2\sigma^4+16a^3\sigma^2+4a^4)(97\sigma^8+284a\sigma^6+334a^2\sigma^4+192a^3\sigma^2+48a^4)}.$$

The limits of the joint AREs as $a \to \infty$ equal 1 in both cases, while as $a \to 0$ they equal $9/32 \approx 28\%$ and $6561/496640 \approx 1\%$ for $\gamma = 0$ and $\gamma = 2$, respectively. For intermediate a, the individual and joint AREs are complicated functions of a and σ^2, as plotted in Fig. 4.2.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

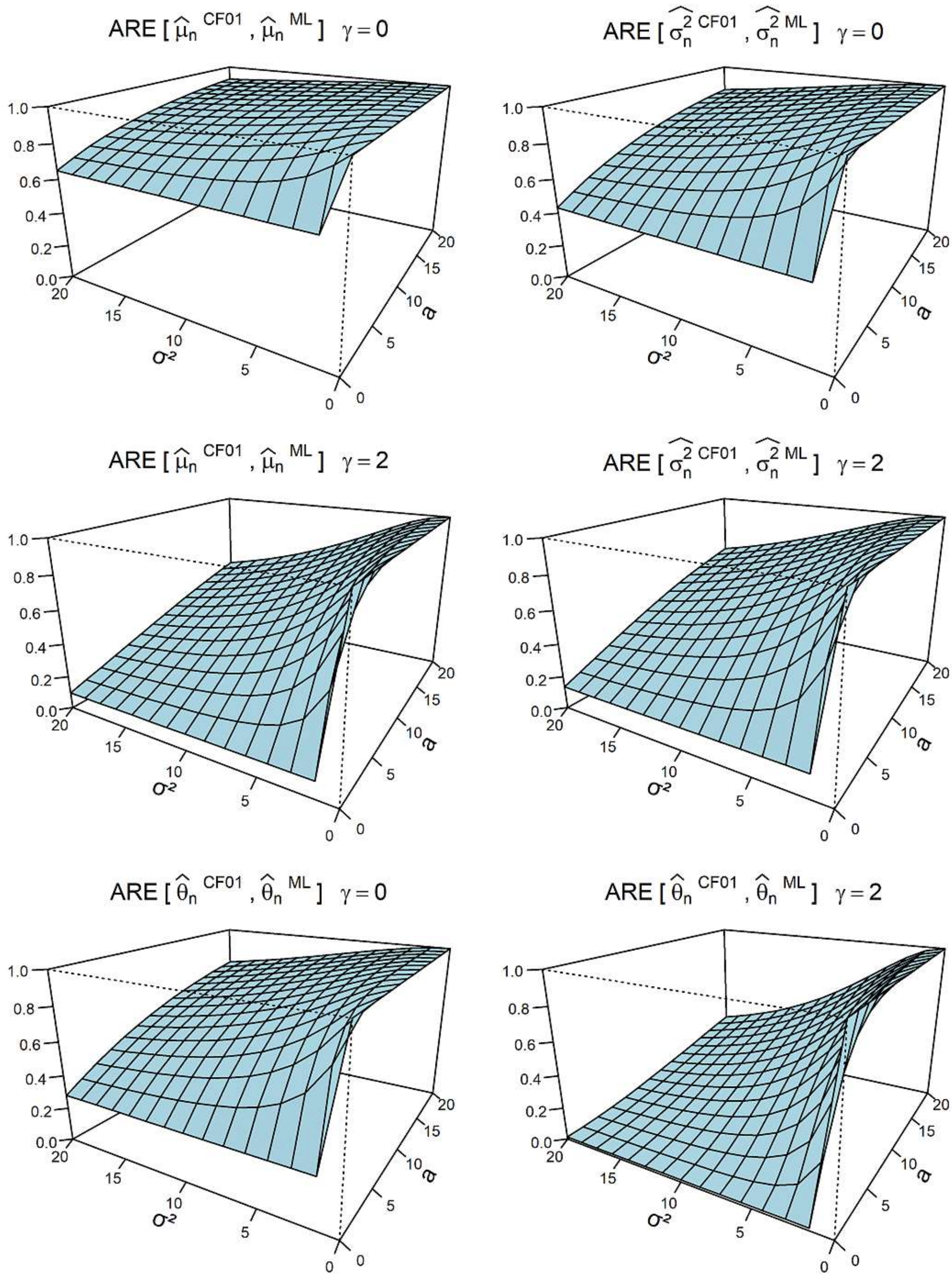


Figure 4.2: Estimation in the normal(μ,σ^2) distribution: Individual and joint asymptotic relative efficiencies (AREs) of the proposed estimators, $\widehat{\mu}_n^{\text{CF}\,0,1}$ and $\widehat{\sigma}_n^{2\,\text{CF}\,0,1}$, with respect to the ML estimators.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

4.5.1.3 Robustness-efficiency trade-offs

The above graphs reveal that a blind choice of the scale parameter of the weight function a w.r.t. the true σ^2 may degrade the efficiency of the proposed estimators. For the setting $\gamma = 0$, the consequences are not that drastic: the efficiency may be then, at worst, comparable with that of sample median for μ and the squared median absolute deviation (MAD) for σ (these are popular robust estimators of location and scale having, respectively, 63% and 37% ARE at the normal distribution; see, e.g., Rousseeuw and Croux (1993)). However, for $\gamma = 2$, the degradation of our estimators may be far more severe. Of course, we may set a extremely large to get ARE close to 1 with high probability, but, according to IFs in Fig. 4.1, this would produce estimates completely unrobust against large outliers.

Fortunately, a combined analysis of Figs. 4.1 and 4.2 suggests a solution. A reasonable trade-off between efficiency and robustness can be achieved by setting a as *linearly proportional* to σ^2, that is

$$a = k\sigma^2 \,, \tag{4.44}$$

where k is a proportionality constant. In this way, the estimator can operate along the lines of constancy of ARE, which can be seen in Fig. 4.2. For example, observe the top left panel of Fig. 4.2 to note that, with $a = \sigma^2$ ($k = 1$), the estimator $\widehat{\mu}_n^{\mathrm{CF}\,0,1}$ with $\gamma = 0$ has efficiency equal to about 90%. On the other hand, the top left panel of Fig. 4.1 (where $\sigma^2 = 2$) shows that setting $a = 2$ produces a very satisfactory IF. Putting $k > 1$ increases efficiency but sacrifices some robustness, and vice versa.

Since σ^2 is unknown, we propose using an ancillary estimator of it to find good value of a. This estimator must be neither very precise nor robust and should provide only a rough idea about the value of σ^2. Instead, it should be easily computable so that the effort associated with computing (4.38) and (4.39) is not significantly increased. Consequently, we consider two variants:

$$a = kS_n^2 \,, \tag{4.45}$$

with S_n^2 being the sample variance, and

$$a = k(\mathrm{MAD}_n)^2 \,. \tag{4.46}$$

Here, MAD_n stands for the *median absolute deviation* computed by

$$\mathrm{MAD}_n = b \cdot \mathrm{med}\Big(|X_1 - \widehat{\mathrm{med}}_n|,\ \ldots, |X_n - \widehat{\mathrm{med}}_n|\Big) \,, \tag{4.47}$$

where $\widehat{\mathrm{med}}_n$ is the sample median and b is a consistency factor equal to 1.483 for the normal distribution.

The constant k remains controlled by the user and decides about the degree of the trade-off.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

4.5.1.4 Finite-sample performance: simulation analysis

The usefulness of data-driven tuning of the estimators designed in asymptotic setting must be now confirmed by simulations in finite samples. Figs. 4.3 and 4.4 show results of estimation of μ and σ^2 using $\widehat{\mu}_n^{\text{CF}\,0,1}$, in (4.38), and $\widehat{\sigma_n^2}^{\text{CF}\,0,1}$, in (4.39), for 5000 samples[1] of size $n = 50$, with and without outliers, and for two very different values of $\sigma^2 = 0.01$ and $\sigma^2 = 100$, whereby $\mu = 0$ in each case. Fig. 4.3 reports the case when a is estimated by sample variance (4.45), while Fig. 4.4 reports situation when a is estimated by MAD (4.46), each with $k = 1, 2$ and 3.

The proposed estimators are compared with ML estimators and with two robust alternatives. For μ these alternatives are: 1) the sample median and 2) the Hodges-Lehmann estimator[2] defined as the median of the averages of the $n(n-1)/2$ possible pairs of observations; see Hodges and Lehmann (1963) and Hampel et al. (1986). The Hodges-Lehmann estimator is a universal non-parametric estimator of median with 97% ARE at the normal model. For σ^2, the two robust alternatives are: 1) the square of MAD given in (4.47) and 2) the square of the so-called Q_n estimator[3] expressed by

$$Q_n = 2.222 \cdot \Big\{|X_i - X_j|, i < j\Big\}_{(r)}, \tag{4.48}$$

where $r = \binom{\lfloor n/2\rfloor+1}{2}$ and $\{\cdot\}_{(r)}$ denotes the r-th order statistic of the set. The Q_n estimator, introduced by Rousseeuw and Croux (1993), has 82% ARE at the normal distribution.

Each estimator is represented by two box plots. The right one refers to pure samples, while the left one corresponds to the same samples contaminated by a single outlier $X_0 = \mu + 5\widehat{\sigma}_n$, where $\widehat{\sigma}_n$ is the sample standard deviation of the original sample $X_1, ..., X_{50}$ (X_0 replaces X_1 in each sample).

Let us analyze the results. In general, the proposed data-driven estimators exhibit a good robustness-variance trade-off. Their spread is comparable to that of ML estimators, and their bias is not greater than that of the robust alternatives when an outlier is present. As anticipated previously, $\gamma = 0$ gives higher robustness than $\gamma = 2$. Additionally, as we observe now, $\gamma = 2$ introduces a slight negative bias, unrelated to contamination. However, this effect is also observed for sample median and the Q_n estimator. Moreover, we notice that differences in performance of our estimators when driven by (4.45) or (4.46) are minor. The latter makes the estimator with $\gamma = 0$ more slightly more robust when k is increased from 1 to 3. A particularly good performance occurs for $\gamma = 0$, $a = k(\text{MAD}_n)^2$ and $k = 1$. For the estimation of μ, this estimator emerges as a strong competitor to the celebrated Hodges-Lehmann estimator.

[1] To provide reproducible results, the simulation seed in R was fixed using set.seed(1). This seed was used for all simulations in the thesis.

[2] To compute the Hodges-Lehmann estimator, we used the function HodgesLehmann(·) from the R package DescTools.

[3] To compute the Q_n estimator, we used the function Qn(·) from the package robustbase in R.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Fig. 4.5 presents the estimation of $\mu = 0$ and $\sigma^2 = 100$ under more severe contamination. In this case, samples now contain three outliers: $X_{0,1} = \mu + 3\widehat{\sigma}_n$, $X_{0,2} = \mu + 4\widehat{\sigma}_n$ and $X_{0,3} = \mu + 5\widehat{\sigma}_n$. We employed tuning with $a = k(\mathrm{MAD}_n)^2$. The resistance of the proposed estimators to such a batch of outliers remains very good, especially with $\gamma = 0$ and $k = 1$.

4.5.1.5 Estimation under misspecification

The previous section addressed point contamination of the normal model. Now, we investigate the behavior of estimators under full misspecification of the model, meaning that the sample comes from a completely different distribution.

Typically, the impact of misspecification on estimator is assessed analytically by computing the loss of its efficiency caused by the wrong assumption (this is reasonable if the parameter is shared by the true model and the wrongly assumed one). Due to complexity of our estimators, such analysis is not tractable. Therefore, we assess the effect through simulations, focusing only the estimator of location $\widehat{\mu}_n^{\mathrm{CF}\,0,1}$ with $\gamma = 0$. We apply the estimator to samples generated from the following symmetric unimodal distributions: normal(μ, σ^2), Laplace(μ, β), logistic(μ, β), Cauchy (μ, β), all with $\mu = 0$ and $\sigma^2, \beta = 1$ as well as Student-t(ν) with $\nu = 3$ degrees of freedom. In simulations, the variance of our estimator is compared with that of ML, sample median and the Hodges-Lehmann estimators used in the same scenarios for sample sizes from $n = 30$ to 500. The results are displayed in Table 4.1, where bold font indicates the best result (smallest empirical variance) per sample size.

We observe that our estimator performs very well in case of the light-tailed models (Laplace, logistic, Student-t) across all sample sizes. Notably, in the Student-t case, it even performs slightly better than the Hodges-Lehmann estimator in small samples. Conversely, for the heavy-tailed Cauchy data, the estimator is clearly suboptimal compared to the sample median and the Hodges-Lehmann estimator, although it still outperforms the sample mean, which is inconsistent in this scenario. This arises because $\widehat{\mu}_n^{\mathrm{CF}\,0,1}$ is partially based on the first derivative of CF near $s = 0$ which may take extreme values due to the lack of the mean in the Cauchy distribution. Fig. 4.6 displays kernel-smoothed densities of the estimators for the Laplace and Cauchy models, illustrating more extreme realizations of $\widehat{\mu}_n^{\mathrm{CF}\,0,1}$ in the Cauchy case.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

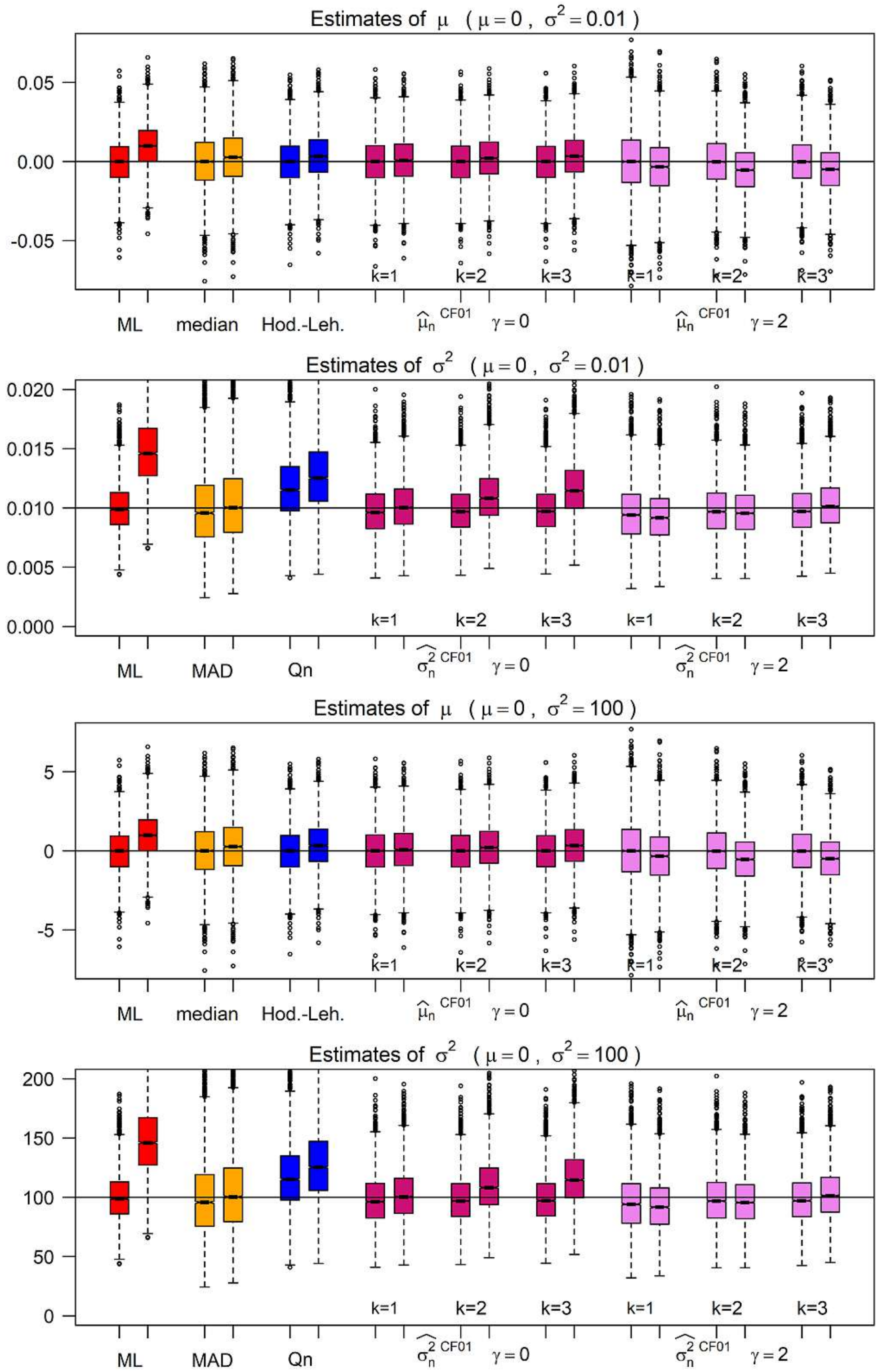


Figure 4.3: Estimation of μ and σ^2 in the normal(μ,σ^2) model for $\mu = 0$ and $\sigma^2 = 0.01$ or $\sigma^2 = 100$, without and with one outlier $x_0 = \mu + 5\widehat{\sigma}_n$ (right and left box plot, resp.); $n = 50$; $a = kS_n^2$, $k = 1, 2, 3$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

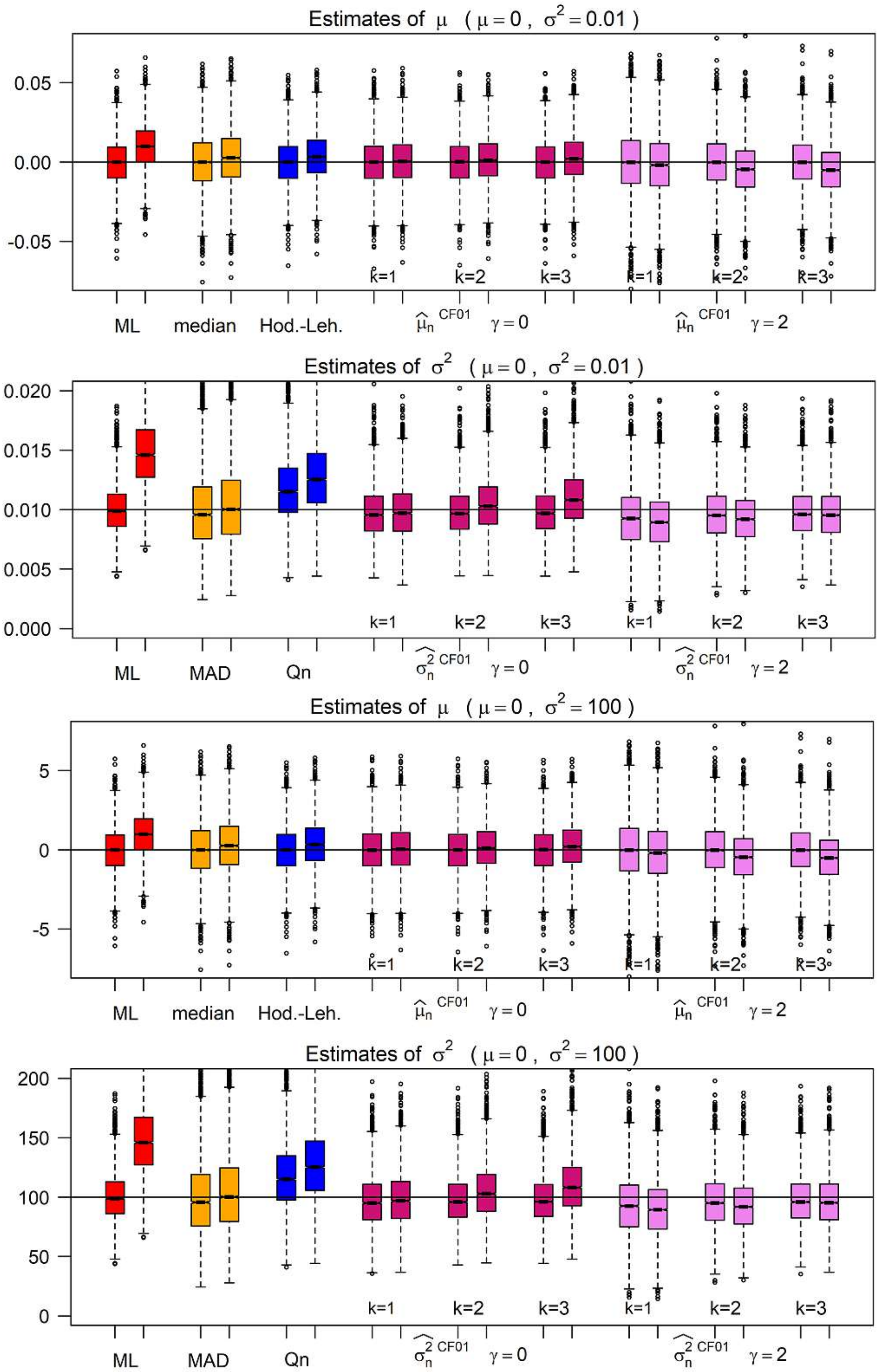


Figure 4.4: Estimation of μ and σ^2 in the normal(μ,σ^2) model for $\mu = 0$ and $\sigma^2 = 0.01$ or $\sigma^2 = 100$, without and with one outlier $x_0 = \mu + 5\widehat{\sigma}_n$ (right and left box plot, resp.); $n = 50$; $a = k(\text{MAD}_n)^2$, $k = 1, 2, 3$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

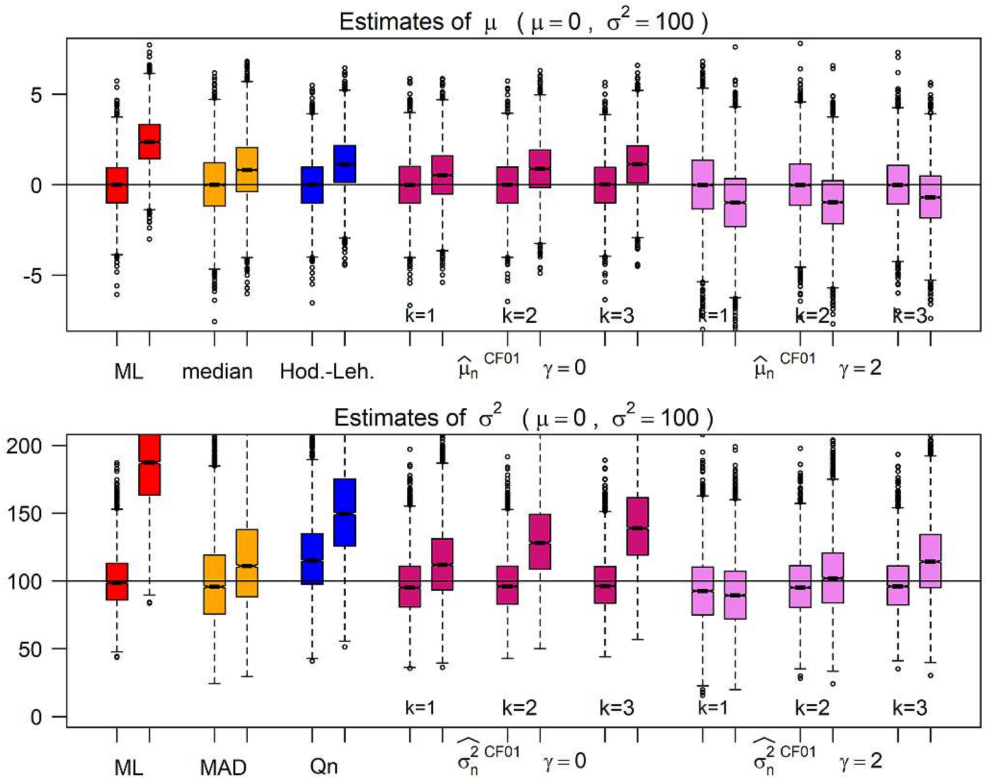


Figure 4.5: Estimation of μ and σ^2 in the normal(μ,σ^2) model for $\mu = 0$ and $\sigma^2 = 100$, without and with three outliers $x_{0,1} = \mu + 3\widehat{\sigma}_n$, $x_{0,2} = \mu + 4\widehat{\sigma}_n$, $x_{0,3} = \mu + 5\widehat{\sigma}_n$ (right and left box plot, resp.); $n = 50$; $a = k(\mathrm{MAD}_n)^2$, $k = 1, 2, 3$.

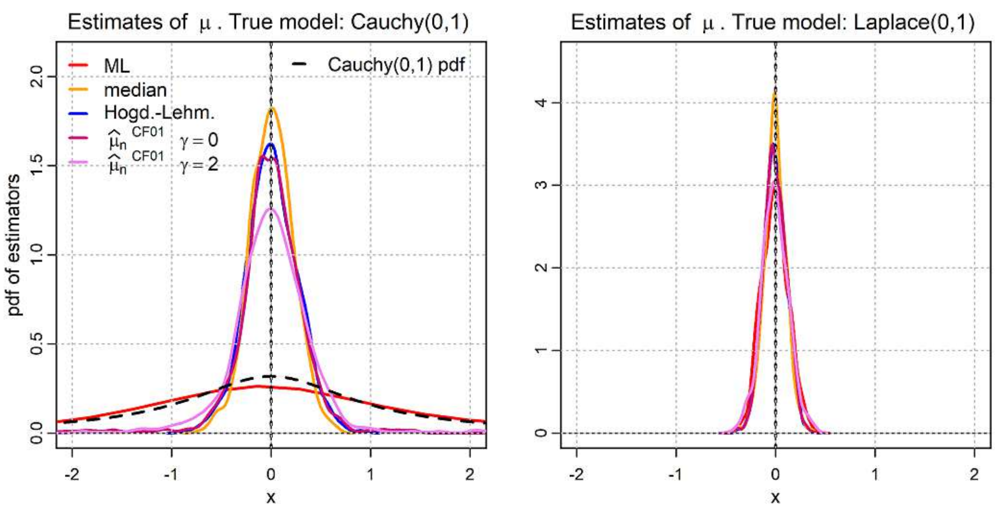


Figure 4.6: Smoothed densities of the estimators of μ and σ^2 for the normal(μ,σ^2) distribution under model misspecification. True models: Cauchy(0,1) and Laplace(0,1).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Table 4.1: Variances of the estimators of location μ under model misspecification.

True model: Normal(0,1)

n	sample mean	sample median	Hodges-Lehmann	$\widehat{\mu}_n^{\mathrm{CF}\,0,1}$, $(\gamma = 0)$
30	**0.033**	0.050	0.035	0.037
50	**0.020**	0.031	0.022	0.023
70	**0.014**	0.022	0.015	0.015
100	**0.010**	0.016	0.011	0.011
200	**0.005**	0.008	**0.005**	0.006
500	**0.002**	0.003	**0.002**	**0.002**

True model: Laplace(0,1)

n	sample mean	sample median	Hodges-Lehmann	$\widehat{\mu}_n^{\mathrm{CF}\,0,1}$, $(\gamma = 0)$
30	0.033	**0.021**	0.023	0.023
50	0.020	**0.012**	0.014	0.014
70	0.014	**0.008**	0.010	0.010
100	0.010	**0.006**	0.007	0.007
200	0.005	**0.003**	**0.003**	**0.003**
500	0.002	**0.001**	**0.001**	**0.001**

True model: Logistic(0,1)

n	sample mean	sample median	Hodges-Lehmann	$\widehat{\mu}_n^{\mathrm{CF}\,0,1}$, $(\gamma = 0)$
30	0.109	0.131	**0.101**	0.104
50	0.064	0.076	**0.059**	0.061
70	0.048	0.057	**0.044**	0.045
100	0.033	0.040	**0.031**	**0.031**
200	0.016	0.020	**0.015**	**0.015**
500	0.007	0.008	**0.006**	**0.006**

True model: Student-t(3)

n	sample mean	sample median	Hodges-Lehmann	$\widehat{\mu}_n^{\mathrm{CF}\,0,1}$, $(\gamma = 0)$
30	0.095	0.062	0.055	**0.053**
50	0.059	0.037	0.032	**0.031**
70	0.043	0.026	0.023	**0.022**
100	0.029	0.019	**0.016**	**0.016**
200	0.016	0.009	**0.008**	**0.008**
500	0.006	0.004	**0.003**	**0.003**

True model: Cauchy(0,1)

n	sample mean	sample median	Hodges-Lehmann	$\widehat{\mu}_n^{\mathrm{CF}\,0,1}$, $(\gamma = 0)$
30	7026	**0.086**	0.123	25.46
50	2115	**0.051**	0.071	3.350
70	28083	**0.036**	0.050	20.42
100	1524	**0.025**	0.034	0.530
200	69222	**0.012**	0.017	6.340
500	12731	**0.005**	0.007	0.174

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

4.5.2 Exponential distribution

The exponential distribution has density

$$f_{\boldsymbol{\theta}}(x) \quad = \quad \lambda e^{-\lambda x}\,, \quad x \geq 0, \tag{4.49}$$

where $\lambda > 0$ is the rate parameter.

For this distribution, we analyze an entire family of estimators based on differential equations satisfied by both LT and CF. In deriving these estimators, we make use of equations up to the second order, including the 'zero-order' equation representing the transform expression itself. In addition to the $\mathcal{L}^2$ estimators, we also consider the transform-matching estimators of the form (4.8).

Recall that LT and CF of the exponential distribution are

$$L_{\boldsymbol{\theta}}(s) = \frac{\lambda}{\lambda + s}, \qquad \phi_{\boldsymbol{\theta}}(s) = \frac{\lambda}{\lambda - is}, \tag{4.50}$$

and the real and imaginary parts of the latter are given by

$$u_{\boldsymbol{\theta}}(s) = \frac{\lambda^2}{\lambda^2 + s^2}, \qquad v_{\boldsymbol{\theta}}(s) = \frac{\lambda s}{\lambda^2 + s^2}.$$

The LT and CF satisfy the following differential equations: for $s > -\lambda$,

$$L_{\boldsymbol{\theta}}(s) - \lambda/(\lambda + s) \;=\; 0\,, \tag{4.51}$$

$$(\lambda + s)L'_{\boldsymbol{\theta}}(s) + L_{\boldsymbol{\theta}}(s) \;=\; 0\,, \tag{4.52}$$

$$(\lambda + s)L''_{\boldsymbol{\theta}}(s) + 2L'_{\boldsymbol{\theta}}(s) \;=\; 0\,, \tag{4.53}$$

$$(\lambda + s)^2 L''_{\boldsymbol{\theta}}(s) - 2L_{\boldsymbol{\theta}}(s) \;=\; 0\,, \tag{4.54}$$

and, for $s \in \mathbb{R}$,

$$\phi_{\boldsymbol{\theta}}(s) - \lambda/(\lambda - is) \;=\; 0\,, \tag{4.55}$$

$$(s + i\lambda)\phi'_{\boldsymbol{\theta}}(s) + \phi_{\boldsymbol{\theta}}(s) \;=\; 0\,, \tag{4.56}$$

$$(s + i\lambda)\phi''_{\boldsymbol{\theta}}(s) + 2\phi'_{\boldsymbol{\theta}}(s) \;=\; 0\,, \tag{4.57}$$

$$(s + i\lambda)^2\phi''_{\boldsymbol{\theta}}(s) - 2\phi_{\boldsymbol{\theta}}(s) \;=\; 0\,. \tag{4.58}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.5.2.1 Considered $\mathcal{L}^2$ estimators

Using the eight equations mentioned above, we define the $\mathcal{L}^2$ estimators of the types (4.1) and (4.3), incorporating the weight functions (4.4) and (4.5), as follows:

$$\widehat{\lambda}_n^{\mathrm{LT}\,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_0^\infty \Big[\widehat{L}_n(s) - \frac{\lambda}{\lambda+s}\Big]^2 s^\gamma e^{-as} \mathrm{d}s\,, \tag{4.59}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,1,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_0^\infty \Big[(\lambda+s)\widehat{L}_n'(s) + \widehat{L}_n(s)\Big]^2 s^\gamma e^{-as} \mathrm{d}s\,, \tag{4.60}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,2,1} = \underset{\lambda>0}{\operatorname{argmin}} \int_0^\infty \Big[(\lambda+s)\widehat{L}_n''(s) + 2\widehat{L}_n'(s)\Big]^2 s^\gamma e^{-as} \mathrm{d}s, \,, \tag{4.61}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,2,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_0^\infty \Big[(\lambda+s)^2\widehat{L}_n''(s) - 2\widehat{L}_n(s)\Big]^2 s^\gamma e^{-as} \mathrm{d}s\,, \tag{4.62}$$

for $\gamma = 0, 1, 2, ...,$ and

$$\widehat{\lambda}_n^{\mathrm{CF}\,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_{-\infty}^\infty \Big|\widehat{\phi}_n(s) - \lambda/(\lambda - is)\Big|^2 s^\gamma e^{-as^2} \mathrm{d}s\,, \tag{4.63}$$

$$\widehat{\lambda}_n^{\mathrm{CF}\,1,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_{-\infty}^\infty \Big|(s+i\lambda)\widehat{\phi}_n'(s) + \widehat{\phi}_n(s)\Big|^2 s^\gamma e^{-as^2} \mathrm{d}s\,, \tag{4.64}$$

$$\widehat{\lambda}_n^{\mathrm{CF}\,2,1} = \underset{\lambda>0}{\operatorname{argmin}} \int_{-\infty}^\infty \Big|(s+i\lambda)\widehat{\phi}_n''(s) + 2\widehat{\phi}_n'(s)\Big|^2 s^\gamma e^{-as^2} \mathrm{d}s\,, \tag{4.65}$$

$$\widehat{\lambda}_n^{\mathrm{CF}\,2,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_{-\infty}^\infty \Big|(s+i\lambda)^2\widehat{\phi}_n''(s) - 2\widehat{\phi}(s)\Big|^2 s^\gamma e^{-as^2} \mathrm{d}s\,, \tag{4.66}$$

for $\gamma = 0, 2, 4, ...,$ with $a > 0$ in all the cases.

The estimators (4.59) and (4.63) do not admit closed-form expressions for any γ. For $\gamma = 0$, integration can be performed in (4.59) to obtain

$$\widehat{\lambda}_n^{\mathrm{LT}\,0} = \underset{\lambda>0}{\operatorname{argmin}} \Big[\lambda - \frac{2\lambda}{n}\sum_{j=1}^n e^{\lambda(X_j+a)} E_1(\lambda(X_j+a)) - a\lambda^2 e^{\lambda a} E_1(\lambda a)\Big]\,, \tag{4.67}$$

where $E_1(z) = \int_z^\infty e^{-t}/t\,\mathrm{d}t$ is the exponential integral.

In turn, the integrand in the estimator (4.63) can also be written in somewhat simpler form, so that

$$\widehat{\lambda}_n^{\mathrm{CF}\,0} = \underset{\lambda>0}{\operatorname{argmin}} \int_{-\infty}^{\infty} \Big[\Big[\sum_{j=1}^n \cos(sX_j) - \frac{\lambda^2}{\lambda^2+s^2}\Big]^2 + \Big[\sum_{j=1}^n \sin(sX_j) - \frac{\lambda s}{\lambda^2+s^2}\Big]^2\Big] s^\gamma e^{-as^2} \mathrm{d}s. \tag{4.68}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

The remaining $\mathcal{L}^2$-type estimators can be represented as follows:

$$\widehat{\lambda}_n^{\mathrm{LT}\,1,0} = \frac{\displaystyle\sum_{j,k=1}^{n} \frac{X_j^2 + X_k^2 - 2\gamma X_j X_k + a(X_j + X_k)}{(X_j + X_k + a)^{\gamma+2}}}{\displaystyle\sum_{j,k=1}^{n} \frac{2X_j X_k}{(X_j + X_k + a)^{\gamma+1}}}, \tag{4.69}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,2,1} = \frac{\displaystyle\sum_{j,k=1}^{n} \frac{X_j X_k [X_j^2 + X_k^2 + a(X_j + X_k) + (1-\gamma) X_j X_k]}{(X_j + X_k + a)^{\gamma+2}}}{\displaystyle\sum_{j,k=1}^{n} \frac{X_j^2 X_k^2}{(X_j + X_k + a)^{\gamma+1}}}, \tag{4.70}$$

$$\begin{aligned}\widehat{\lambda}_n^{\mathrm{LT}\,2,0} &= \underset{\lambda>0}{\operatorname{argmin}} \Bigg[\lambda^4 \sum_{j,k=1}^{n} \frac{\gamma! X_j^2 X_k^2}{(X_j + X_k + a)^{\gamma+1}} + \lambda^3 \sum_{j,k=1}^{n} \frac{4(\gamma+1)! X_j^2 X_k^2}{(X_j + X_k + a)^{\gamma+2}} \\ &+ \lambda^2 \sum_{j,k=1}^{n} \frac{6(\gamma+2)! X_j^2 X_k^2 - 2\gamma! (X_j^2 + X_k^2)(X_j + X_k + a)^2}{(X_j + X_k + a)^{\gamma+3}} \\ &+ 4\lambda \sum_{j,k=1}^{n} \frac{(\gamma+3)! X_j^2 X_k^2 - (\gamma+1)! (X_j^2 + X_k^2)(X_j + X_k + a)^2}{(X_j + X_k + a)^{\gamma+4}} \Bigg],\end{aligned} \tag{4.71}$$

all for $\gamma = 0, 1, 2, \ldots$, and

$$\widehat{\lambda}_n^{\mathrm{CF}\,1,0} = \begin{cases} \dfrac{\displaystyle\sum_{j,k=1}^{n} (X_j + X_k) \exp\Big(-(X_j - X_k)^2/4a\Big)}{\displaystyle 2\sum_{j,k=1}^{n} X_j X_k \exp\Big(-(X_j - X_k)^2/4a\Big)}, & \gamma = 0, \\[2ex] \dfrac{\displaystyle\sum_{j,k=1}^{n} (X_j + X_k) \exp\Big(-(X_j - X_k)^2/4a\Big) \Big[(X_j - X_k)^2/a - 2\Big]}{\displaystyle 2\sum_{j,k=1}^{n} X_j X_k \exp\Big(-(X_j - X_k)^2/4a\Big) \Big[(X_j - X_k)^2/a - 2\Big]}, & \gamma = 2, \end{cases} \tag{4.72}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

$$\widehat{\lambda}_n^{\mathrm{CF}\,2,1} = \begin{cases} \dfrac{\sum_{j,k=1}^{n} (X_j^2 X_k + X_j X_k^2) \exp\Big(-(X_j - X_k)^2/4a\Big)}{\sum_{j,k=1}^{n} X_j^2 X_k^2 \exp\Big(-(X_j - X_k)^2/4a\Big)}, & \gamma = 0\,, \\[2ex] \dfrac{\sum_{j,k=1}^{n} (X_j^2 X_k + X_j X_k^2) \exp\Big(-(X_j - X_k)^2/4a\Big)\Big[(X_j - X_k)^2/a - 2\Big]}{2\sum_{j,k=1}^{n} X_j^2 X_k^2 \exp\Big(-(X_j - X_k)^2/4a\Big)\Big[(X_j - X_k)^2/a - 2\Big]}, & \gamma = 2\,, \end{cases} \tag{4.73}$$

$$\begin{aligned} \widehat{\lambda}_n^{\mathrm{CF}\,2,0} &= \operatorname*{argmin}_{\lambda > 0} \Big[\lambda^4 \sum_{j,k=1}^{n} X_j^2 X_k^2 \exp\Big(-(X_j - X_k)^2/4a\Big) \\ &+ \frac{\lambda^2}{2a^2} \sum_{j,k=1}^{n} \Big[2(2a^2 + X_j^3 X_k^3) - X_j^2(2a^2 + X_k^4) - X_k^2(2a^2 + X_j^4)\Big] \exp\Big(-(X_j - X_k)^2/4a\Big) \\ &- \frac{2\lambda}{a} \sum_{j,k=1}^{n} (X_j + X_k)(X_j - X_k)^2 \exp\Big(-(X_j - X_k)^2/4a\Big)\Big], \qquad \gamma = 0\,. \end{aligned} \tag{4.74}$$

Derivations of (4.70) and (4.72) are shown in Appendix B.4.

4.5.2.2 Considered transform-matching estimators

Based on the same differential equations, we also consider the LT-based transform-matching estimators of the type (4.8) (with $p = 1$), being the solvers of

$$\widehat{L}_n(s) - \frac{\lambda}{\lambda + s} = 0\,, \tag{4.75}$$

$$(\lambda + s)\widehat{L}_n'(s) + \widehat{L}_n(s) = 0\,, \tag{4.76}$$

$$(\lambda + s)\widehat{L}_n''(s) + 2\widehat{L}_n'(s) = 0\,, \tag{4.77}$$

$$(\lambda + s)^2\widehat{L}_n''(s) - 2\widehat{L}_n(s) = 0\,, \tag{4.78}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

for $s > 0$. These are:

$$\widehat{\lambda}_n^{\mathrm{LT}\,0}(s) = \frac{s\widehat{L}_n(s)}{1-\widehat{L}_n(s)}\,, \tag{4.79}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,1,0}(s) = -\frac{\widehat{L}_n(s)}{\widehat{L}'_n(s)} - s\,, \tag{4.80}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,2,1}(s) = -\frac{2\widehat{L}'_n(s)}{\widehat{L}''_n(s)} - s\,, \tag{4.81}$$

$$\widehat{\lambda}_n^{\mathrm{LT}\,2,0}(s) = \sqrt{\frac{2\widehat{L}_n(s)}{\widehat{L}''_n(s)}} - s\,, \tag{4.82}$$

whereby $\widehat{\lambda}_n^{\mathrm{LT}\,2,0}(s)$ is the larger (and always positive) solver of the quadratic equation (4.78).

Among the CF-based estimators, we consider only the solver of

$$\widehat{\phi}_n(s) - \frac{\lambda}{\lambda - is} = 0\,, \tag{4.83}$$

which, by separation of the real and imaginary parts, is expressed as

$$\widehat{\lambda}_n^{\mathrm{CF}\,0}(s) = s\,\frac{\widehat{u}_n(s)}{\widehat{v}_n(s)}, \quad s \neq 0. \tag{4.84}$$

4.5.2.3 Asymptotic normality

We begin with the asymptotic normality of the transform-matching estimators.

Theorem 9. *The transform-matching estimators expressed by (4.79)-(4.82) and (4.84) are strongly consistent and asymptotically normal, whereby*

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,0}(s) - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \frac{\lambda(\lambda+s)^2}{\lambda+2s}\Big), \tag{4.85}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,1,0}(s) - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \frac{(\lambda+s)^4(\lambda^2+2\lambda s+2s^2)}{\lambda(\lambda+2s)^3}\Big), \tag{4.86}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,2,1}(s) - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \frac{2(\lambda+s)^6(\lambda^2+\lambda s+s^2)}{\lambda(\lambda+2s)^5}\Big), \tag{4.87}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,2,0}(s) - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \frac{(\lambda+s)^4(5\lambda^4+20\lambda^3 s+34\lambda^2 s^2+32\lambda s^3+14s^4)}{4\lambda(\lambda+2s)^5}\Big), \tag{4.88}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{CF}\,0}(s) - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0\,, \frac{\lambda^6+4\lambda^4 s^2+5\lambda^2 s^4+2s^6}{\lambda^2(\lambda^2+4s^2)}\Big). \tag{4.89}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Proof: The estimators of λ are continuous functions of the empirical transforms, which are strongly consistent non-parameteric estimators of the model transforms. Therefore, by the Continuous Mapping Theorem, the estimators of λ are also strongly consistent. Their asymptotic variances can be determined using the multivariate Delta Method, Casella and Berger (2002). To illustrate, let us consider $\widehat{\lambda}_n^{\mathrm{LT},1,0}(s)$. From the preceding chapter, we are aware that

$$\sqrt{n}\Big(\widehat{L}_n(s) - L(s)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\bigg(0, \frac{\lambda s^2}{(\lambda+s)^2(\lambda+2s)}\bigg), \tag{4.90}$$

$$\sqrt{n}\Big(\widehat{L}_n'(s) - L'(s)\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\bigg(0, \frac{\lambda(\lambda^4+2\lambda^3 s+2s^4)}{(\lambda+s)^4(\lambda+2s)^3}\bigg). \tag{4.91}$$

Let us denote the two asymptotic variances above as $\mathbb{V}_{\mathrm{as}}[\widehat{L}_n(s)]$ and $\mathbb{V}_{\mathrm{as}}[\widehat{L}_n'(s)]$, respectively. Applying the Delta Method, we can find the asymptotic variance of $\widehat{L}_n(s)/\widehat{L}_n'(s)$ as $n \to \infty$,

$$\mathbb{V}_{\mathrm{as}}\left[\frac{\widehat{L}_n(s)}{\widehat{L}_n'(s)}\right] = \left(\frac{L(s)}{L'(s)}\right)^2 \left(\frac{\mathbb{V}_{\mathrm{as}}[\widehat{L}_n(s)]}{[L(s)]^2} + \frac{\mathbb{V}_{\mathrm{as}}[\widehat{L}_n'(s)]}{[L'(s)]^2} - \frac{\mathbb{C}\mathrm{ov}[\widehat{L}_n(s), \widehat{L}_n'(s)]}{L(s)L'(s)}\right), \tag{4.92}$$

where the covariance term, computed in a similar way as (B.1), equals $L'(2s)-L(s)L'(s)$. Inserting $L(s) = \lambda/(\lambda+s)$ and $L'(s) = -\lambda/(\lambda+s)^2$ gives then

$$\mathbb{V}_{\mathrm{as}}\left[\frac{\widehat{L}_n(s)}{\widehat{L}_n'(s)}\right] = \frac{(\lambda+s)^4(\lambda^2+2\lambda s+2s^2)}{(\lambda+2s)^3}. \tag{4.93}$$

The derivations of the other asymptotic variances follow a similar procedure. The covariance term necessary for computing the variance in (4.89) is obtained using (2.37). □

The next theorem asserts the asymptotic normality of the $\mathcal{L}^2$ counterparts.

Theorem 10. *The $\mathcal{L}^2$ LT-based estimators defined in (4.59)-(4.62) are asymptotically normal, such that*

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,0} - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \sigma^2_{\mathrm{LT}\,0}\Big), \tag{4.94}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,1,0} - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \sigma^2_{\mathrm{LT}\,1,0}\Big), \tag{4.95}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,2,1} - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \sigma^2_{\mathrm{LT}\,2,1}\Big), \tag{4.96}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{LT}\,2,0} - \lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \sigma^2_{\mathrm{LT}\,2,0}\Big), \tag{4.97}$$

TU Bibliothek, Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar. The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

where

$$\sigma^2_{\mathrm{LT}\,0} = \frac{\lambda \displaystyle\int_0^\infty\!\!\int_0^\infty \frac{(st)^{\gamma+2}e^{-a(s+t)}\,\mathrm{d}s\,\mathrm{d}t}{(\lambda+s)^3\,(\lambda+t)^3\,(\lambda+s+t)}}{\left(\displaystyle\int_0^\infty \frac{s^{\gamma+2}e^{-as}}{(\lambda+s)^4}\mathrm{d}s\right)^2},$$

$$\sigma^2_{\mathrm{LT}\,1,0} = \frac{\displaystyle\int_0^\infty\!\!\int_0^\infty \frac{\lambda^2+\lambda(s+t)+2st}{(\lambda+s)^2\,(\lambda+t)^2\,(\lambda+s+t)^3}(st)^{\gamma}e^{-a(s+t)}\,\mathrm{d}s\,\mathrm{d}t}{\lambda\left(\displaystyle\int_0^\infty \frac{s^{\gamma}e^{-as}}{(\lambda+s)^4}\mathrm{d}s\right)^2},$$

$$\sigma^2_{\mathrm{LT}\,2,1} = \frac{\displaystyle\int_0^\infty\!\!\int_0^\infty \frac{2\lambda^2+\lambda(s+t)-(s^2-4st+t^2)}{(\lambda+s)^3\,(\lambda+t)^3\,(\lambda+s+t)^5}(st)^{\gamma}e^{-a(s+t)}\,\mathrm{d}s\,\mathrm{d}t}{\lambda\left(\displaystyle\int_0^\infty \frac{s^{\gamma}e^{-as}}{(\lambda+s)^6}\mathrm{d}s\right)^2},$$

$$\sigma^2_{\mathrm{LT}\,2,0} = \frac{\displaystyle\int_0^\infty\!\!\int_0^\infty \frac{6(\lambda+s)^2(\lambda+t)^2+(2st-\lambda^2)(\lambda+s+t)^2}{(\lambda+s)^3\,(\lambda+t)^3\,(\lambda+s+t)^5}(st)^{\gamma}e^{-a(s+t)}\,\mathrm{d}s\,\mathrm{d}t}{4\lambda\left(\displaystyle\int_0^\infty \frac{s^{\gamma}e^{-as}}{(\lambda+s)^4}\mathrm{d}s\right)^2}.$$

Proof. For the estimators $\widehat{\lambda}_n^{\mathrm{LT}\,1,0}$, $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}$, and $\widehat{\lambda}_n^{\mathrm{LT}\,2,0}$, the theorem follows from Theorem 1 in a scalar setting, i.e. with $\boldsymbol{\theta} = \lambda$. In particular, in Theorem 1

- for $\widehat{\lambda}_n^{\mathrm{LT}\,1,0}$ we put $r = 1$, $m = 0$ and $h(\boldsymbol{\theta}, s) = \lambda + s$;
- for $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}$ we put $r = 2$, $m = 1$ and $h(\boldsymbol{\theta}, s) = \frac{1}{2}(\lambda + s)$;
- for $\widehat{\lambda}_n^{\mathrm{LT}\,2,0}$ we put $r = 2$, $m = 0$ and $h(\boldsymbol{\theta}, s) = -\frac{1}{2}(\lambda + s)^2$.

The asymptotic normality of $\widehat{\lambda}_n^{\mathrm{LT}\,0}$ requires a separate proof. This proof is simpler as it uses only CLT along with Taylor expansion; it does not necessitate V-statistic theory, and we omit it. □

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Finally, we have the theorem concerning the asymptotic distribution of the $\mathcal{L}^2$ CF-based estimators.

Theorem 11. *The $\mathcal{L}^2$ CF-based estimators defined by (4.63)-(4.66) are asymptotically normal, whereby*

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{CF}\,0}-\lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0,\sigma^2_{\mathrm{CF}\,0}\Big), \tag{4.98}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{CF}\,1,0}-\lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0,\sigma^2_{\mathrm{CF}\,1,0}\Big), \tag{4.99}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{CF}\,2,1}-\lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0,\sigma^2_{\mathrm{CF}\,2,1}\Big), \tag{4.100}$$

$$\sqrt{n}\Big(\widehat{\lambda}_n^{\mathrm{CF}\,2,0}-\lambda\Big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0,\sigma^2_{\mathrm{CF}\,2,0}\Big), \tag{4.101}$$

where

$$\sigma^2_{\mathrm{CF}\,0} = \frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\lambda^2(\lambda^4+2\lambda^2s^2+2\lambda^2t^2+s^2+t^2)(st)^{\gamma+2}e^{-a(s^2+t^2)}\,\mathrm{d}s\,\mathrm{d}t}{(\lambda^2+s^2)^2\,(\lambda^2+t^2)^2\,(\lambda^2+(s+t)^2)\,(\lambda^2+(s-t)^2)}}{\left(\displaystyle\int_{-\infty}^{\infty}\frac{s^{\gamma+2}e^{-as^2}}{(\lambda^2+s^2)^2}\,\mathrm{d}s\right)^2},$$

$$\sigma^2_{\mathrm{CF}\,1,0} = \frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{2P_1(s,t)\,(st)^{\gamma}e^{-a(s^2+t^2)}\,\mathrm{d}s\,\mathrm{d}t}{(\lambda^2+s^2)^2\,(\lambda^2+t^2)^2\,(\lambda+(s+t)^2)^3\,(\lambda+(s-t)^2)^3}}{\left(\displaystyle\int_{0}^{\infty}\frac{s^{\gamma}e^{-as^2}}{(\lambda+s)^4}\,\mathrm{d}s\right)^2},$$

$$\sigma^2_{\mathrm{CF}\,2,1} = \frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{P_2(s,t)\,(st)^{\gamma}e^{-a(s^2+t^2)}\,\mathrm{d}s\,\mathrm{d}t}{(\lambda^2+s^2)^3\,(\lambda^2+t^2)^3\,(\lambda+(s+t)^2)^5\,(\lambda+(s-t)^2)^5}}{\left(\displaystyle\int_{0}^{\infty}\frac{s^{\gamma}e^{-as^2}}{(\lambda^2+s^2)^3}\,\mathrm{d}s\right)^2},$$

$$\sigma^2_{\mathrm{CF}\,2,0} = \frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{64P_3(s,t)\,(st)^{\gamma}e^{-a(s^2+t^2)}\,\mathrm{d}s\,\mathrm{d}t}{(\lambda^2+s^2)^3\,(\lambda^2+t^2)^3\,(\lambda^2+(s+t)^2)^5\,(\lambda^2+(s-t)^2)^5}}{\left(\displaystyle\int_{0}^{\infty}\frac{3s^2-2\lambda^2}{(\lambda^2+s^2)^3}s^{\gamma}e^{-as^2}\,\mathrm{d}s\right)^2},$$

and

$$\begin{aligned}
P_1(s,t) &= \lambda^{14}+\lambda^{12}(6s^2+6t^2)+\lambda^{10}(15s^4+31s^2t^2+15t^4)+\lambda^8(20s^6+59s^4t^2+59s^2t^4+20t^6)\\
&+ \lambda^6(15s^8+46s^6t^2+110s^4t^4+46s^2t^6+15t^8)\\
&+ \lambda^4(6s^{10}+4s^8t^2+94s^6t^4+94s^4t^6+4s^2t^8+6t^{10})\\
&+ \lambda^2(s^{12}-13s^{10}t^2+19s^8t^4+50s^6t^6+19s^4t^8-13s^2t^{10}+t^{12})\\
&- 5s^{12}t^2-9s^{10}t^4+14s^8t^6+14s^6t^8-9s^4t^{10}-5s^2t^{12},
\end{aligned}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

while $P_2(s,t)$ and $P_3(s,t)$ are even polynomials of yet higher orders.

Proof. As for the estimators $\widehat{\lambda}_n^{\mathrm{CF}\,1,0}$, $\widehat{\lambda}_n^{\mathrm{CF}\,2,1}$, and $\widehat{\lambda}_n^{\mathrm{CF}\,2,0}$, the theorem follows from Theorem 3 in the scalar setting with $\boldsymbol{\theta} = \lambda$. In particular, to obtain the limiting variances in (4.99), (4.100) and (4.101), we put respectively the following:

- for $\widehat{\lambda}_n^{\mathrm{CF}\,1,0}$: $r=1$, $m=0$ and $h(\boldsymbol{\theta}, s) = s+i\lambda$, $\mathrm{Re}h(\boldsymbol{\theta}, s)=s$, $\mathrm{Im}h(\boldsymbol{\theta}, s)=\lambda$;
- for $\widehat{\lambda}_n^{\mathrm{CF}\,2,1}$: $r=2$, $m=1$ and $h(\boldsymbol{\theta}, s) = \frac{1}{2}(s+i\lambda)$, $\mathrm{Re}h(\boldsymbol{\theta}, s) = \frac{1}{2}s$, $\mathrm{Im}h(\boldsymbol{\theta}, s) = \frac{1}{2}\lambda$;
- for $\widehat{\lambda}_n^{\mathrm{CF}\,2,0}$: $r = 2$, $m = 0$ and $h(\boldsymbol{\theta}, s) = -\frac{1}{2}(s+i\lambda)^2$, $\mathrm{Re}h(\boldsymbol{\theta}, s) = \frac{1}{2}(\lambda^2 - s^2)$, $\mathrm{Im}h(\boldsymbol{\theta}, s) = -\lambda s$.

As for the estimator $\widehat{\lambda}_n^{\mathrm{CF}\,0}$, its asymptotic normality and the variance expression follow from sec. 2 of Heathcote (1977). □

We see that the asymptotic variances of the CF-based estimators are more complicated than those of the LT-based counterparts. These expressions become less and less tractable as the order of a differential equation increases.

4.5.2.4 Asymptotic relative efficiencies

Given the asymptotic variances of our estimators, we can compute their AREs w.r.t. the ML estimator $\widehat{\lambda}_n^{\mathrm{ML}} = \overline{X}_n$, in case of which

$$\sqrt{n}\left(\widehat{\lambda}_n^{\mathrm{ML}} - \lambda\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \lambda^2\right). \tag{4.102}$$

Hence, the AREs are generally expressed by

$$\mathrm{ARE}\left[\widehat{\lambda}_n^{*}; \widehat{\lambda}_n^{\mathrm{ML}}\right] = \frac{\lambda^2}{\mathbb{V}_{\mathrm{as}}[\widehat{\lambda}_n^{*}]}, \tag{4.103}$$

where for $\mathbb{V}_{\mathrm{as}}[\widehat{\lambda}_n^{*}]$ we put the asymptotic variances in (4.85)-(4.89), (4.94)-(4.97) and (4.98)-(4.101). The results are complicated functions of the true λ and the tuning parameters s or a and γ. Despite this complexity, we can identify lines of constant AREs by analyzing their three-dimensional plots. The AREs of the transform-matching estimators (4.79)-(4.82) and (4.84) are depicted in Fig. 4.7. We observe that such lines are determined by *linear* relations between s and λ on the s-λ-plane. As for $\mathcal{L}^2$ estimators, Fig. 4.8 plots AREs for the LT-based ones (4.59)-(4.62) for $\gamma = 0, 1, 2$. Here, in turn, we see that the lines of constancy of ARE lines are determined by a *inversely proportional* relation between a and λ on the a-λ-plane, irrespectively of the value of γ. The same was found in case of the CF-based estimators whose AREs are qualitatively similar for $\gamma = 0, 2, 4$ (the plots are omitted). These observations will be used in sec. 4.5.2.6 for data-driven selection of the s- and a-values.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

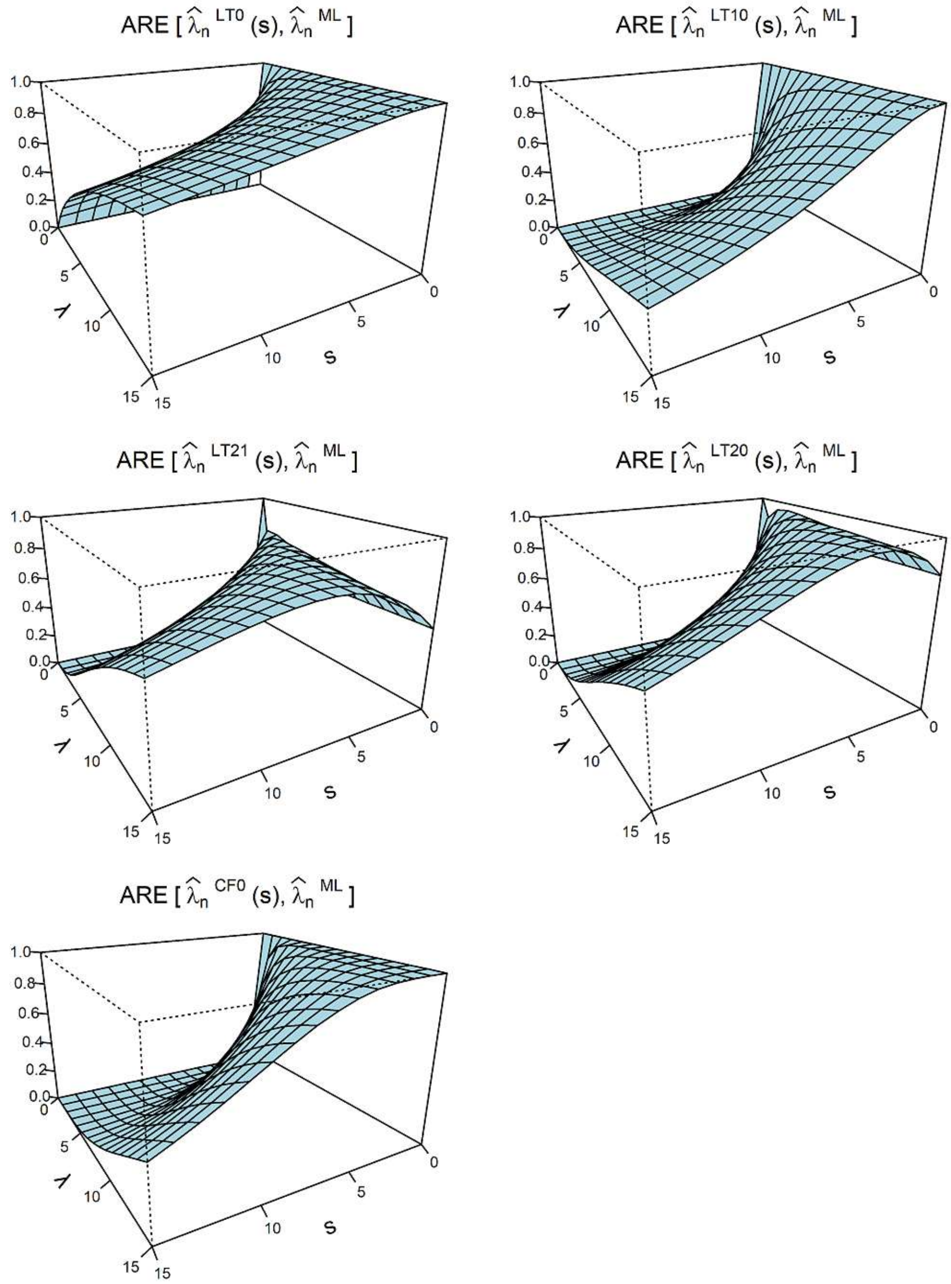


Figure 4.7: Estimation in the exponential(λ) distribution: asymptotic relative efficiencies (AREs) of the transform matching estimators (4.79)-(4.82) and (4.84) w.r.t. the ML estimator.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

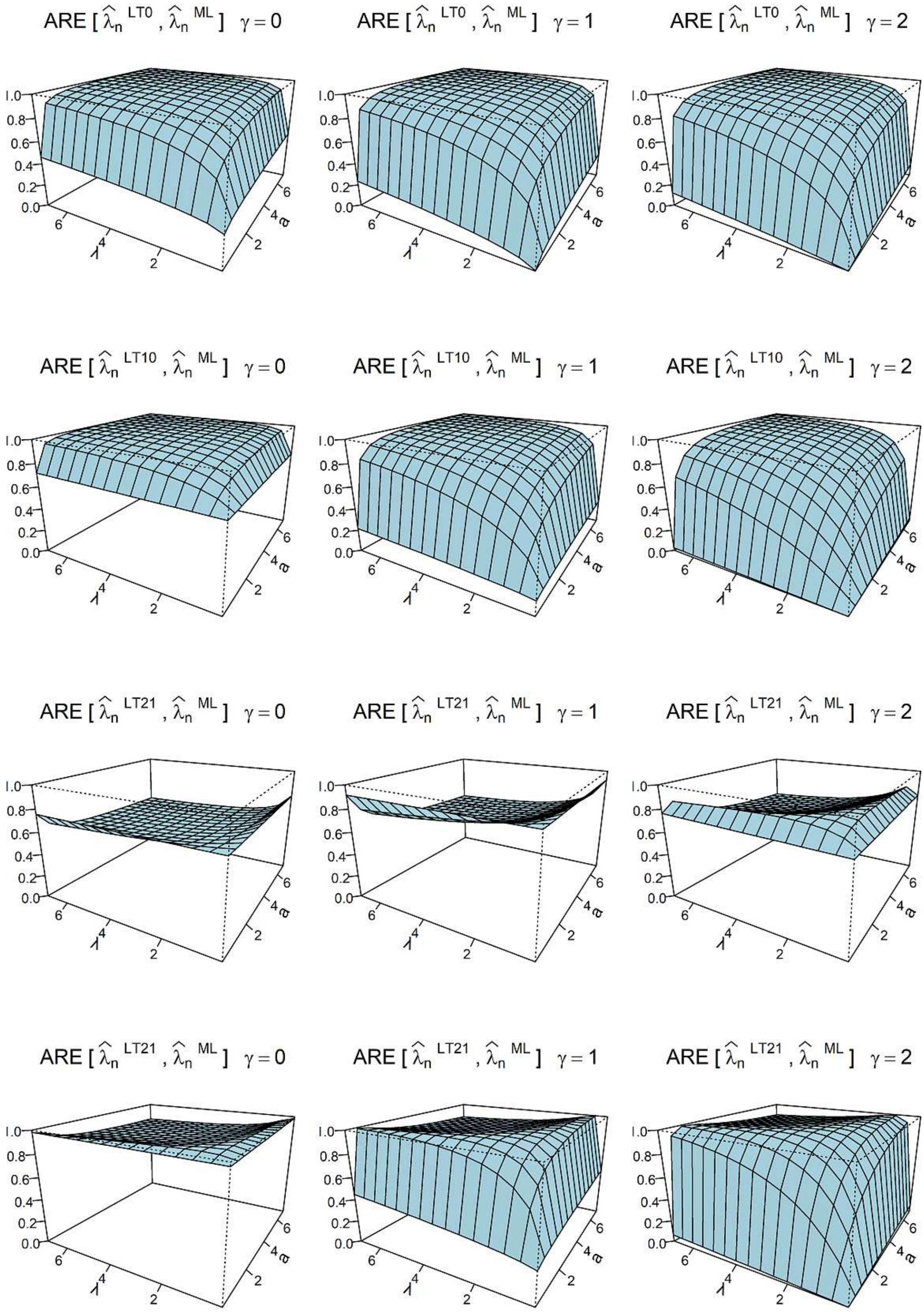


Figure 4.8: Estimation in the exponential(λ) distribution: asymptotic relative efficiencies (AREs) of the $\mathcal{L}^2$ LT-based estimators (4.59)-(4.62) w.r.t. the ML estimator.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.2.5 Influence functions

In this section, we derive and analyze the IFs of the proposed estimators of λ. The results are summarized in the three following theorems. The first one is:

Theorem 12. *The functionals corresponding to the transform-matching estimators (4.79)-(4.82) and (4.84) have the following influence functions*

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,0}(s)\Big] = \frac{\lambda+s}{s}\left[(\lambda+s)e^{-sx}-\lambda\right], \tag{4.104}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,1,0}(s)\Big] = \frac{(\lambda+s)^2}{\lambda}\left[1-(\lambda+s)x\right]e^{-sx}, \tag{4.105}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,2,1}(s)\Big] = \frac{(\lambda+s)^3}{2\lambda}\left[2-(\lambda+s)x\right]xe^{-sx}, \tag{4.106}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,2,0}(s)\Big] = \frac{(\lambda+s)^2}{4\lambda}\left[2-(\lambda+s)^2x^2\right]e^{-sx}, \tag{4.107}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{CF}\,0}(s)\Big] = \frac{\lambda^2+s^2}{\lambda s}\left[s\cos(sx)-\lambda\sin(sx)\right]. \tag{4.108}$$

Proof. These IFs are obtained from the univariate or multivariate chain rule, as given in (2.47) and (2.48). In the latter variant, we set $p=2$ and replace $T_1(F)$ and $T_2(F)$ with either $L(s)$ and $L'(s)$, or $L(s)$ and $L''(s)$, or $L'(s)$ and $L''(s)$, or $u(s)$ and $v(s)$. □

The second proposition refers to the IFs of their $\mathcal{L}^2$ counterparts.

Theorem 13. *The functionals corresponding to the $\mathcal{L}^2$ LT-based estimators (4.59)-(4.62) have the following influence functions:*

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,0}\Big] = -\frac{\displaystyle\int_0^\infty \frac{s}{(\lambda+s)^2}\left(e^{-sx}-\frac{\lambda}{\lambda+s}\right)s^\gamma e^{-as}\,\mathrm{d}s}{\displaystyle\int_0^\infty \frac{s^2}{(\lambda+s)^4}s^\gamma e^{-as}\,\mathrm{d}s}, \tag{4.109}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,1,0}\Big] = -\frac{\displaystyle\int_0^\infty \frac{\lambda}{\lambda+s}e^{-sx}\left(x-\frac{1}{\lambda+s}\right)s^\gamma e^{-as}\,\mathrm{d}s}{\displaystyle\int_0^\infty \frac{\lambda^2}{(\lambda+s)^4}s^\gamma e^{-as}\,\mathrm{d}s}, \tag{4.110}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,2,1}\Big] = -\frac{\displaystyle\int_0^\infty \frac{\lambda}{\lambda+s}xe^{-sx}\left(\frac{x}{2}-\frac{1}{\lambda+s}\right)s^\gamma e^{-as}\,\mathrm{d}s}{\displaystyle\int_0^\infty \frac{\lambda^2}{(\lambda+s)^6}s^\gamma e^{-as}\,\mathrm{d}s}, \tag{4.111}$$

$$\mathrm{IF}\Big[x;\lambda^{\mathrm{LT}\,2,0}\Big] = -\frac{\displaystyle\int_0^\infty e^{-sx}\left(\lambda x^2-\frac{2\lambda}{(\lambda+s)^2}\right)s^\gamma e^{-as}\,\mathrm{d}s}{\displaystyle\int_0^\infty \frac{4\lambda^2}{(\lambda+s)^4}s^\gamma e^{-as}\,\mathrm{d}s}. \tag{4.112}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Proof. IFs (4.110)-(4.112) follow from Theorem 4, where the values of r, m and $h(\boldsymbol{\theta}, s)$ are set as in the proof of Theorem 10. The IF (4.109) requires a separate derivation which is similar. □

Finally, we present the statement concerning the $\mathcal{L}^2$ counterparts based on CF.

Theorem 14. *The functionals corresponding to the $\mathcal{L}^2$ CF-based estimators (4.63)-(4.66) have the following influence functions:*

$$\mathrm{IF}\Big[x;\lambda^{\,\mathrm{CF}\,0}\Big] = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{s[(s^2-\lambda^2)\sin(sx) - s\lambda(1-2\cos(sx))]}{(\lambda^2+s^2)^2} s^\gamma e^{-as^2}\,\mathrm{d}s}{\displaystyle\int_{-\infty}^{\infty} \frac{s^2}{(\lambda^2+s^2)^2} s^\gamma e^{-as^2}\mathrm{d}s}, \tag{4.113}$$

$$\mathrm{IF}\Big[x;\lambda^{\,\mathrm{CF}\,1,0}\Big] = \frac{\displaystyle\lambda\int_{-\infty}^{\infty} \frac{s[2\lambda-(3\lambda^2-s^2)x]\sin(sx) + [\lambda^2-s^2-\lambda(\lambda^2-3s^2)x]\cos(sx)}{(\lambda^2+s^2)^2} s^\gamma e^{-as^2}\mathrm{d}s}{\displaystyle\int_{-\infty}^{\infty} \frac{\lambda^2}{(\lambda^2+s^2)^2} s^\gamma e^{-as^2}\mathrm{d}s}, \tag{4.114}$$

$$\mathrm{IF}\Big[x;\lambda^{\,\mathrm{CF}\,2,1}\Big] = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{S_{2,1}(s,\lambda)\sin(sx) + C_{2,1}(s,\lambda)\cos(sx)}{(\lambda^2+s^2)^2} s^\gamma e^{-as^2}\,\mathrm{d}s}{\displaystyle\int_{-\infty}^{\infty} \frac{\lambda^2}{(\lambda^2+s^2)^3} s^\gamma e^{-as^2}\,\mathrm{d}s}, \tag{4.115}$$

$$\mathrm{IF}\Big[x;\lambda^{\,\mathrm{CF}\,2,0}\Big] = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{\lambda\Big[S_{2,0}(s,\lambda)\sin(sx) + C_{2,0}(s,\lambda)\cos(sx) + 8\lambda^3 s^2\Big]}{(\lambda^2+s^2)^3} s^\gamma e^{-as^2}\,\mathrm{d}s}{\displaystyle\int_{-\infty}^{\infty} \frac{4\lambda^4}{(\lambda^2+s^2)^3} s^\gamma e^{-as^2}\,\mathrm{d}s}, \tag{4.116}$$

where

$$\begin{aligned}
S_{2,1}(s,\lambda) &= [3\lambda^2 - s^2 - (2\lambda^2 - 2s^2)\lambda x]xs\lambda\,,\\
C_{2,1}(s,\lambda) &= \tfrac{1}{2}[2\lambda^3 - 6\lambda s^2 - (\lambda^4 - 6\lambda^2 s^2 + s^4)x]x\lambda\,,\\
S_{2,0}(s,\lambda) &= 4s\lambda[(3s^2x^2+1)\lambda^2 + s^2 - \lambda^4 x^2]\,,\\
C_{2,0}(s,\lambda) &= (9s^2x^2+2)\lambda^4 - (\lambda^6 + 7\lambda^2 s^4 + s^6)x^2 - 2s^4\,.
\end{aligned}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Proof. The IFs (4.114)-(4.116) result from Theorem 6, wherein the values of r, m and $h(\boldsymbol{\theta}, s)$ are set as in the proof of Theorem 11. The IF (4.113) is derived using the theorem given in sec. 3 of Campbell (1993). Note that with $\gamma = 0, 2$ and 4, all the integrands are even functions of s and this fact saves time when integrating numerically. □

Figs. 4.9, 4.10 and 4.11 display the IFs from Theorems 12, 13 and 14 for $\lambda = 2$ and various settings of s, a and γ. The IF of the ML estimator is plotted on each graph as a reference.

We begin by analyzing the IFs of the transform-matching estimators shown in Fig. 4.9. Note that IF of $\widehat{\lambda}_n^{\mathrm{CF}\,0}(s)$ is bounded in x but oscillating. Such IF shape is not desirable, implying that a large outlier may exert a greater impact on the estimator than a smaller one. As for the LT-based estimators of this group, we can easily compute the following limits:

$$\begin{aligned}
\lim_{x\to 0} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,0}(s)\Big] &= \lambda + s, & \lim_{x\to \infty} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,0}(s)\Big] &= -\frac{\lambda(\lambda+s)}{s},\\
\lim_{x\to 0} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,1,0}(s)\Big] &= \frac{(\lambda+s)^2}{\lambda}, & \lim_{x\to \infty} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,1,0}(s)\Big] &= 0,\\
\lim_{x\to 0} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,2,1}(s)\Big] &= 0, & \lim_{x\to \infty} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,2,1}(s)\Big] &= 0,\\
\lim_{x\to 0} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,2,0}(s)\Big] &= \frac{(\lambda+s)^2}{2\lambda}, & \lim_{x\to \infty} \mathrm{IF}\Big[x; \lambda^{\mathrm{LT}\,2,0}(s)\Big] &= 0.
\end{aligned}$$

Upon examining them as $x \to \infty$, we observe the advantages of constructing estimators based on differential equations of the transform. While the IF of the estimator based on LT alone, that is $\lambda^{\mathrm{LT}\,0}(s)$, is only bounded in this direction, the IFs of the three estimators based on differential equations of LT asymptotically tend to zero. Examining the limits as $x \to 0$, we also observe the advantage of employing equations of order higher than one. Specifically, the limiting value of the IF for $\lambda^{\mathrm{LT},2,0}(s)$ is twice as small as that for $\lambda^{\mathrm{LT},1,0}(s)$, and the IF of $\lambda^{\mathrm{LT},2,1}(s)$ exhibits an 'ideal' shape, rapidly approaching zero in both limits. It is important to note that, of course, the rates of this decay depend on the user-chosen parameter s and the true value of λ.

Now, let us inspect the IFs of the $\mathcal{L}^2$ estimators in Figs. 4.10 and 4.11. The above described effect of the order of equation carries over here. That is, estimators based on LT and CF alone have IFs only bounded as $x \to \infty$, while those based on differential equations manifest IFs descending to zero in this limit. And, similarly, equations involving the first and second derivatives of LT or CF yield estimators with zero IFs as $x \to 0$

Regarding the choices of a and γ, note that their different settings result in more or less robust estimates. Generally, selecting a smaller a causes all the IFs to stabilize more quickly as $x \to \infty$. The impact of γ is more nuanced. For the LT-based estimators, increasing γ (for a fixed a) accelerates the drop of IFs to zero as $x \to \infty$. However, this increase also enlarges the IFs of $\lambda^{\mathrm{LT}\,1,0}$ and $\lambda^{\mathrm{LT}\,2,1}$ as $x \to 0$. In the case of CF-based estimators, a larger γ does not result in more favorable IFs as $x \to \infty$ (they oscillate more rapidly before stabilizing). However, a larger γ does not degrade IFs as $x \to 0$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

In view of the IF plots and computational advantages, some of the proposed estimators are clearly preferred. For further study we select:

- the transform-matching LT-based estimators $\widehat{\lambda}_n^{\mathrm{LT}\,0}(s)$, $\widehat{\lambda}_n^{\mathrm{LT}\,1,0}(s)$, $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}(s)$ and $\widehat{\lambda}_n^{\mathrm{LT}\,2,0}(s)$;
- the $\mathcal{L}^2$ LT-based estimators $\widehat{\lambda}_n^{\mathrm{LT}\,1,0}$ with $\gamma = 1$, $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}$ with $\gamma = 2$, $\widehat{\lambda}_n^{\mathrm{LT}\,2,0}$ with $\gamma = 1$, and the $\mathcal{L}^2$ CF-based estimator $\widehat{\lambda}_n^{\mathrm{CF}\,1,0}$ with $\gamma = 0$.

4.5.2.6 Robustness-efficiency trade-offs

We now combine our findings of secs. 4.5.2.4 and 4.5.2.5 to propose a data-driven choice of s or a, providing a compromise between efficiency and robustness. From Figs. 4.7 and 4.8 we know that to keep the ARE approximately constant over the space of λ, we should set $s = k\widehat{\lambda}_n$ or $a = k/\widehat{\lambda}_n$, where $\widehat{\lambda}_n$ is some auxiliary, easily computable, estimator of λ. Suitable candidates are the ML estimator $\widehat{\lambda}_n^{\mathrm{ML}} = 1/\overline{X}_n$ and the median-based estimator $\widehat{\lambda}_n^{\mathrm{med}} = \log(2)/\widehat{\mathrm{med}}_n$; see Staudte and Sheather (1990). The constant k gives control over the robustness-efficiency trade-off. To ensure good trade-offs, based on the above ARE and IFs plots, along with some initial simulations, we recommend using: $k = 0.3$ for $\widehat{\lambda}_n^{\mathrm{LT}\,0}(s)$, $k = 0.3$ for $\widehat{\lambda}_n^{\mathrm{LT}\,1,0}(s)$, $k = 0.7$ for $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}(s)$, $k = 0.5$ for $\widehat{\lambda}_n^{\mathrm{LT}\,2,0}(s)$, $k = 1$ for $\widehat{\lambda}_n^{\mathrm{LT}\,1,0}$, $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}$, $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}$ and $\widehat{\lambda}_n^{\mathrm{CF}\,1,0}$.

4.5.2.7 Finite-sample performance: simulation analysis

The above asymptotic considerations are now validated by finite-sample simulations. We present results only for the median-driven choices of s and a, which provide slightly more robust estimates. The proposed estimators are compared with the sample mean- and median-based estimators defined above. Fig. 4.12 displays box plots of estimates drawn from 5000 simulated samples of size $n = 50$, considering $\lambda = 0.01$ and $\lambda = 100$. For each estimator, the left box plot pertains to estimation from pure sample. The middle and right box plots depict estimation in presence of one large and two small outliers, respectively. The large outlier is generated by $X_0 = \overline{X}_n + 10/\lambda$ (replacing X_1 in the original sample). Meanwhile, the two small outliers are $X_{0\text{-}1} = \overline{X}_n/100$ and $X_{0\text{-}2} = \overline{X}_n/1000$ (replacing X_1 and X_2).

We make here the following observations. Firstly, the proposed data-driven choice of s and a adapts all the estimators to changes in the scale of data. Namely, when λ changes by factor 10^5, the box plots stretch accordingly. Secondly, all the transform-based estimators compare well with the ML estimator in pure samples. In the presence of outliers, our estimators can be more biased than the median-based estimator $\widehat{\lambda}_n^{\mathrm{med}}$, but, simultaneously, they have greater precision. The $\mathcal{L}^2$ estimator $\widehat{\lambda}_n^{\mathrm{LT},1,0}$ stands out as an exception, being highly sensitive to small outliers. The estimators based on the second-order equations of LT are more robust against small outliers, consistently with expectations from their IFs. It is of interest that the transform-matching estimators tend to be more reliable then their $\mathcal{L}^2$ counterparts. The estimator $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}(s)$ emerges as a clear winner in this comparison, remaining nearly unbiased for the two types of outliers.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

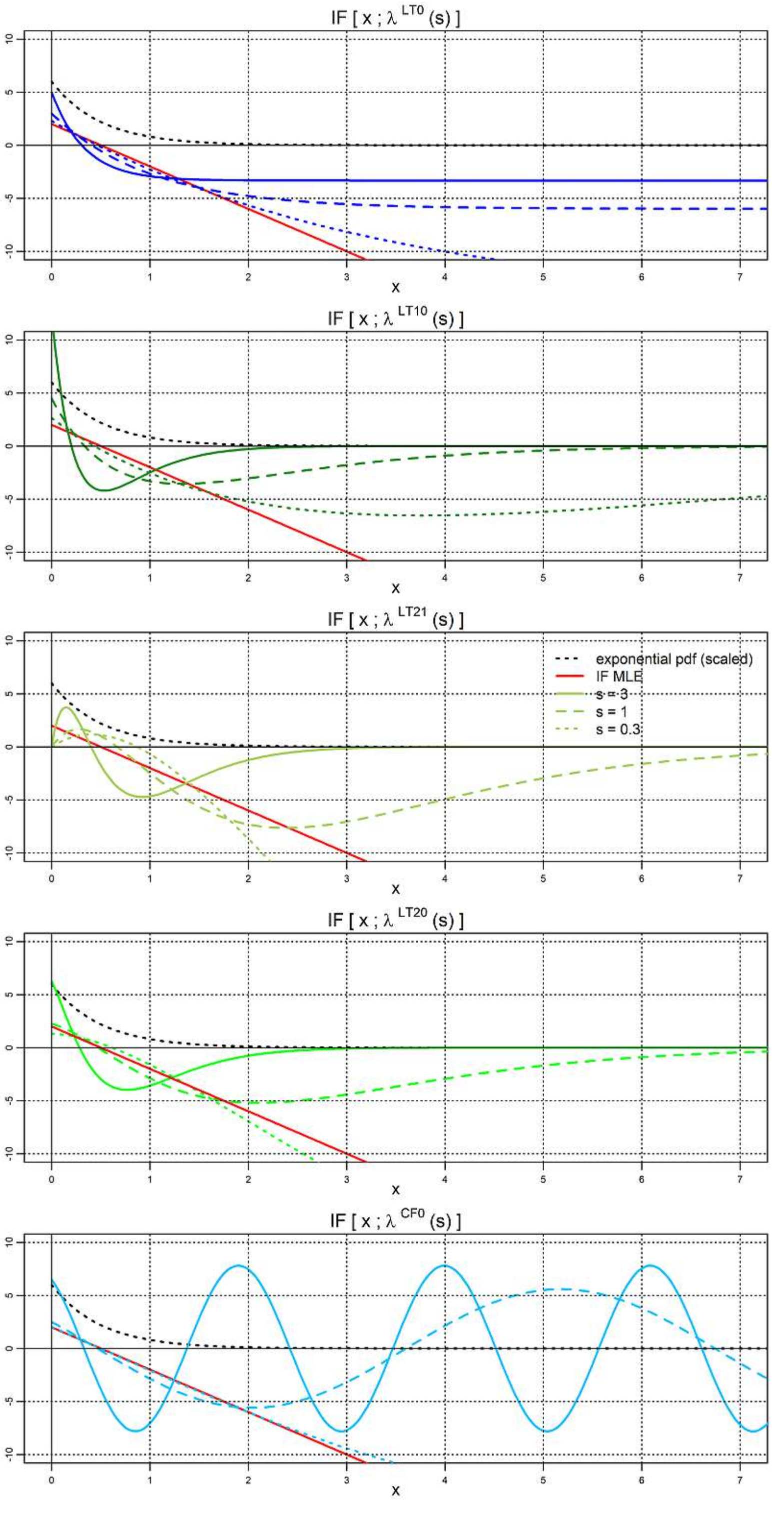


Figure 4.9: IFs of the transform-matching estimators of λ in the exponential(λ) distribution from Theorem 12, for $\lambda = 2$ and $s = 3, 1$ and 0.3.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

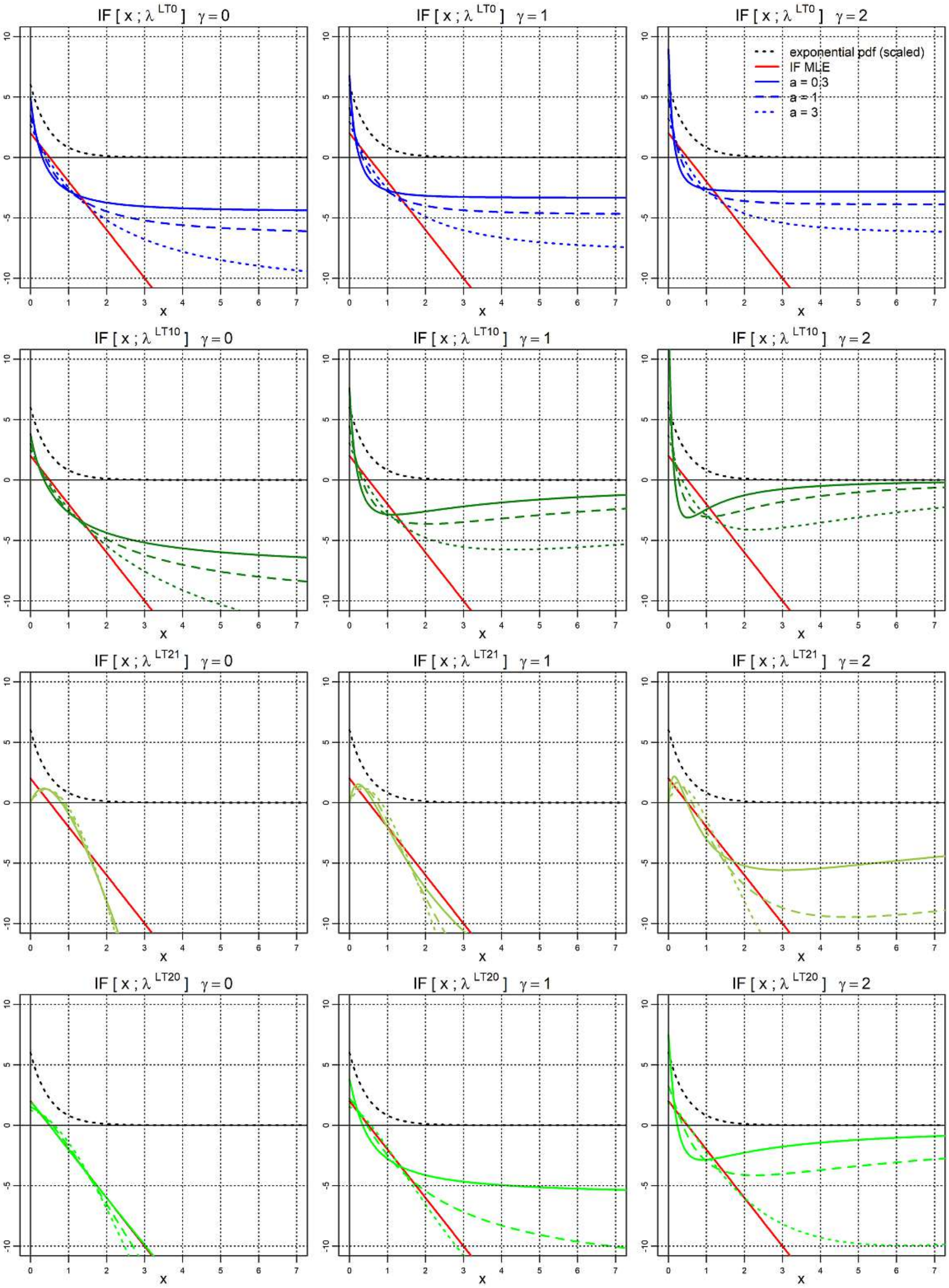


Figure 4.10: IFs of the $\mathcal{L}^2$ LT-based estimators of λ in the exponential(λ) distribution from Theorem 13, for $\lambda = 2$ and combinations of $\gamma = 0, 1, 2$ with $a = 0.3, 1, 3$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

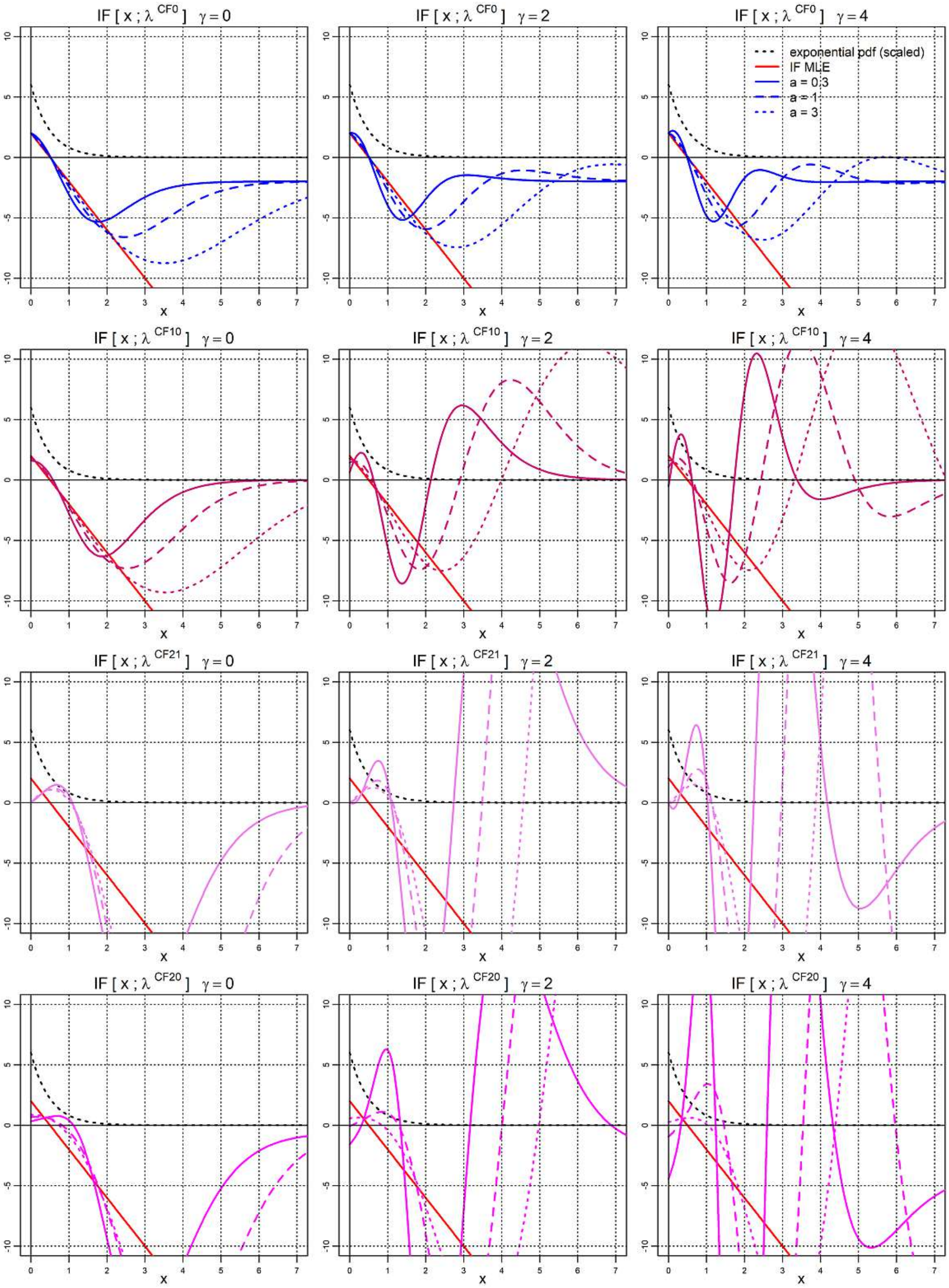


Figure 4.11: IFs of the $\mathcal{L}^2$ CF-based estimators of λ in the exponential(λ) distribution from Theorem 14, for $\lambda = 2$ and combinations of $\gamma = 0, 2, 4$ with $a = 0.3, 1, 3$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

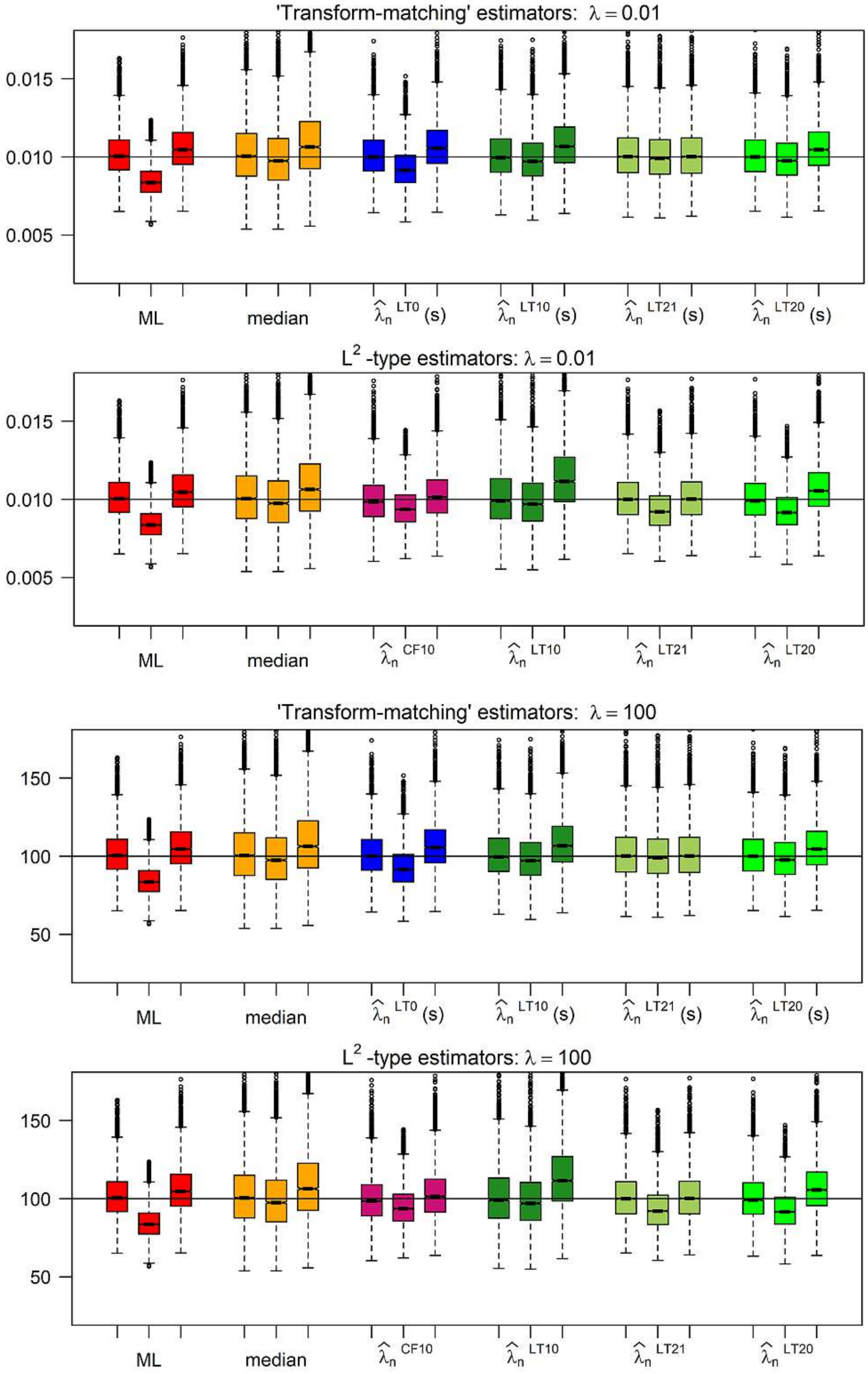


Figure 4.12: Estimation of λ in the exponential(λ) model for $\lambda = 0.01$ (the two upper panels) and $\lambda = 100$ (the two lower panels) under mild contamination; $n = 50$, 5000 replications. Comparisons of the ML, median-based and selected transform-based estimators. Left box plot: pure sample; middle box plot: sample with one large outlier; right box plot: sample with two small outliers.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

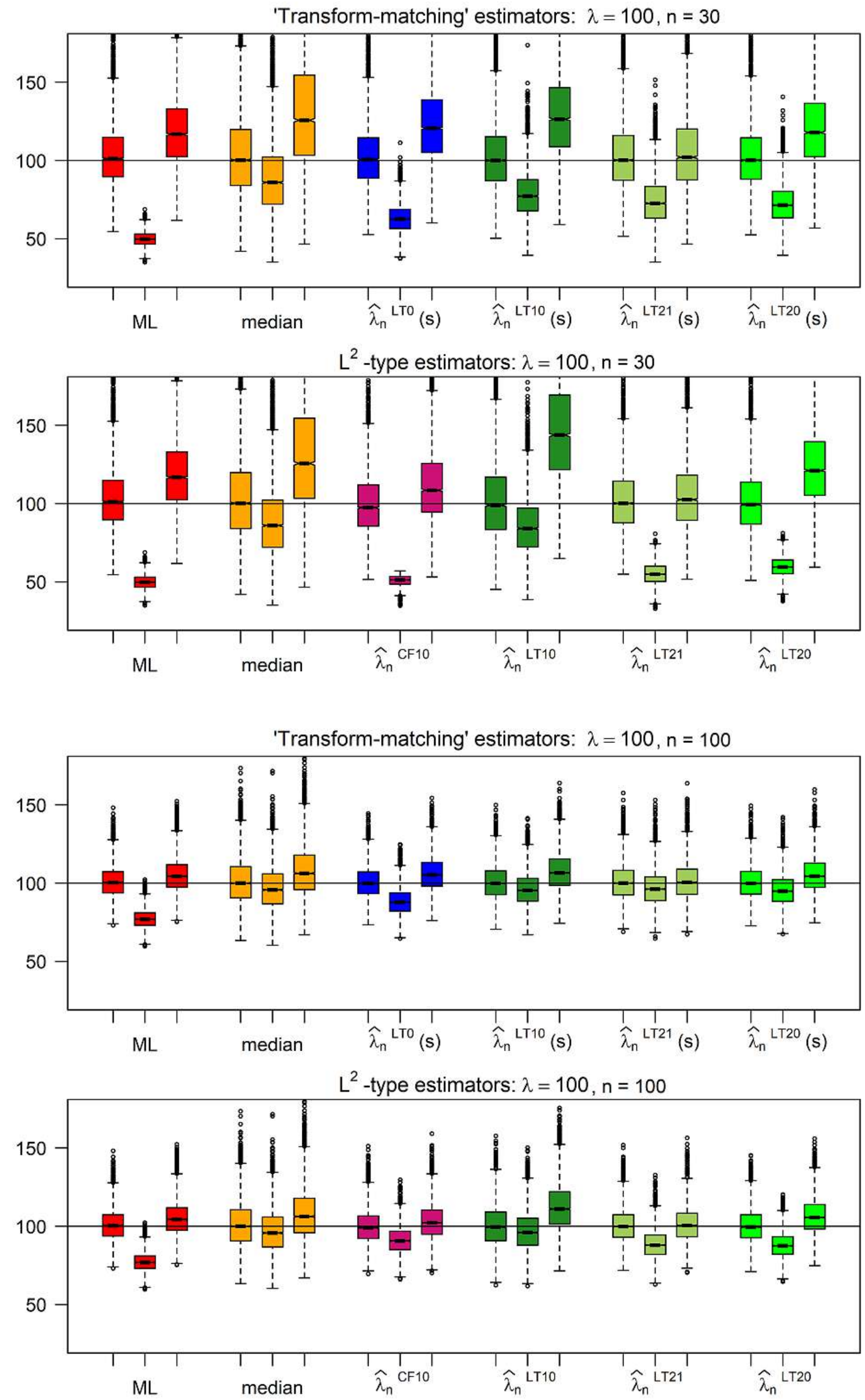


Figure 4.13: Estimation of λ in the exponential(λ) model under strong contamination; $\lambda = 100$ for $n = 30$ (the two upper panels) and $n = 100$ (the two lower panels); 5000 replications. Comparisons of the ML, median-based and selected transform-based estimators. Left box plot: pure sample; middle box plot: sample with three large outlier; right box plot: sample with four small outliers.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.3 Rayleigh distribution

We now deal with a distribution that lacks closed-form expression for the transforms (in terms of elementary functions) but for which estimation based on the differential equations satisfied by the transforms is a viable alternative.

The Rayleigh cumulative distribution and its density function are given by:

$$F_{\boldsymbol{\theta}}(x) = 1 - \exp(-x^2/2\vartheta^2)\,, \tag{4.117}$$

$$f_{\boldsymbol{\theta}}(x) = \frac{x}{\vartheta^2}\exp(-x^2/2\vartheta^2)\,, \tag{4.118}$$

$x \geq 0$ and $\vartheta^2 > 0$. It is a unimodal scale family with $\mathbb{E}[X]=\vartheta\sqrt{\pi/2}$, $\mathbb{V}[X]=\vartheta^2(4-\pi)/2$, $\mathrm{Med}[X]=\vartheta\sqrt{2\log 2}$ and the mode located at $x=\vartheta$.

In fact, the corresponding LT is expressible as

$$L_{\boldsymbol{\theta}}(s) = L_\vartheta(s) = \int_0^\infty e^{-sx}\mathrm{d}F(x) = 1 - s\vartheta\sqrt{\frac{\pi}{2}}\exp\left(\frac{\vartheta^2 s^2}{2}\right)\mathrm{erfc}\left(\frac{\vartheta s\sqrt{2}}{2}\right), \tag{4.119}$$

$s \in \mathbb{R}$, where $\mathrm{erfc}(\cdot)$ is the complementary error function; see Meintanis and Iliopoulos (2003b). But this representation requires a fully numerical approach if we would like to estimate the parameter directly from it.

On the other hand, by the method described in sec. 3.2, we can show that

$$L''_{\boldsymbol{\theta}}(s) - \vartheta^2 s L'_{\boldsymbol{\theta}}(s) - 2\vartheta^2 L_{\boldsymbol{\theta}}(s) = 0\,, \tag{4.120}$$

and

$$L'''_{\boldsymbol{\theta}}(s) - \vartheta^2 s L''_{\boldsymbol{\theta}}(s) - 3\vartheta^2 L'_{\boldsymbol{\theta}}(s) = 0\,. \tag{4.121}$$

The next sections analyze two estimators of $\boldsymbol{\theta} = \vartheta^2$ based on (4.120) and two based on (4.121). The parameter-uniqueness within the model, ensuring consistency of the estimators, has already been verified in sec. 3.4.2.

In this example, we do not consider the corresponding CF-based estimators. Instead, we carry out comparisons with a greater number of traditional robust and non-robust estimators.

4.5.3.1 Considered estimators

The empirical versions of (4.120) and (4.121) yield the following transform-matching estimators:

$$\widehat{\vartheta^2_n}^{\,\mathrm{LT}\,2,1,0}(s) = \frac{\widehat{L}''_n(s)}{2\widehat{L}_n(s) + s\widehat{L}'_n(s)}\,, \tag{4.122}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

and

$$\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,3,2,1}(s) = \frac{\widehat{L}_n'''(s)}{3\widehat{L}_n'(s) + s\widehat{L}_n''(s)}\,. \tag{4.123}$$

In turn, the corresponding $\mathcal{L}^2$ estimators are given by

$$\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,2,1,0} = \underset{\vartheta^2>0}{\operatorname{argmin}} \int_0^\infty \left[\widehat{L}_n''(s) - \vartheta^2 s\widehat{L}_n'(s) - 2\vartheta^2\widehat{L}_n(s)\right]^2 s^\gamma e^{-as}\mathrm{d}s\,, \tag{4.124}$$

$$\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,3,2,1} = \underset{\vartheta^2>0}{\operatorname{argmin}} \int_0^\infty \left[\widehat{L}_n'''(s) - \vartheta^2 s\widehat{L}_n''(s) - 3\vartheta^2\widehat{L}_n'(s)\right]^2 s^\gamma e^{-as}\mathrm{d}s\,. \tag{4.125}$$

and, for any $\gamma = 0, 1, 2, ...$, they admit the explicit expressions

$$\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,2,1,0} = \frac{\displaystyle\sum_{j,\,k=1}^{n} \frac{2(X_j^3+X_k^3) + 2a(X_j^2+X_k^2) + (1-\gamma)X_jX_k(X_j+X_k)}{(X_j+X_k+a)^{\gamma+2}}}{\displaystyle 2\sum_{j,\,k=1}^{n} \frac{(\gamma^2-\gamma+6)X_jX_k + 2(1-\gamma)(X_j^2+X_k^2) + 2a(3-\gamma)(X_j+X_k) + 4a}{(X_j+X_k+a)^{\gamma+3}}},$$

$$\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,3,2,1} = \frac{\displaystyle\sum_{j,\,k=1}^{n} \frac{3(X_j^2+X_k^2)(X_j+X_k+a) - (\gamma+1)X_jX_k(X_j+X_k)}{(X_j+X_k+a)^{\gamma+2}/X_jX_k}}{\displaystyle 2\sum_{j,\,k=1}^{n} \frac{(\gamma^2-3\gamma+14)X_jX_k + 3(2-\gamma)(X_j^2+X_k^2) + 3a(5-\gamma)(X_j+X_k) + 9a^2}{(X_j+X_k+a)^{\gamma+3}/X_jX_k}}.$$

Derivations of the two expressions are tedious but similar to those for the exponential distribution, sketched in Appendix B.4.

4.5.3.2 Asymptotic normality

Asymptotic normality of the estimators (4.122) and (4.123) follows from the multivariate Delta Method. It is addressed in the following theorem:

Theorem 15. *The transform-matching estimators (4.122) and (4.123) are strongly consistent and asymptotically normal, whereby*

$$\sqrt{n}\left(\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,2,1,0}(s) - \vartheta^2\right) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \sigma^{2\,\mathrm{LT}\,2,1,0}(s)\Big), \tag{4.126}$$

$$\sqrt{n}\left(\widehat{\vartheta_n^2}^{\,\mathrm{LT}\,3,2,1}(s) - \vartheta^2\right) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \sigma^{2\,\mathrm{LT}\,3,2,1}(s)\Big), \tag{4.127}$$

where

$$\sigma^{2\,\mathrm{LT}\,2,1,0}(s) = \vartheta^4\left[\frac{L_\vartheta^{(4)}(2s)-[L''(s)]^2}{[L_\vartheta''(s)]^2} - \frac{4L_\vartheta''(2s)+2sL_\vartheta'''(2s) - 2L_\vartheta''(s)[2L_\vartheta(s)+sL_\vartheta'(s)]}{L_\vartheta''(s)[2L_\vartheta(s) + sL_\vartheta'(s)]} + \frac{4L_\vartheta(2s)+4sL_\vartheta'(2s)+s^2L_\vartheta''(2s) - [2L_\vartheta(s)+sL_\vartheta'(s)]^2}{[2L_\vartheta(s) + sL_\vartheta'(s)]^2}\right],$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

and

$$\sigma^{2\,\mathrm{LT}\,3,2,1}(s) = \vartheta^4\Big[\frac{L_\vartheta^{(6)}(2s)-[L_\vartheta'''(s)]^2}{[L_\vartheta'''(s)]^2} + \frac{9L_\vartheta''(2s)+6sL_\vartheta'''(2s)+s^2L_\vartheta^{(4)}(2s)-[3L_\vartheta'(s)+sL_\vartheta''(s)]^2}{[3L_\vartheta'(s)+sL_\vartheta''(s)]^2} - \frac{6L_\vartheta^{(4)}(2s)+2sL_\vartheta^{(5)}(2s)-2L_\vartheta'''(s)[3L_\vartheta'(s)+sL_\vartheta''(s)]}{L_\vartheta'''(s)[3L_\vartheta'(s)+sL_\vartheta''(s)]}\Big].$$

The asymptotic variances require derivatives of LT up to the sixth order. Since $L_\vartheta(s)$ is numerically available from (4.119), we can calculate the required derivatives recursively. By routine calculations, we obtain

$$\begin{aligned}
L_\vartheta'(s) &= \vartheta^2 s-(1+\vartheta^2s^2)K_\vartheta(s),\\
L_\vartheta''(s) &= \vartheta^2(2+s^2\vartheta^2)-\vartheta^2 s(3+\vartheta^2s^2)K_0(s),\\
L_\vartheta'''(s) &= \vartheta^4 s(5+s^2\vartheta^2)-\vartheta^2(3+6\vartheta^2s^2+\vartheta^4s^4)K_\vartheta(s),\\
L_\vartheta^{(4)}(s) &= \vartheta^2(8\vartheta^2+9\vartheta^4s^2+\vartheta^6s^4)-\vartheta^2(15\vartheta^2s+10\vartheta^4s^3+\vartheta^6s^5)K_\vartheta(s),\\
L_\vartheta^{(5)}(s) &= \vartheta^6 s(14\vartheta^2s^2+\vartheta^4s^4+33)-\vartheta^4(15+45\vartheta^2s^2+15\vartheta^4s^4+\vartheta^6s^6)K_\vartheta(s),\\
L_\vartheta^{(6)}(s) &= \vartheta^6(87\vartheta^2s^2+20\vartheta^4s^4+\vartheta^6s^6+48)-\vartheta^6 s(105+105\vartheta^2s^2+21\vartheta^4s^4+\vartheta^6s^6)K_\vartheta(s),
\end{aligned}$$

where $K_\vartheta(s) = (1-L_\vartheta(s))/s\vartheta$.

Asymptotic normality of the $\mathcal{L}^2$ estimators (4.124) and (4.125) is asserted by Theorem 2 with $p = 1$. However, the resulting asymptotic variances are overly complicated. It is easier to compute them numerically from the associated IFs derived below.

4.5.3.3 Influence functions

The first theorem concerning the abovementioned IFs is as follows:

Theorem 16. *The influence functions of the functionals corresponding to the transform-matching estimators (4.122) and (4.123) are given by:*

$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{LT}\,2,1,0}(s)\Big] = \frac{-L_\vartheta''(s)[(2-sx)e^{-sx}-2L(s)-sL'(s)][2L_\vartheta(s)+sL_\vartheta'(s)]+x^2e^{-sx}-L''(s)}{[2L_\vartheta(s)+sL_\vartheta'(s)]^2}, \tag{4.128}$$

$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{LT}\,3,2,1}(s)\Big] = \frac{L_\vartheta'''(s)[3L_\vartheta'(s)+sL_\vartheta''(s)][(3-sx)xe^{-sx}+3L'(s)+sL''(s)]-x^3e^{-sx}-L'''(s)}{[3L_\vartheta'(s)+sL_\vartheta''(s)]^2}. \tag{4.129}$$

Proof. These IFs are obtained from the multivariate chain rule (2.48) with $p = 3$. There, for $T_1(F)$, $T_2(F)$ and $T_3(F)$ we have to put $L(s)$, $L'(s)$ and $L''(s)$ in the case of the estimator (4.122), or $L'(s)$, $L''(s)$ and $L'''(s)$ in the case of the estimator (4.123). □

The second theorem, concerning the IFs of the $\mathcal{L}^2$ counterparts, is the following:

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Theorem 17. *The influence functions of the functionals corresponding to the $\mathcal{L}^2$ estimators (4.124) and (4.125) are given by:*

$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{LT}\,2,1,0}\Big] = \frac{\displaystyle\int_0^\infty \left[\begin{array}{c}\Big[2\vartheta^2 L_0(s) - L_\vartheta''(s)\Big](2-sx)e^{-sx} \\ -L_0(s)\Big[x^2e^{-sx} - 2L_\vartheta''(s) + 2\vartheta^2 L_0(s)\Big]\end{array}\right] s^\gamma e^{-as}\mathrm{d}s}{\displaystyle\int_0^\infty L_0^2(s)s^\gamma e^{-as}\mathrm{d}s}, \quad (4.130)$$

where $L_0(s) = 2L_\vartheta(s) + sL'_\vartheta(s)$*, and by*

$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{LT}\,2,1,0}\Big] = \frac{\displaystyle\int_0^\infty \left[\begin{array}{c}\Big[2\vartheta^2 L_0(s) - L_\vartheta'''(s)\Big](sx-3)xe^{-sx} \\ +L_0(s)\Big[x^3e^{-sx} + 2L_\vartheta'''(s) - 2\vartheta^2 L_0(s)\Big]\end{array}\right] s^\gamma e^{-as}\mathrm{d}s}{\displaystyle\int_0^\infty L_0^2(s)s^\gamma e^{-as}\mathrm{d}s}, \quad (4.131)$$

where $L_0(s) = 3L'_\vartheta(s) + sL''_\vartheta(s)$*.*

Proof. We apply Theorem 5. For the estimator (4.124), we have to put $r = 2$, $m = 1$, $\ell = 0$ with $h(\boldsymbol{\theta}, s) = -1/2\vartheta^2$, $g(\boldsymbol{\theta}, s) = s/2$, while for the estimator (4.125) $r = 3$, $m = 2$, $\ell = 1$ with $h(\boldsymbol{\theta}, s) = -1/3\vartheta^2$, $g(\boldsymbol{\theta}, s) = s/3$. Routine calculations and rearrangements lead then to (4.130) and (4.131). □

The IFs from Theorems 16 and 17 are visualized in Figs. 4.14 and 4.15 for $\vartheta = 1$. Their shapes resemble those of the LT-based estimators of the parameter λ in the exponential distribution. Namely, smaller s in (4.122) and (4.123), as well as larger γ but smaller a in (4.124) and (4.125), cause all the IFs to drop more rapidly to zero as $x \to \infty$. Conversely, these settings lead to less favorable IFs when $x \to 0$ for the estimators involving the zero-order derivative, i.e., (4.122) and (4.124). The estimators based on third-order equation that do not employ the zero-order derivative, namely (4.123) and (4.125), are the most attractive in terms of infinitesimal robustness: their IFs drop to zero as $x \to 0$ for any s, or γ and a.

4.5.3.4 Alternative estimators

The four new estimators will be compared with the following distribution-based estimators of ϑ^2:

- the ML estimator

$$\widehat{\vartheta^2_n}^{\mathrm{ML}} = \frac{1}{2n}\sum_{j=1}^n X_j^2; \quad (4.132)$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

- the method-of-moments (MM) estimator

$$\widehat{\vartheta_n^2}^{\,\mathrm{MM}} = \frac{2}{\pi}\Big(\frac{1}{n}\sum_{j=1}^{n} X_j\Big)^2 ; \tag{4.133}$$

- the median-based estimator

$$\widehat{\vartheta_n^2}^{\,\mathrm{med}} = \frac{\widehat{\mathrm{med}}_n^{\,2}}{2\log 2} ; \tag{4.134}$$

- the minimum Cramér-von Mises distance estimator (CM)

$$\widehat{\vartheta_n^2}^{\,\mathrm{CM}} = \operatorname*{argmin}_{\vartheta^2>0} \int_0^{\infty} \Big(\widehat{F}_n(x) - F_\vartheta(x)\Big)^2 \mathrm{d}F_\vartheta(x) ; \tag{4.135}$$

- the $\mathcal{L}^2$-type minimum distance estimator based on the density

$$\widehat{\vartheta_n^2}^{\,\mathrm{Sc}} = \operatorname*{argmin}_{\vartheta^2>0} \Big\{ \int_0^{\infty} [f_\vartheta(x)]^2 \mathrm{d}x - \frac{2}{n}\sum_{j=1}^{n} f_\vartheta(X_j) \Big\}, \tag{4.136}$$

  introduced by Scott (2001). It is a special case of the so-called minimum density power divergences estimator; see Basu et al. (2011).

All the five alternative estimators are asymptotically normal. In particular,

$$\sqrt{n}\,(\widehat{\vartheta_n^2}^{\,\mathrm{ML}} - \vartheta^2) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \vartheta^4\Big) , \tag{4.137}$$

$$\sqrt{n}\,(\widehat{\vartheta_n^2}^{\,\mathrm{MM}} - \vartheta^2) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, 4\vartheta^4(4-\pi)/\pi\Big) , \tag{4.138}$$

$$\sqrt{n}\,(\widehat{\vartheta_n^2}^{\,\mathrm{Med}} - \vartheta^2) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \vartheta^4/\log^2 2\Big) , \tag{4.139}$$

$$\sqrt{n}\,(\widehat{\vartheta_n^2}^{\,\mathrm{CM}} - \vartheta^2) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, 657\vartheta^4/500\Big) , \tag{4.140}$$

$$\sqrt{n}\,(\widehat{\vartheta_n^2}^{\,\mathrm{Sc}} - \vartheta^2) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \frac{16\vartheta^4}{49}\Big(\frac{512}{27\pi} - 1\Big)\Big) . \tag{4.141}$$

The asymptotic variances of the ML and MM estimators are derived using the Delta Method, see Casella and Berger (2002). The asymptotic variance of the median-based estimator is obtained by leveraging the asymptotic variance of the empirical q-quantile, namely $\mathbb{V}_{\mathrm{as}}[X_{(nq)}] = q(1-q)/f(F^{-1}(q))$; see Staudte and Sheather (1990). The asymptotic normality of the CM and the Scott's estimators was established by Duchesne et al. (1997) and Scott (2001), respectively. The variance in (4.140) was obtained using eq. 2.28 of Duchesne et al. (1997), while that in (4.141) from Proposition 3.1 of Scott (2001).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Taking ratios of the asymptotic variances (4.138)-(4.141) to that of ML in (4.137), we obtain:

$$\mathrm{ARE}\Big[\widehat{\vartheta^2_n}^{\mathrm{MM}}, \widehat{\vartheta^2_n}^{\mathrm{ML}}\Big] = \frac{\pi/4}{(4-\pi)} \approx 91.5\%\,, \tag{4.142}$$
$$\mathrm{ARE}\Big[\widehat{\vartheta^2_n}^{\mathrm{Med}}, \widehat{\vartheta^2_n}^{\mathrm{ML}}\Big] = \log^2(2) \approx 48.0\%\,, \tag{4.143}$$
$$\mathrm{ARE}\Big[\widehat{\vartheta^2_n}^{\mathrm{CM}}, \widehat{\vartheta^2_n}^{\mathrm{ML}}\Big] = \frac{500}{657} \approx 76.1\%\,, \tag{4.144}$$
$$\mathrm{ARE}\Big[\widehat{\vartheta^2_n}^{\mathrm{Sc}}, \widehat{\vartheta^2_n}^{\mathrm{ML}}\Big] = \frac{1323\pi}{16(512-27\pi)} \approx 60.8\%\,. \tag{4.145}$$

Note that these AREs are uniform over the parameter space.

Regarding robustness, we obtain the following IFs of the alternative estimators:

$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{ML}}\Big] = x^2/2-\vartheta^2\,, \tag{4.146}$$
$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{MM}}\Big] = 2x\vartheta\sqrt{2/\pi}-2\vartheta^2\,, \tag{4.147}$$
$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{Med}}\Big] = \begin{cases} -\dfrac{\vartheta^2}{\log 2}, & x<\vartheta\sqrt{2\log 2}\,,\\ 0, & x=\vartheta\sqrt{2\log 2}\,,\\ \dfrac{\vartheta^2}{\log 2}, & x>\vartheta\sqrt{2\log 2}\,,\end{cases} \tag{4.148}$$
$$\mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{CM}}\Big] = \frac{15\vartheta^2}{8}-\frac{27\vartheta^2}{8}\Big(1+\frac{x^2}{\vartheta^2}\Big)\exp\Big(\frac{-x^2}{\vartheta^2}\Big)\,, \tag{4.149}$$
$$\mathrm{IF}\Big[x,\vartheta^{2\,\mathrm{Sc}}\Big] = \frac{64}{7}\Big[\frac{\vartheta x}{\sqrt{\pi}}\Big(\frac{x^2}{2\vartheta^2}-1\Big)\exp\Big(\frac{-x^2}{\vartheta^2}\Big)+\frac{\vartheta^2}{16}\Big]\,. \tag{4.150}$$

The IFs (4.146) and (4.147) are straightforward to obtain, and for (4.148), refer to eq. 3.2.3 of Staudte and Sheather (1990). The IF of the CM estimator was computed using eq. 2.16 in Duchesne et al. (1997), whereas that of the Scott's estimator was obtained by eq. 9.22 in Basu et al. (2011). All the IFs are visualized on the left panel of Fig. 4.14. It is evident that the median-based, CM, and the Scott's estimators are robust against large outliers, whereas ML and MM, based on sample moments, completely lack robustness. Note that the robust alternatives have significantly lower ARE w.r.t. ML than MM.

4.5.3.5 Asymptotic relative efficiency

Given the asymptotic variance of the ML estimator, equal to ϑ^4/n, we also compute ARE of our proposed estimators. Firstly, for the transform-matching estimators, we obtain

$$\mathrm{ARE}\Big[\widehat{\vartheta^2}_n^{\mathrm{LT}\,2,1,0}(s), \widehat{\vartheta^2}_n^{\mathrm{ML}}\Big] = \frac{\vartheta^4}{\sigma^{2\,\mathrm{LT}\,2,1,0}(s)}, \qquad \mathrm{ARE}\Big[\widehat{\vartheta^2}_n^{\mathrm{LT}\,3,2,1}(s), \widehat{\vartheta^2}_n^{\mathrm{ML}}\Big] = \frac{\vartheta^4}{\sigma^{2\,\mathrm{LT}\,3,2,1}(s)}\,,$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

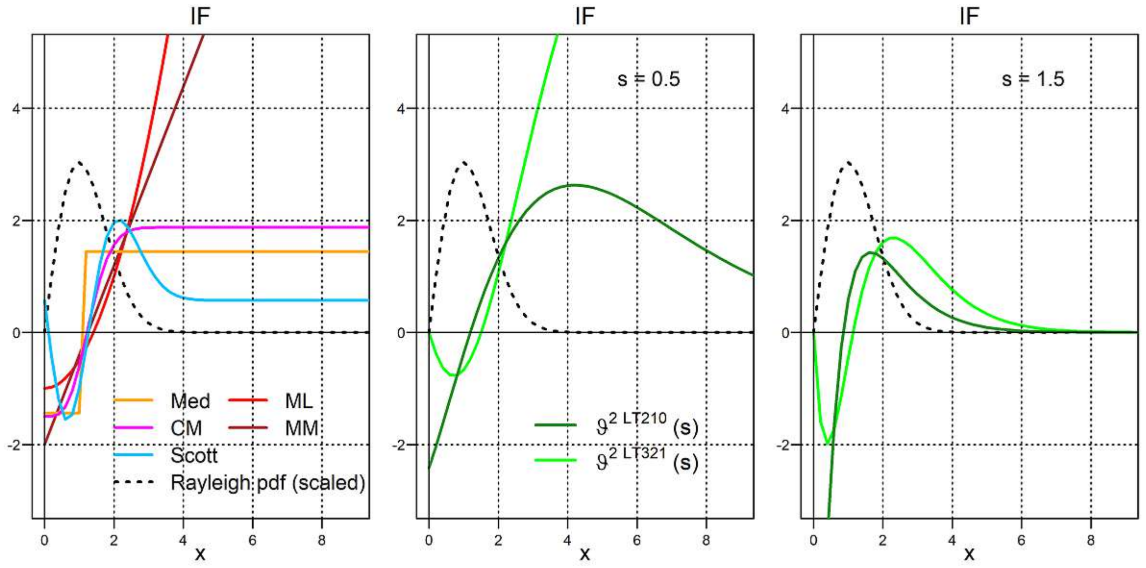


Figure 4.14: Estimation of ϑ^2 in the Rayleigh(ϑ) model. Left panel: IFs of the ML, MM, median-based, Cramér-von Mises and Scott estimators given, respectively, by (4.146)-(4.150). Middle and right panels: IFs (4.128) and (4.129) of the transform-matching estimators from Theorem 16 for $s = 0.5$ and 1.5; $\vartheta = 1$ in all the cases.

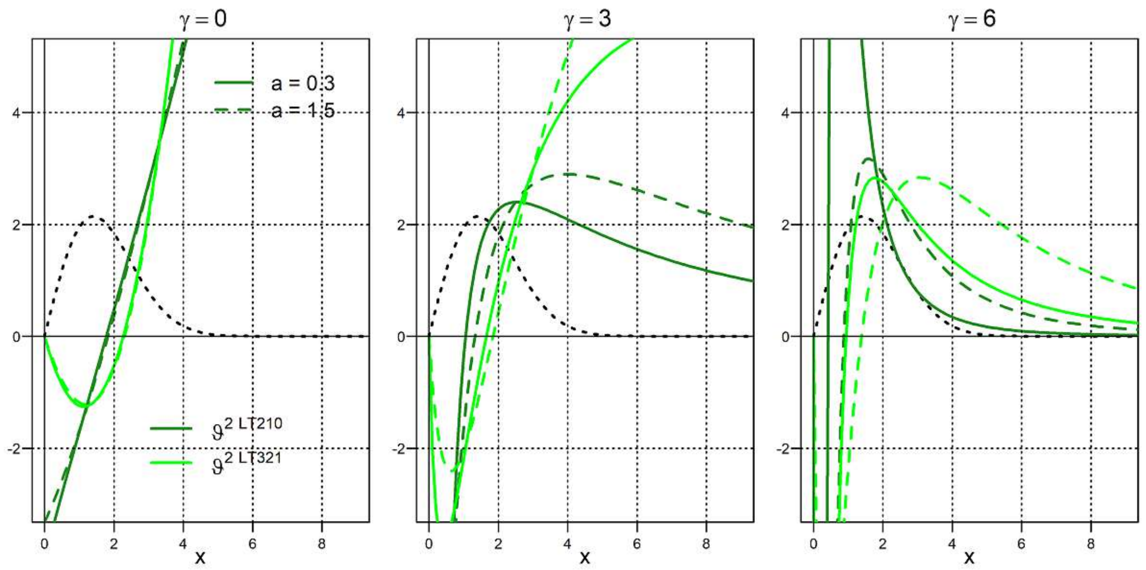


Figure 4.15: Estimation of ϑ^2 in the Rayleigh(ϑ) model: IFs (4.130) and (4.131) of the $\mathcal{L}^2$ estimators from Theorem 17 for combinations of $a = 0.3, 1.5$ and $\gamma = 0, 3, 6$; $\vartheta = 1$ in all the cases.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

where $\sigma^{2\,\mathrm{LT}\,2,1,0}(s)$ and $\sigma^{2\,\mathrm{LT}\,3,2,1}(s)$ are given by (4.128) and (4.128), respectively.

Secondly, for the $\mathcal{L}^2$ estimators, we have

$$\mathrm{ARE}\Big[\widehat{\vartheta^2}_n^{\,\mathrm{LT}\,2,1,0}, \widehat{\vartheta^2}_n^{\,\mathrm{ML}}\Big] = \frac{\vartheta^4}{\int_0^\infty \mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{LT}\,2,1,0}\Big]^2 \mathrm{d}F_\vartheta(x)}, \tag{4.151}$$

$$\mathrm{ARE}\Big[\widehat{\vartheta^2}_n^{\,\mathrm{LT}\,3,2,1}, \widehat{\vartheta^2}_n^{\,\mathrm{ML}}\Big] = \frac{\vartheta^4}{\int_0^\infty \mathrm{IF}\Big[x;\vartheta^{2\,\mathrm{LT}\,3,2,1}\Big]^2 \mathrm{d}F_\vartheta(x)}, \tag{4.152}$$

where the integrated IFs are given in Theorem 17. These AREs are displayed in Figs. 4.16-4.18.

The ARE behaves similarly to that of the LT-based estimators in the exponential model, as illustrated in Figs. 4.7 and 4.8. (Note that the shapes of the ARE surfaces differ due to the Rayleigh scale being ϑ^2 while the exponential scale is $1/\lambda$). Specifically, AREs decrease to zero as $s \to \infty$ or as $a \to 0$ while integer γ increases. Also, for the estimators that do not employ zero-order derivatives, the AREs reach a maximum for some intermediate value of s or a within the interval $(0, \infty)$, considering a fixed ϑ^2.

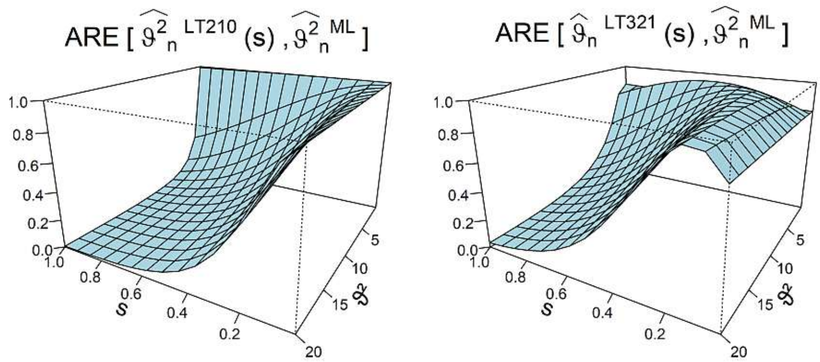


Figure 4.16: Estimation of ϑ^2 in the Rayleigh(ϑ) model: AREs of the transform-matching estimators (4.122) and (4.123) w.r.t. the ML estimator in function of ϑ^2 and s.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

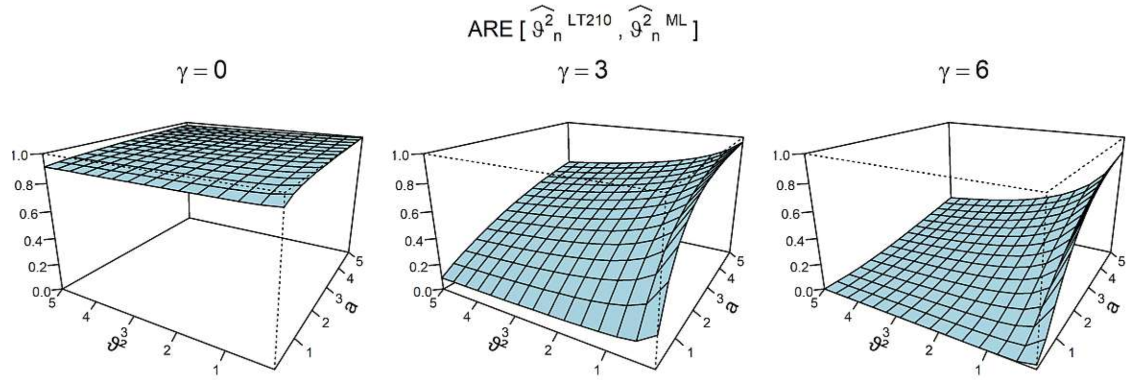


Figure 4.17: Estimation of ϑ^2 in the Rayleigh(ϑ) model: AREs of the $\mathcal{L}^2$ estimator (4.124) w.r.t. to the ML estimator in function of ϑ^2 and a, for $\gamma = 0, 3, 6$.

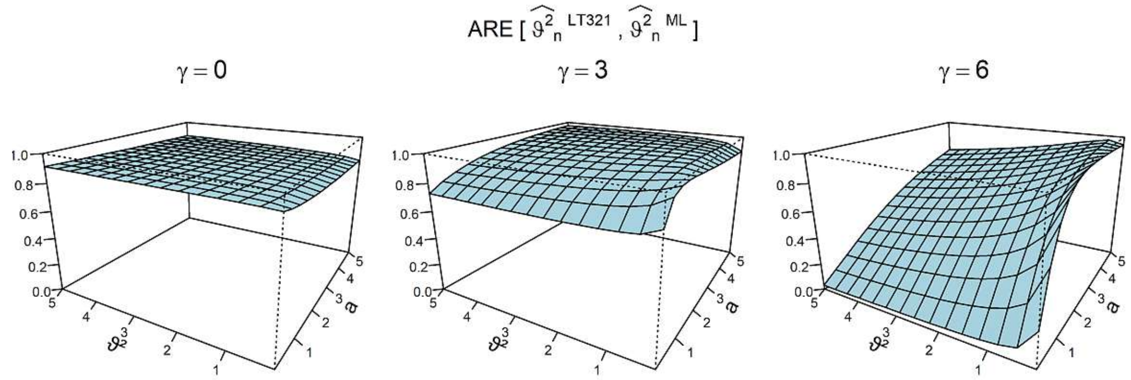


Figure 4.18: Estimation of ϑ^2 in the Rayleigh(ϑ) model: AREs of the $\mathcal{L}^2$ estimator (4.125) w.r.t. to the ML estimator in function of ϑ^2 and a, for $\gamma = 0, 3, 6$.

4.5.3.6 Robustness-efficiency trade-offs

Analyzing the IFs and lines of constancy of ARE observed in Figs. 4.16-4.18 (determined by linear or inversely proportional relations on the s-ϑ^2- and a-ϑ^2-planes), we recommend the following data-driven choice of s or a.

For the transform-matching estimators, we recommend putting

$$s = k/\overline{X}_n\,, \qquad \text{or} \qquad s = k/\widehat{\mathrm{med}}_n\,,$$

where $k = 0.75$ for $\widehat{\vartheta^2}_n^{\mathrm{LT}\,2,1,0}(s)$ and $k = 1.5$ for $\widehat{\vartheta^2}_n^{\mathrm{LT}\,3,2,1}(s)$.

As for the $\mathcal{L}^2$ estimators, we find that using $\widehat{\vartheta^2}_n^{\mathrm{LT}\,2,1,0}$ with $\gamma = 3$ and $\widehat{\vartheta^2}_n^{\mathrm{LT}\,3,2,1}(s)$ with $\gamma = 6$ provides robust yet relatively precise estimates over the entire parameter range.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Specifically, in this setting, we recommend selecting

$$a = k\overline{X}_n\,, \qquad \text{or} \qquad a = k\,\widehat{\mathrm{med}}_n,$$

with $k = 1$ in both cases.

The median-driven selection of s or a yields slightly more robust estimates then the mean-driven.

4.5.3.7 Finite-sample performance: simulation analysis

We now simulate the four LT-based estimators in the median-driven variants and compare them with the five competitors described in sec. 4.5.3.4. In Fig. 4.19, we present box plots of estimates drawn from 5000 samples of size $n = 50$, for $\vartheta^2 = 0.1$ and 10. The box plots for the Scott's estimator for $\vartheta^2 = 0.1$ is missing due to numerical problems in optimizing of its loss function for such a small parameter value. All the estimators are computed for: 1) pure samples; 2) samples contaminated with one large outlier $X_0 = \overline{X}_n + 2\max(X_1, ..., X_n)$; 3) samples contaminated with two small outliers $X_{0\text{-}1} = \overline{X}_n/10$ and $X_{0\text{-}2} = \overline{X}_n/100$.

Let us examine the results in pure samples first. The data-driven selection of s or a ensures that our estimators are invariant to changes in scale. The ratios of the sample variance of the ML estimator to the sample variances of the MM, median-based, CM and Scott's estimator were 0.91, 0.48, 0.74 and 0.59, respectively, indicating that the ARE results (4.142)-(4.145) hold nearly exactly in this finite-sample setting. The corresponding ratios for $\widehat{\vartheta 2}_n^{\mathrm{LT}\,2,1,0}(s)$, $\widehat{\vartheta 2}_n^{\mathrm{LT}\,3,2,1}(s)$, $\widehat{\vartheta 2}_n^{\mathrm{LT}\,2,1,0}$, and $\widehat{\vartheta 2}_n^{\mathrm{LT}\,3,2,1}$ were 0.7, 0.68, 0.72 and 0.64, respectively. Thus, the loss of precision of the proposed estimators in pure samples (traded for robustness) is average as compared to the standard robust alternatives.

Considering the estimation with outliers, the robustness of $\widehat{\vartheta 2}_n^{\mathrm{LT}\,2,1,0}(s)$ and $\widehat{\vartheta 2}_n^{\mathrm{LT}\,2,1,0}$ against small outliers is the weakest among all the considered estimators (as could be anticipated from shape of their IFs in Figs. 4.14 and 4.15 as $x \to 0$). On the other hand, the estimators $\widehat{\vartheta 2}_n^{\mathrm{LT}\,3,2,1}(s)$ and $\widehat{\vartheta 2}_n^{\mathrm{LT}\,3,2,1}$ clearly outperform all the competitors in terms of robustness to the two types of outliers. Between these two estimators, $\widehat{\vartheta 2}_n^{\mathrm{LT}\,3,2,1}(s)$ has slightly smaller bias than its $\mathcal{L}^2$ counterpart. In view of the quite simple expression and the potential numerical problems of the Scott's estimator, $\widehat{\vartheta 2}_n^{\mathrm{LT}\,3,2,1}(s)$ emerges as the winner of the comparison.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

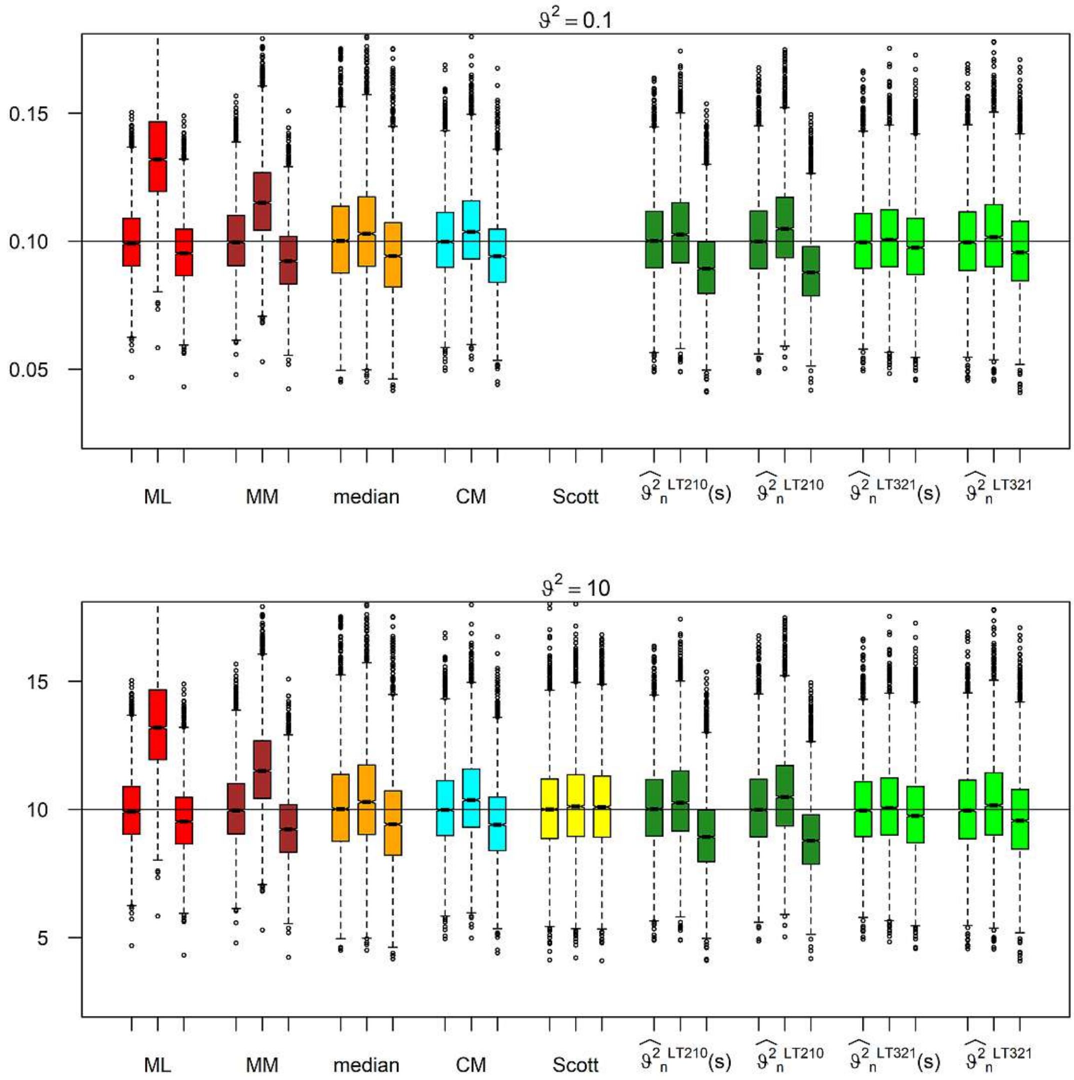


Figure 4.19: Estimation of ϑ^2 in the Rayleigh(ϑ) model from 5000 simulated samples of size $n = 50$, for $\vartheta^2 = 0.1$ (upper panel) and $\vartheta^2 = 10$ (lower panel). Comparison of the four LT-based and five standard estimators. Left box plot: pure sample; middle box plot: sample with one large outlier; right box plot: sample with two small outliers.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.4 Gamma distribution

The three distributions estimated so far were either a location-scale or a scale family. The model considered in this section, the gamma distribution, is a shape-scale family. We estimate its both parameters deriving $\mathcal{L}^2$ estimators based on the first- and second-order differential equations of the associated LT. Due to space limitation, we do not go into details of the efficiency-robustness analysis. Instead, we focus on performance of the estimators under model misspecification.

The density of the distribution is

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)\,, \quad x \geq 0, \tag{4.153}$$

where $\alpha, \beta > 0$ are the shape and rate parameters, respectively. The mean and variance are given by $\mathbb{E}[X] = \alpha/\beta$ and $\mathbb{V}[X] = \alpha/\beta^2$.

The ML estimation (1.1) is problematic in this case since the likelihood equations

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} \log X_j = -\log \widehat{\beta}_n^{\mathrm{ML}} + \psi(\widehat{\alpha}_n^{\mathrm{ML}}), \\ \overline{X}_n = \widehat{\alpha}_n^{\mathrm{ML}} / \widehat{\beta}_n^{\mathrm{ML}}, \end{cases} \tag{4.154}$$

involve the digamma function and require a numerical solution. Therefore, one often prefers the explicit, though less efficient, MM estimators,

$$\left(\widehat{\alpha}_n^{\mathrm{MM}}, \widehat{\beta}_n^{\mathrm{MM}}\right)^\top = \left(\frac{\overline{X}_n^2}{S_n^2}, \frac{S_n^2}{\overline{X}_n}\right). \tag{4.155}$$

A large body of literature is devoted to the properties and computational aspects of these ML and MM estimators and their modifications; see Bowman and Shenton (1988), sec. 17.7 of Johnson et al. (1994), or Hwang and Huang (2002) and reference therein. Useful approximate formulas for the ML estimators are used by Henze et al. (2012) and Ye and Chen (2017). Robust method for fitting the gamma model has been studied, for example, by Clarke et al. (2012). However, their estimator (a B-optimal M-estimator) is not explicit and also necessitates a numerical algorithm.

Explicit LT-based estimators can be obtained from

$$L(s) = \left(\frac{\beta}{\beta+s}\right)^\alpha, \qquad s > -\beta\,, \tag{4.156}$$

which satisfies the two differential equations:

$$(\beta+s)L'(s) + \alpha L(s) \;=\; 0, \tag{4.157}$$

$$(\beta+s)L''(s) + (\alpha+1)L'(s) \;=\; 0\,. \tag{4.158}$$

Based on the considerations in sec. 3.4.2, both equations are solved by the gamma LT with a unique setting of α and β, and therefore, can provide consistent estimators.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

4.5.4.1 Considered estimators

We consider only the integrated $\mathcal{L}^2$ estimators, based on (4.157) and (4.158), namely

$$\left(\widehat{\alpha}_n^{\mathrm{LT}\,1,0}, \widehat{\beta}_n^{\mathrm{LT}\,1,0}\right)^\top = \underset{\alpha,\beta>0}{\operatorname{argmin}} \int_0^\infty \left[(\beta+s)\widehat{L}_n'(s) + \alpha \widehat{L}_n(s)\right]^2 s^\gamma e^{-as} \mathrm{d}s\,, \tag{4.159}$$

$$\left(\widehat{\alpha}_n^{\mathrm{LT}\,2,1}, \widehat{\beta}_n^{\mathrm{LT}\,2,1}\right)^\top = \underset{\alpha,\beta>0}{\operatorname{argmin}} \int_0^\infty \left[(\beta+s)\widehat{L}_n''(s) + (\alpha+1) \widehat{L}_n'(s)\right]^2 s^\gamma e^{-as} \mathrm{d}s\,, \tag{4.160}$$

$a > 0$, $\gamma = 0, 1, \ldots$. They admit the following closed-form expressions:

$$\widehat{\alpha}_n^{\mathrm{LT}\,1,0} = \frac{C_n^2 D_n - 2A_n C_n E_n}{2A_n(4A_nB_n - C_n^2)} + \frac{D_n}{2A_n}, \tag{4.161}$$

$$\widehat{\beta}_n^{\mathrm{LT}\,1,0} = \frac{C_n D_n - 2A_n E_n}{4A_nB_n - C_n^2}\,, \tag{4.162}$$

where

$$A_n = \sum_{j,k=1}^n \frac{\gamma!}{(X_j+X_k+a)^{\gamma+1}}\,, \qquad B_n = \sum_{j,k=1}^n \frac{\gamma! X_j X_k}{(X_j+X_k+a)^{\gamma+1}}\,,$$

$$C_n = \sum_{j,k=1}^n \frac{\gamma!(X_j+X_k)}{(X_j+X_k+a))^{\gamma+1}}\,, \qquad D_n = \sum_{j,k=1}^n \frac{(\gamma+1)!(X_j+X_k)}{(X_j+X_k+a)^{\gamma+2}}\,,$$

$$E_n = \sum_{j,k=1}^n \frac{2(\gamma+1)! X_j X_k}{(X_j+X_k+a)^{\gamma+2}}\,,$$

and

$$\widehat{\alpha}_n^{\mathrm{LT}\,2,1} = \frac{2A_nC_nD_n + B_nC_n^2}{2A_n(C_n^2 - 4A_nE_n)} - \frac{B_n}{2A_n}, \tag{4.163}$$

$$\widehat{\beta}_n^{\mathrm{LT}\,2,1} = \frac{2A_nD_n + B_nC_n}{C_n^2 - 4A_nE_n}\,, \tag{4.164}$$

where

$$A_n = \sum_{j,k=1}^n \frac{\gamma!}{(X_j+X_k+a)^{\gamma+1}}\,, \qquad B_n = \sum_{j,k=1}^n \frac{\gamma! X_j X_k[(1-\gamma)(X_j+X_k)+2a]}{(X_j+X_k+a)^{\gamma+2}}\,,$$

$$C_n = \sum_{j,k=1}^n \frac{\gamma! X_j X_k (X_j+X_k)}{(X_j+X_k+a)^{\gamma+1}}\,, \qquad D_n = \sum_{j,k=1}^n \frac{\gamma! X_j X_k[2\gamma X_j X_k - X_j^2 - X_k^2 - a(X_j+X_k)]}{(X_j+X_k+a)^{\gamma+2}},$$

$$E_n = \sum_{j,k=1}^n \frac{\gamma! X_j^2 X_k^2}{(X_j+X_k+a)^{\gamma+1}}.$$

These expressions are obtained as unique minimizers of certain quadratic target functions. The derivations are tedious but routine.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.4.2 Finite-sample simulations

Joint asymptotic normality and joint IFs of (4.159) and (4.160) follows from Theorems 1 and 4. The analysis of ARE and IFs reveals that a larger value of a makes the estimators more efficient, while a larger value of γ increases their robustness against tail outliers. Similar to the previous models, a data-driven choice of a is needed to achieve robustness-efficiency trade-off and invariance to the scale of data. We found that setting $a = \alpha/\beta$ works very well for this purpose. Since $\alpha/\beta = \mathbb{E}[X]$, we estimate a using the sample mean directly, that is, we set $a = \overline{X}_n$.

Fig. 4.20 presents simulations of (4.159) with $\gamma = 2$ and (4.160) with $\gamma = 4$, compared to the ML and MM estimators. We consider samples of size $n = 50$ in three variants: 1) pure samples, 2) samples contaminated by one large outlier $X_0 = 2\max(X_1,...,X_n)$, and, 3) sample contaminated with one small outlier $X_0 = 0.01\overline{X}_n$. Three settings of $(\alpha,\ \beta)$ were examined: (3, 0.1), (3, 10) and (0.3, 0.1).

We note that the estimator (4.160), based on the equation with first and second derivative of LT (but not including the LT itself), exhibits greater robustness against small outliers compared to (4.159). Therefore, our earlier observation concerning the robustness of estimators within scale families (exponential, Rayleigh) persists within the shape-scale gamma family.

For $\alpha > 1$, the estimator (4.160) performs optimally when aiming for low variance and robustness simultaneously, irrespective of the presence and type of outliers. The scenario differs when $\alpha < 1$. While all the LT-based estimators are little biased, three of them, $\widehat{\beta}_n^{\mathrm{LT}\,1,0}$, $\widehat{\alpha}_n^{\mathrm{LT}\,2,1}$ and $\widehat{\beta}_n^{\mathrm{LT}\,2,1}$, display relatively large variances. Moreover, as indicated by the box plots, these estimates may even become negative, yielding meaningless results. To quantify this effect, the top part of Table 4.4 shows the percentages of negative estimates for sample sizes $n = 30, 50, 70, 100$. It is observed that the fraction of negative realizations is approximately 10 percent in very small samples and decreases as the sample size grows.

Based on this simulation, we recommend $\widehat{\alpha}_n^{\mathrm{LT}\,2,1}$ and $\widehat{\beta}_n^{\mathrm{LT}\,2,1}$ when density is known to be modal ($\alpha > 1$) prior to estimation. If the shape is unknown, then using $\widehat{\alpha}_n^{\mathrm{LT}\,1,0}$ alone still performs quite well. Notably, for $\alpha < 1$, it is more precise than the MM estimator in pure samples and more robust against large outliers. In addition, it competes favorably with the ML estimator due to its closed-form expression.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

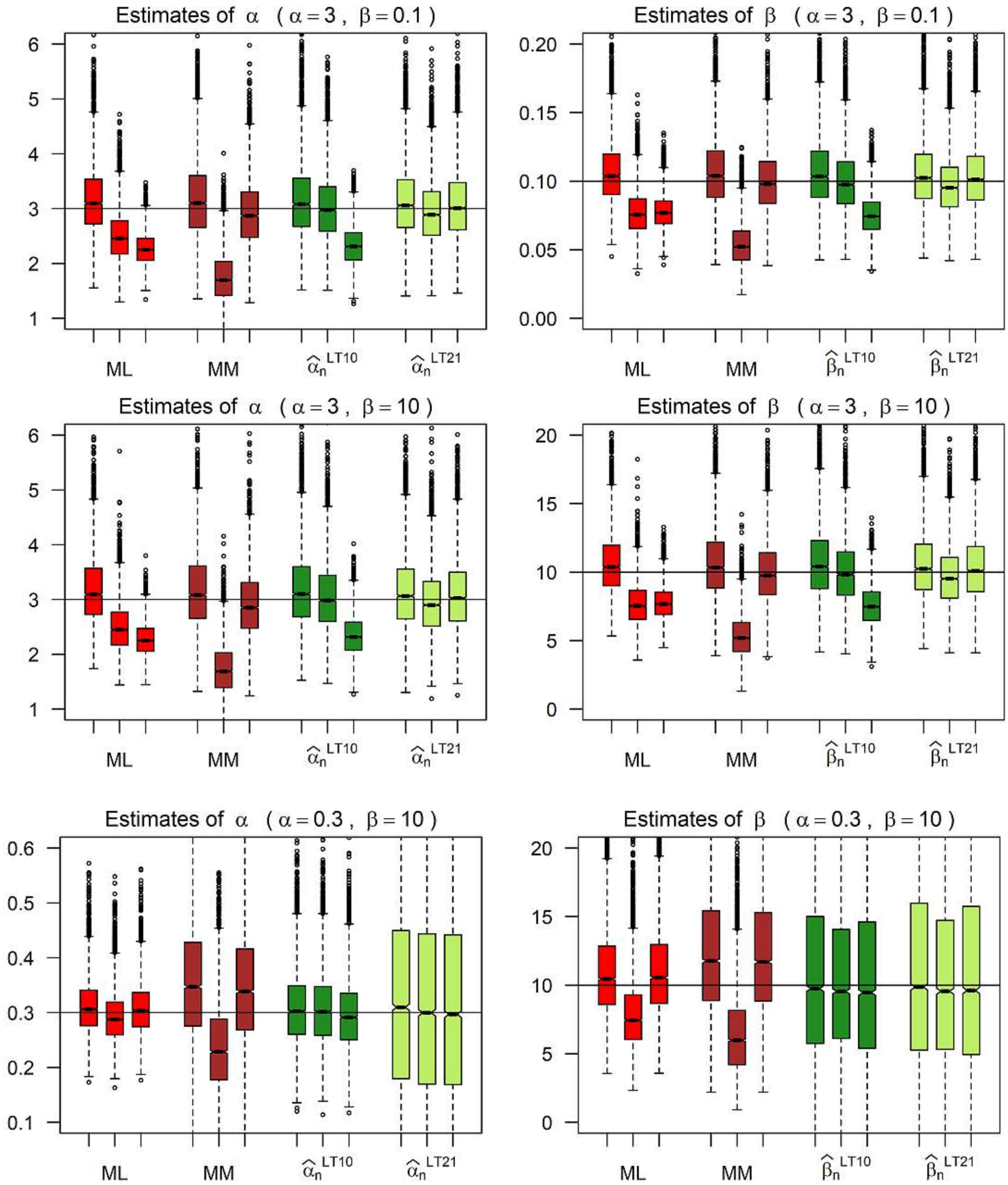


Figure 4.20: Simultaneous estimation of α and β in the gamma(α, β) model from 5000 simulated samples of size $n = 50$ for three combinations of the parameters. The $\mathcal{L}^2$ LT-based estimators are compared with the ML and MM estimators. Left box plot: pure sample; middle box plot: sample with one large outlier; right box plot: sample with one small outlier.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

4.5.4.3 Estimation under misspecification

The assumption regarding the gamma model may be incorrect if the sample actually originates from another one-sided distribution with a similar shape. Common examples of such distributions include the log-normal, Weibull, Rayleigh, and inverse Gaussian. Assessing the impact of misspecification on estimators in skewed distributions differs from symmetric location-scale models, where the gain in the variance of the estimators can be computed analytically or simulated. In skewed models, the parameters do not have the same 'meaning' across different distributions. Therefore, we quantify the misspecification effect here using three non-parametric measures:

1. Mean integrated squared error (MISE)

$$\text{MISE} = \mathbb{E}\int_0^\infty \Big[f_{\text{true}}(x) - f(x|\widehat{\alpha}_n, \widehat{\beta}_n)\Big]^2 \mathrm{d}x; \tag{4.165}$$

2. Mean integrated absolute error (MIAE)

$$\text{MIAE} = \mathbb{E}\int_0^\infty \Big|f_{\text{true}}(x) - f(x|\widehat{\alpha}_n, \widehat{\beta}_n)\Big| \mathrm{d}x; \tag{4.166}$$

3. Mean Kullback-Leibler divergence (MKLD)

$$\text{MKLD} = \mathbb{E}\int_0^\infty f_{\text{true}}(x) \log\left(\frac{f_{\text{true}}(x)}{f(x|\widehat{\alpha}_n, \widehat{\beta}_n)}\right) \mathrm{d}x. \tag{4.167}$$

In these definitions, $f_{\text{true}}(x)$ denotes the true density that generates data, whereas $f(x|\widehat{\alpha}_n, \widehat{\beta}_n)$ is the wrongly assumed gamma density, estimated using some estimators of α and β. In addition, we explore the impact of misspecification on the parametric estimation of median and 0.9-quantile of the true distribution, using the mean squared error (MSE)

$$\begin{aligned} \text{MSE}[x_q] &= \mathbb{E}\Big[F^{-1}_{\text{true}}(q) - F^{-1}(q|\widehat{\alpha}_n, \widehat{\beta}_n)\Big]^2 \qquad (4.168)\\ &= F^{-1}_{\text{true}}(q) - \mathbb{E}\Big[F^{-1}(q|\widehat{\alpha}_n, \widehat{\beta}_n)\Big] + \mathbb{V}\Big[F^{-1}(q|\widehat{\alpha}_n, \widehat{\beta}_n)\Big], \qquad q = 0.5, 0.9. \end{aligned}$$

Unfortunately, none of the four aforementioned measures is analytically computable, even in asymptotically as $n \to \infty$. Therefore, we resort to simulations of the estimators and numerical computations of the measures. In these simulations, we employed 5000 samples drawn from the following true distributions:

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

1. Rayleigh(ϑ^2) with $\vartheta^2 = 0.5$ and 3,
2. log-normal(μ,σ) with $\sigma = 0.5$ and 1.5 ($\mu = 0$),
3. Weibull(c, λ) with $c = 0.5$ and 2 ($\lambda = 0$),
4. inverse Gaussian(μ, λ) with $\mu = 1, \lambda = 3$ and $\mu = 3, \lambda = 1$.

Considered were small and moderate sample sizes, n = 30, 50, 70 and 100 (for which misspecification is more likely to happen than for large samples). We found that our LT-based estimators may occasionally have negative realizations; this happened for the log-normal(0,1.5), Weibull(0.5,1) and inverse Gaussian(3,1) cases for smaller samples. Since negative estimates of α and β hinder the computation of the misspecification measures, the samples which produced negative estimates were excluded from the ongoing simulation. In particular, every sample for which at least one of the considered estimators was negative was replaced with a new one, and the estimates were computed afresh. The exact percentages of negative realizations (and replaced samples) are reported in Table 4.4.

The results of simulation conducted in this way are presented in Tables 4.2 and 4.3. Bold font is used to indicate the best fit to the particular true model (log-normal, Weibull, Rayleigh and inverse Gaussian) in terms of each measure. An analysis of the tables suggest the following. When the true distribution is modal (Rayleigh, inverse Gaussian and Weibull with $c = 2$), none of the estimators is superior. The proposed LT-based estimators tend to provide a better fit despite misspecification, as indicated by the smallest values of MISE, MIAE, and MSE of the median. On the other hand, the ML estimator performs best regarding MKLD. This aligns with expectations, as the ML estimators minimize Kullback-Leibler divergences. In turn, the ML and MM estimators generally yield better estimations of the 0.9-quantile, although there are exceptions (see the log-normal case). The LT-based estimators certainly fail for the long-tail variant of the Weibull distribution with $c = 0.5$, that is when $f(0^+) = \infty$. Note that the gamma density can adapt to this shape with $\alpha < 1$, but, as we know from Fig. 4.20, the LT-based estimators have quite large variances in this parameter setting. This explains large MISE and MSE of 0.9-quantile relative to ML and MM in this particular scenario.

The presented study is not exhaustive, but its goal was to demonstrate that the proposed estimators generally do not fail under misspecification. On the contrary, we can conclude that they are competitive and exhibit robustness in this context as well.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Table 4.2: Loss due to misspecification under assumption of the gamma model.
True model: Rayleigh(ϑ^2) with $\vartheta^2 = 0.5$ (left) and $\vartheta^2 = 3$ (right)

measure	n	ML		MM		$(\widehat{\alpha}_n^{\mathrm{LT}\,1,0}, \widehat{\beta}_n^{\mathrm{LT}\,1,0})$		$(\widehat{\alpha}_n^{\mathrm{LT}\,2,1}, \widehat{\beta}_n^{\mathrm{LT}\,2,1})$	
MISE	30	0.031	0.013	0.033	0.014	0.031	0.012	**0.029**	**0.011**
	50	0.021	0.009	0.024	0.010	0.020	0.009	**0.019**	**0.008**
	70	0.017	0.007	0.020	0.008	0.016	0.007	**0.015**	**0.006**
	100	0.015	0.005	0.017	0.007	0.013	0.005	**0.012**	**0.005**
MIAE	30	**0.202**	**0.202**	0.205	0.205	0.217	0.217	0.203	0.203
	50	0.171	0.171	0.175	0.175	0.182	0.182	**0.167**	**0.167**
	70	0.156	0.156	0.160	0.160	0.164	0.164	**0.148**	**0.148**
	100	0.146	0.145	0.151	0.151	0.152	0.152	**0.137**	**0.137**
MKLD	30	**0.056**	**0.058**	0.067	0.067	0.066	0.066	0.062	0.062
	50	**0.039**	**0.039**	0.048	0.048	0.046	0.046	0.043	0.043
	70	**0.031**	**0.032**	0.040	0.040	0.038	0.038	0.035	0.035
	100	**0.027**	**0.027**	0.036	0.036	0.033	0.033	0.030	0.030
MSE[$x_{0.5}$]	30	**0.008**	**0.047**	**0.008**	**0.050**	**0.008**	0.049	**0.008**	0.049
	50	**0.005**	**0.029**	0.006	0.033	**0.005**	**0.029**	**0.005**	**0.029**
	70	0.004	0.021	0.004	0.026	**0.003**	**0.020**	**0.003**	**0.020**
	100	0.003	0.017	0.004	0.021	0.003	0.116	**0.002**	**0.115**
MSE[$x_{0.9}$]	30	**0.020**	**0.124**	0.024	0.145	0.074	0.442	0.043	0.260
	50	**0.012**	**0.072**	0.015	0.088	0.052	0.311	0.028	0.170
	70	**0.008**	**0.050**	0.010	0.064	0.042	0.252	0.023	0.130
	100	**0.006**	**0.037**	0.008	0.050	0.035	0.211	0.018	0.107

True model: log-normal(μ,σ) with $\sigma = 0.5$ (left) and $\sigma = 1.5$ (right), for $\mu = 0$

measure	n	ML		MM		$(\widehat{\alpha}_n^{\mathrm{LT}\,1,0}, \widehat{\beta}_n^{\mathrm{LT}\,1,0})$		$(\widehat{\alpha}_n^{\mathrm{LT}\,2,1}, \widehat{\beta}_n^{\mathrm{LT}\,2,1})$	
MISE	30	0.027	0.049	0.035	0.130	0.027	**0.039**	**0.026**	0.154
	50	0.019	0.045	0.026	0.140	**0.017**	**0.030**	**0.017**	0.144
	70	0.016	0.053	0.023	0.180	**0.014**	**0.032**	**0.014**	0.175
	100	0.013	0.053	0.020	0.194	0.012	**0.030**	**0.011**	0.178
MIAE	30	0.208	0.353	0.230	0.491	0.205	**0.296**	**0.202**	0.507
	50	0.177	0.320	0.200	0.500	0.174	**0.248**	**0.170**	0.435
	70	0.163	0.376	0.186	0.628	0.159	**0.280**	**0.153**	0.494
	100	0.152	0.376	0.174	0.666	0.146	**0.270**	**0.140**	0.476
MKLD	30	**0.064**	0.091	0.073	0.247	0.075	0.046	0.068	**0.026**
	50	**0.045**	0.084	0.054	0.284	0.055	**0.021**	0.050	0.206
	70	**0.037**	0.101	0.046	0.378	0.046	**0.015**	0.041	0.218
	100	**0.032**	0.101	0.040	0.419	0.040	**0.007**	0.035	0.198
MSE[$x_{0.5}$]	30	**0.010**	0.318	0.011	0.518	**0.010**	**0.175**	**0.010**	0.391
	50	**0.006**	0.387	0.008	0.504	**0.006**	**0.204**	**0.006**	0.410
	70	0.005	0.310	0.006	0.379	0.005	**0.043**	**0.004**	0.290
	100	0.004	0.344	0.005	0.348	**0.003**	**0.029**	**0.003**	0.276
MSE[$x_{0.9}$]	30	0.071	17.19	**0.053**	17.23	0.056	**11.30**	**0.053**	23.81
	50	0.046	16.75	**0.034**	17.92	0.040	**13.58**	0.036	16.58
	70	0.034	9.145	**0.024**	10.34	0.032	**7.409**	0.027	8.028
	100	0.024	7.467	**0.017**	7.993	0.026	**7.196**	0.021	7.407

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Table 4.3: Loss due to misspecification under assumption of the gamma model (2). True model: Weibull(c, λ) with $c = 0.5$ (left) and $c = 2$ (right) for $\lambda = 1$

measure	n	ML		MM		$(\widehat{\alpha}_n^{\mathrm{LT}\,1,0}, \widehat{\beta}_n^{\mathrm{LT}\,1,0})$		$(\widehat{\alpha}_n^{\mathrm{LT}\,2,1}, \widehat{\beta}_n^{\mathrm{LT}\,2,1})$	
MISE	30	**0.245**	0.031	1.407	0.033	0.378	0.031	2.354	**0.029**
	50	**0.187**	0.021	1.677	0.024	0.296	0.021	2.612	**0.019**
	70	**0.166**	0.017	1.858	0.020	0.266	0.016	2.691	**0.015**
	100	**0.148**	0.014	2.037	0.017	0.235	0.013	2.738	**0.012**
MIAE	30	**0.216**	**0.201**	0.271	0.205	0.244	0.217	0.374	0.204
	50	**0.190**	0.171	0.253	0.175	0.197	0.182	0.343	**0.167**
	70	0.180	0.156	0.251	0.160	**0.176**	0.164	0.319	**0.150**
	100	0.173	0.146	0.253	0.150	**0.160**	0.152	0.298	**0.137**
MKLD	30	**0.046**	**0.057**	0.114	0.067	0.091	0.065	0.308	0.062
	50	**0.031**	**0.039**	0.116	0.048	0.054	0.046	0.268	0.043
	70	**0.027**	**0.032**	0.124	0.040	0.039	0.038	0.240	0.035
	100	**0.024**	**0.027**	0.130	0.036	0.029	0.033	0.204	0.030
MSE$[x_{0.5}]$	30	**0.108**	0.008	0.110	0.008	0.712	0.008	0.370	0.008
	50	0.072	**0.005**	0.071	**0.005**	**0.050**	**0.005**	0.199	**0.005**
	70	0.061	0.004	0.056	0.004	**0.028**	**0.003**	0.108	**0.003**
	100	0.058	0.003	0.049	0.004	**0.019**	**0.002**	0.097	**0.002**
MSE$[x_{0.9}]$	30	5.581	**0.020**	**5.540**	0.024	195.8	0.074	206.5	0.043
	50	3.412	**0.012**	**3.377**	0.015	8.084	0.052	116.4	0.028
	70	2.588	**0.008**	**2.505**	0.011	16.60	0.042	12.60	0.022
	100	2.021	**0.006**	**1.879**	0.008	3.376	0.035	6.555	0.017

True model: inverse Gaussian(μ, λ) with $\mu = 1, \lambda = 3$ (left) and $\mu = 3, \lambda = 1$ (right)

measure	n	ML		MM		$(\widehat{\alpha}_n^{\mathrm{LT}\,1,0}, \widehat{\beta}_n^{\mathrm{LT}\,1,0})$		$(\widehat{\alpha}_n^{\mathrm{LT}\,2,1}, \widehat{\beta}_n^{\mathrm{LT}\,2,1})$	
MISE	30	0.034	0.088	0.043	1.341	0.034	**0.055**	**0.033**	1.745
	50	0.026	0.088	0.034	1.567	0.024	**0.049**	**0.023**	1.397
	70	0.022	0.088	0.030	1.682	0.020	**0.046**	**0.019**	1.246
	100	0.020	0.087	0.027	1.798	0.018	**0.044**	**0.016**	0.982
MIAE	30	0.222	0.440	0.240	0.524	0.226	**0.394**	**0.218**	0.508
	50	0.196	0.432	0.213	0.540	0.198	**0.375**	**0.186**	0.482
	70	0.183	0.430	0.199	0.551	0.183	**0.366**	**0.171**	0.476
	100	0.173	0.428	0.188	0.562	0.173	**0.360**	**0.160**	0.457
MKLD	30	**0.072**	**0.132**	0.080	0.274	0.089	0.133	0.079	0.286
	50	**0.052**	0.124	0.061	0.299	0.067	**0.099**	0.058	0.235
	70	**0.044**	0.122	0.053	0.315	0.057	**0.085**	0.050	0.218
	100	**0.039**	0.121	0.048	0.330	0.050	**0.075**	0.044	0.200
MSE$[x_{0.5}]$	30	**0.009**	0.268	0.011	0.572	**0.009**	**0.137**	**0.009**	0.343
	50	0.006	0.219	0.007	0.482	0.006	**0.082**	**0.005**	0.283
	70	**0.004**	0.197	0.006	0.443	**0.004**	**0.055**	**0.004**	0.255
	100	**0.003**	0.188	0.005	0.419	**0.003**	**0.040**	**0.003**	0.222
MSE$[x_{0.9}]$	30	0.064	8.998	**0.052**	**6.605**	0.066	10.24	0.059	10.46
	50	0.041	6.210	**0.033**	**4.007**	0.049	9.674	0.041	8.321
	70	0.030	5.098	**0.024**	**2.899**	0.041	9.422	0.032	7.692
	100	0.021	4.267	**0.017**	**2.070**	0.036	9.297	0.026	7.246

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Table 4.4: Percentages of negative realizations of the LT-based estimators in 5000 samples.

distribution	n	$\widehat{\alpha}_n^{\mathrm{LT}\,1,0}$	$\widehat{\beta}_n^{\mathrm{LT}\,1,0}$	$\widehat{\alpha}_n^{\mathrm{LT}\,2,1}$	$\widehat{\beta}_n^{\mathrm{LT}\,2,1}$
gamma(3,0.1)	30	–	–	–	–
	50	–	–	–	–
	70	–	–	–	–
	100	–	–	–	–
gamma(3,10)	30	–	–	–	–
	50	–	–	–	–
	70	–	–	–	–
	100	–	–	–	–
gamma(0.3,10)	30	–	7.92	10.5	13.4
	50	–	2.46	4.68	6.14
	70	–	1.04	2.46	3.44
	100	–	0.26	0.88	1.38
log-normal(0,0.5)	30	–	–	–	–
	50	–	–	–	–
	70	–	–	–	–
	100	–	–	–	–
log-normal(0,1.5)	30	–	–	4.8	0.74
	50	–	–	1.64	0.06
	70	–	–	0.84	–
	100	–	–	0.18	–
Weibull(0.5,1)	30	–	0.96	13.52	9.02
	50	–	0.08	7.0	3.16
	70	–	0.02	3.36	1.26
	100	–	–	1.62	0.44
Weibull(2,1)	30	–	–	–	–
	50	–	–	–	–
	70	–	–	–	–
	100	–	–	–	–
inv. Gaussian(1,3)	30	–	–	–	–
	50	–	–	–	–
	70	–	–	–	–
	100	–	–	–	–
inv. Gaussian(3,1)	30	–	–	1.26	0.10
	50	–	–	0.44	–
	70	–	–	0.06	–
	100	–	–	–	–

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.5 Pareto-Lomax distribution

In this example, we demonstrate the transform-based estimation for a distribution which may not have finite moments. In such cases, due to the assumptions of Theorems 1-3, we cannot employ the $\mathcal{L}^2$ estimators since these would not be consistent over the entire parameter space. However, we can still apply the 'transform-matching' estimators computed at a single s-point.

As a popular example, we consider the Pareto-Lomax distribution (also known as Pareto Type II), which is widely used to model heavy-tailed phenomena in economics, actuarial science, and operations research; see Arnold (2015) or chapter 20 of Johnson et al. (1994). Its cumulative distribution and density functions are given by

$$F_{\alpha,\beta}(x) = 1 - \left(\frac{\beta}{x+\beta}\right)^{\alpha}, \tag{4.169}$$

and

$$f_{\alpha,\beta}(x) = \frac{\alpha\beta^{\alpha}}{(x+\beta)^{\alpha+1}}, \tag{4.170}$$

where $x \geq 0$, $\alpha, \beta > 0$. Depending on the value of the shape parameter α, the distribution may lack moments of certain orders. Specifically, the mean and the variance, expressed by

$$\mathbb{E}[X] = \frac{\beta}{\alpha - 1}, \tag{4.171}$$

and

$$\mathbb{V}[X] = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}, \tag{4.172}$$

exist when $\alpha > 1$ and $\alpha > 2$, respectively.

The traditional methods of estimation encounter difficulties in this case. Firstly, due the potential divergence of sample moments, using the MM estimators yields poor results unless α is relatively large. In turn, the ML estimators, defined in (1.1) as the maximizers of the log-likelihood function

$$\sum_{j=1}^{n} \log f_{\alpha,\beta}(X_j) = n\log\alpha + n\alpha\log\beta - (\alpha+1)\sum_{j=1}^{n}\log(X_j+\beta), \tag{4.173}$$

satisfy the following system:

$$\begin{cases} \widehat{\alpha}_n^{\mathrm{ML}} = \dfrac{n}{\sum_{j=1}^{n} \log\left(1 + X_j/\widehat{\beta}_n^{\mathrm{ML}}\right)} \\ n\widehat{\alpha}_n^{\mathrm{ML}} = (\widehat{\alpha}_n^{\mathrm{ML}}+1)\sum_{j=1}^{n} \dfrac{\widehat{\beta}_n^{\mathrm{ML}}}{X_j + \widehat{\beta}_n^{\mathrm{ML}}}, \end{cases} \tag{4.174}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

which necessitates a numerical solution; see, e.g., Rytgaard (1990).

Among the classical robust estimators, the minimum distance CM estimator takes the following specific form:

$$(\widehat{\alpha}_n^{\text{CM}}, \widehat{\beta}_n^{\text{CM}})^\top = \underset{\alpha,\beta>0}{\text{argmin}} \sum_{j=1}^{n} \left[1 - \left(\frac{\beta}{\beta + X_{(j)}}\right)^{\alpha} - \frac{2j-1}{2n}\right]^2, \tag{4.175}$$

The estimator has been recently applied to the Pareto-Lomax distribution and studied via simulations in Nombebe et al. (2022). While it requires a numerical approach, it appears to be numerically stable over the parameter space.

4.5.5.1 Considered estimators

According to Nadarajah and Kotz (2006), the LT of the distribution can be expressed with the help of the lower incomplete gamma function:

$$L(s) \;=\; \alpha(\beta s)^{\alpha} \exp\{s\beta\}\Gamma(\beta s, -\alpha)\,, \quad s \geq 0. \tag{4.176}$$

While this expression is difficult to work with, in secs. 3.2.1.1 and 3.4.2 it was proved that it uniquely satisfies

$$sL'(s) - (\alpha + \beta s)L(s) + \alpha = 0, \qquad s > 0. \tag{4.177}$$

The same is true for the corresponding second order equation,

$$sL''(s) - (\alpha + \beta s - 1)L'(s) - \beta L(s) = 0, \qquad s > 0. \tag{4.178}$$

We can use their empirical versions to obtain closed-form estimators. According to the approach outlined in sec. 4.1.2, the estimators solve the following systems:

$$\begin{cases} s_1\widehat{L}_n'(s_1) - (\alpha + \beta s_1)\widehat{L}_n(s_1) + \alpha = 0 \\ s_2\widehat{L}_n'(s_2) - (\alpha + \beta s_2)\widehat{L}_n(s_2) + \alpha = 0\;, \end{cases} \tag{4.179}$$

and

$$\begin{cases} s_1\widehat{L}_n''(s_1) - (\alpha + \beta s_1 - 1)\widehat{L}_n'(s_1) - \beta L(s_1) = 0 \\ s_2\widehat{L}_n''(s_2) - (\alpha + \beta s_2 - 1)\widehat{L}_n'(s_2) - \beta L(s_2) = 0\;. \end{cases} \tag{4.180}$$

The corresponding solvers, $\widehat{\alpha}_n^{\text{LT}\,1,0}(s_1, s_2)$, $\widehat{\beta}_n^{\text{LT}\,1,0}(s_1, s_2)$ and $\widehat{\alpha}_n^{\text{LT}\,2,1,0}(s_1, s_2)$, $\widehat{\beta}_n^{\text{LT}\,2,1,0}(s_1, s_2)$, can be presented in limiting form as $s_1 \to s_2 \to s$. This leads to more compact estimators, controlled at a single s-point. After some calculus, we obtain

$$\begin{aligned} \widehat{\alpha}_n^{\text{LT}\,1,0}(s) &= \frac{s[\widehat{\beta}_n^{\text{LT}\,1,0}(s)\widehat{L}_n(s) - \widehat{L}_n'(s)]}{1 - \widehat{L}_n(s)}\,, \\ \widehat{\beta}_n^{\text{LT}\,1,0}(s) &= \frac{[\widehat{L}_n'(s) + s\widehat{L}_n''(s)][1 - \widehat{L}_n(s)] + s[\widehat{L}_n'(s)]^2}{[\widehat{L}_n(s) + s\widehat{L}_n'(s)][1 - \widehat{L}_n(s)] + s[\widehat{L}_n(s)\widehat{L}_n'(s)]}, \end{aligned} \tag{4.181}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

and

$$\widehat{\alpha}_n^{\mathrm{LT}\,2,1,0}(s) = 1 + s\frac{\widehat{L}_n''(s)}{\widehat{L}_n'(s)} - \widehat{\beta}_n^{\mathrm{LT}\,2,1,0}(s)\left(\frac{\widehat{L}_n(s)}{\widehat{L}_n'(s)} + s\right),$$
$$\widehat{\beta}_n^{\mathrm{LT}\,2,1,0}(s) = \frac{s[\widehat{L}_n''(s)]^2 - \widehat{L}_n'(s)\widehat{L}_n''(s) - s\widehat{L}_n'(s)\widehat{L}_n'''(s)}{\widehat{L}_n(s)\widehat{L}_n''(s) - 2[\widehat{L}_n'(s)]^2}. \tag{4.182}$$

Based on the Delta-Method, the obtained estimators are consistent and asymptotically normal, however, due to the special function in (4.176), the expressions for asymptotic covariance matrices are not tractable.

Like for the previously studied distributions, we need to apply a data-driven selection of the s-point to ensure a satisfactory trade-off between accuracy and robustness. Numerous trials have shown that, for the estimator (4.181), a good trade-off occurs when s is inversely proportional to the sample median, i.e. $s = k/\widehat{\mathrm{med}}_n$, with $k > 0$ selected by the user. Unfortunately, no appropriate rule has been found for selecting s for the second estimator (4.182). Nevertheless, since this estimator turns out to be much less precise than (4.181), it is not recommended for practical use and is excluded from the simulations presented below.

4.5.5.2 Finite-sample simulations

This simulation evaluates the performance of $\widehat{\alpha}_n^{\mathrm{LT}\,1,0}(s)$ and $\widehat{\beta}_n^{\mathrm{LT}\,1,0}(s)$ with $s = k/\widehat{\mathrm{med}}_n$ for $k = 1$, by comparing them with the corresponding ML and CM estimators[4].

Considered are samples of size $n = 70$ in three variants: 1) pure samples, 2) samples contaminated by two large outliers, $X_{0\text{-}1} = 5\max(X_1,...,X_n)$, $X_{0\text{-}2} = 10\max(X_1,...,X_n)$, and, 3) samples contaminated with two small outliers, $X_{0\text{-}1} = 0.1\min(X_1,...,X_n)$ and $X_{0\text{-}2} = 0.01\min(X_1,...,X_n)$. The magnitudes of the large outliers reflect heavy-tailedness of the distribution.

The box plots in Fig. 4.21 illustrate the results for two combination of the parameters: $\alpha = 1.5$, $\beta = 2.0$ (light tail) and $\alpha = 0.5$, $\beta = 1.0$ (heavy tail). The performance of our estimators varies with respect to the ML and CM estimators, depending on the parameter values and contamination scenario. In pure samples, they are clearly outperformed by ML and slightly surpassed by CM in terms of variance. However, their robustness is comparable to that of CM in both contamination scenarios and far higher than of ML in the presence of large outliers. In Fig. 4.22, we fix $\beta = 1.0$, and, we examine only $\widehat{\alpha}_n^{\mathrm{LT}\,1,0}(s)$ for extremely heavy tails, that is for $\alpha < 1$ ($\alpha = 0.2,\ 0.4,\ 0.6,\ 0.8$). Our observation from the previous figure hold true here as well. Therefore, we can certainly recommend $\widehat{\alpha}_n^{\mathrm{LT}\,1,0}(s)$ for applications where explicitness is desired, keeping in mind that the cost lies in somewhat increased variance of the estimation.

[4] In this example, the ML estimators were computed using the function flomax from R-package Renext, whereas the CM estimation was implemented by using the R-function optim with the Nelder-Mead method.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

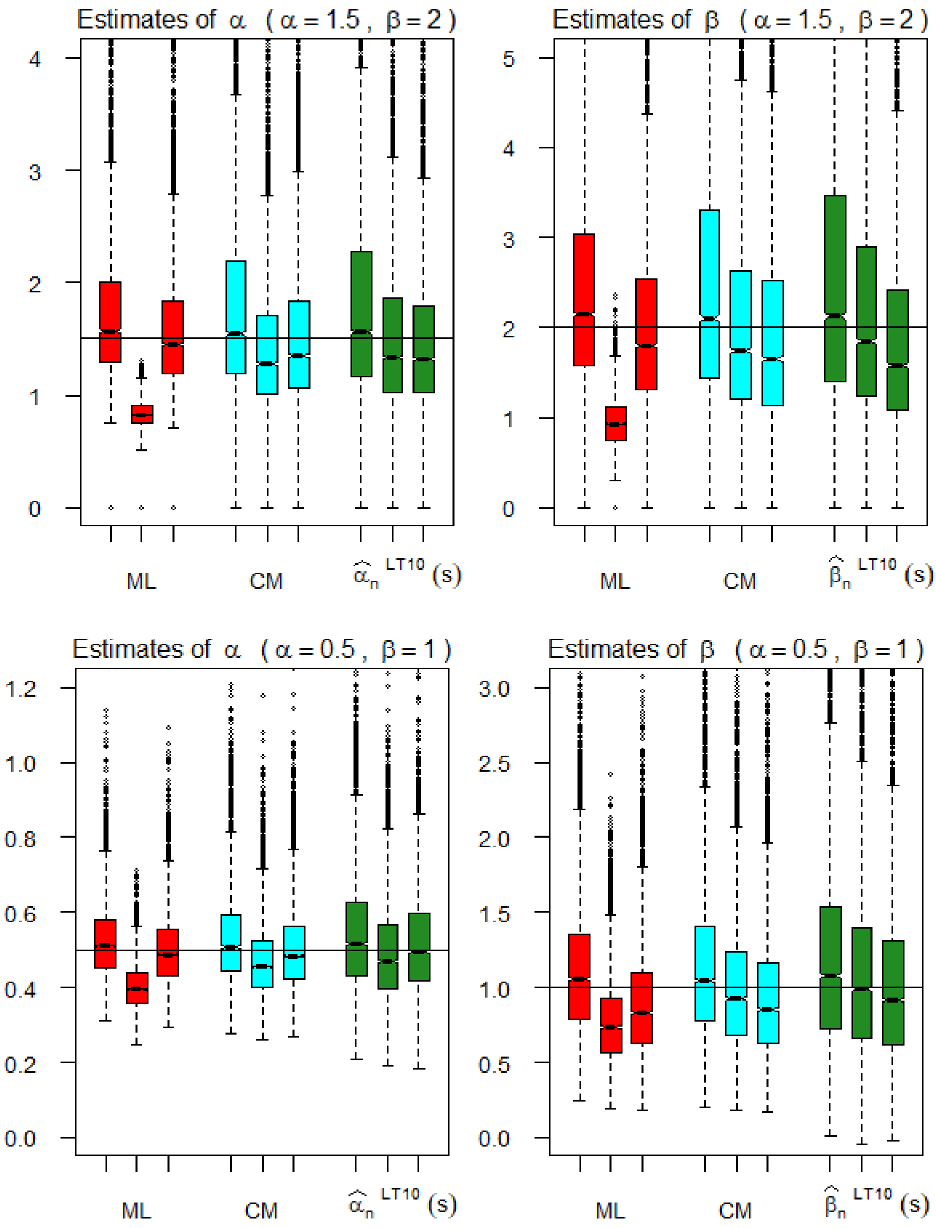


Figure 4.21: Estimation of α and β in the Pareto-Lomax distribution from 5000 generated samples of size $n = 70$, for two combinations of the parameters: $\alpha = 1.5$, $\beta = 2.0$ and $\alpha = 0.5$, $\beta = 1.0$. Left box plot: pure sample; middle box plot: sample with two large outliers; right box plot: sample with two small outliers.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

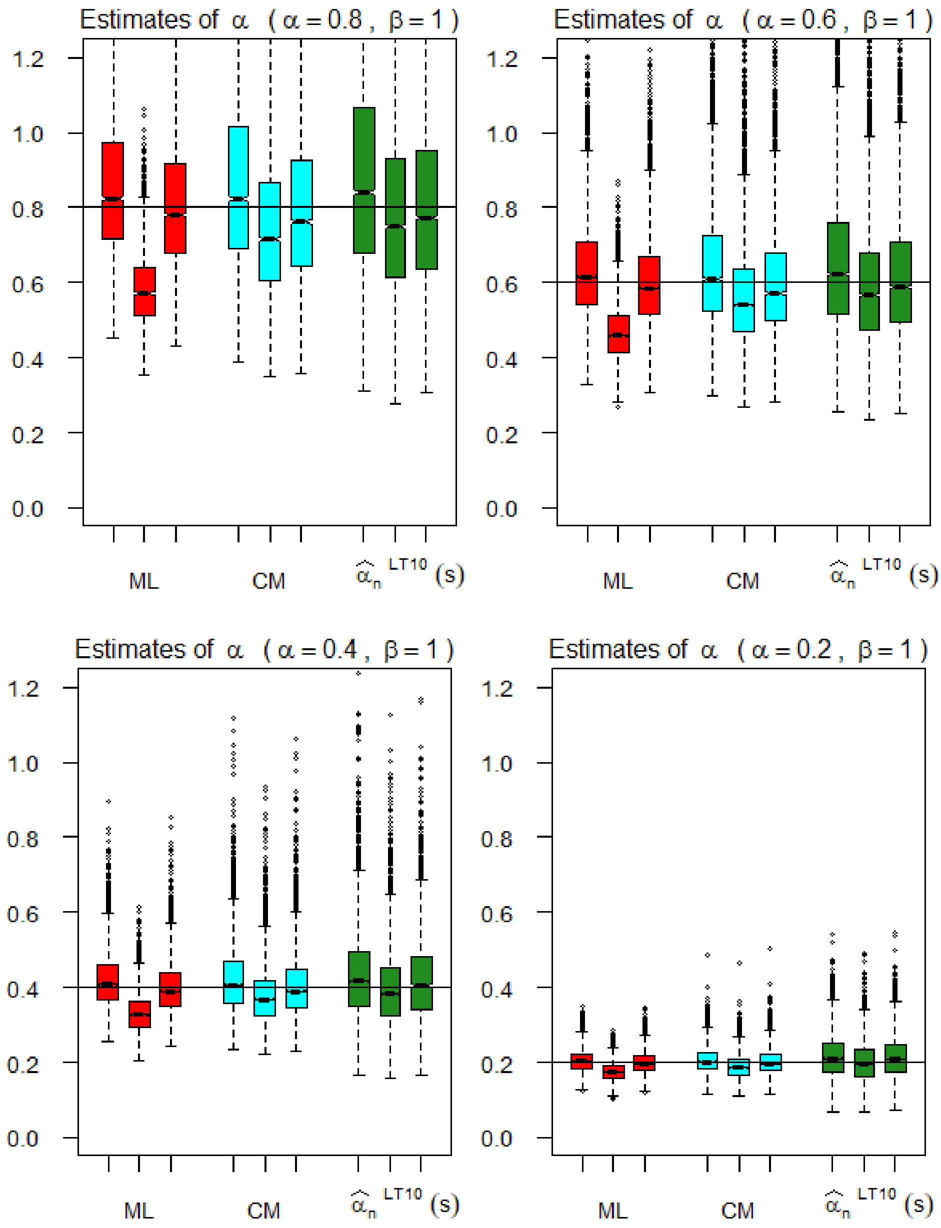


Figure 4.22: Estimation of the shape parameter α in the Pareto-Lomax distribution from 5000 generated samples of size $n = 70$, for four settings of $\alpha = 0.2, 0.4, 0.6, 0.8$; $\beta = 1.0$. Left box plot: pure sample; middle box plot: sample with two large outliers; right box plot: sample with two small outliers.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.6 Normal variance-gamma distribution

The final example in this chapter explores a case when the cumulative distribution and density functions are unknown, however, the corresponding LT or CF has an explicit and tractable form.

A prominent example possessing this property is the mixed normal variance-gamma distribution, detailed already in sec. 3.5.2. As the distribution is supported on the entire real line, we utilize its CF to formulate the estimators. Recall that the CF is expressed as

$$\phi(s) \;=\; \exp\{i\mu s\}\left(\frac{\beta}{\beta+s^2/2}\right)^{\alpha}, \tag{4.183}$$

and satisfies

$$(\beta+s^2/2)\phi'(s)-[i\mu(\beta+s^2/2)-\alpha s]\phi(s)=0\,, \tag{4.184}$$

where $i^2=-1$, $s,\mu\in\mathbb{R}$, $\alpha,\beta>0$.

4.5.6.1 Considered estimators

Epps and Pulley (1985), who studied this case, proposed estimation of the parameters directly from the CF expression (4.183), using the classical $\mathcal{L}^2$ estimator of the type (1.7) evaluated numerically. Our objective is to obtain explicit estimates, and for this reason, we work with (4.184). Unfortunately, using the equation, we cannot derive closed-form estimators of all three parameters due to the product of β with μ. But, note that, for modeling purposes, we may consider β as a constant[5].

We consider then the following $\mathcal{L}^2$ estimator of μ and α (as function of β):

$$\left(\widehat{\mu}_n^{\text{CF }1,0},\widehat{\alpha}_n^{\text{CF }1,0}\right)^{\top} = \operatorname*{argmin}_{\mu\in\mathbb{R},\,\alpha>0}\int_{-\infty}^{\infty}\left|(\beta+s^2/2)\widehat{\phi}_n'(s)-[i\mu(\beta+s^2/2)-\alpha s]\widehat{\phi}_n(s)\right|^2 e^{-as^2}\mathrm{d}s\,. \tag{4.185}$$

Since the underlying distribution has moments of all orders, Theorem 3 from sec. 4.2 implies that the estimator is consistent and asymptotically jointly normal.

Tedious but routine calculations yield the following formulas:

$$\widehat{\alpha}_n^{\text{CF }1,0} = \frac{2\sum\limits_{j,k=1}^{n}(X_j-X_k)^2\left\{\frac{1}{8a^2}\Big[(X_j-X_k)^2-6a\Big]-\beta\right\}\exp\Big(-(X_j-X_k)^2/4a\Big)}{\sum\limits_{j,k=1}^{n}\Big[(X_j-X_k)^2/a-2\Big]\exp\Big(-(X_j-X_k)^2/4a\Big)}, \tag{4.186}$$

[5]The mean and variance of the mixing gamma(α, β) random variable, say Y, are $\mathbb{E}[Y]=\alpha/\beta$ and $\mathbb{V}[Y]=\alpha/\beta^2$. Therefore, they may be controlled only via α if one sets $\beta=1$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

and

$$\widehat{\mu}_n^{\text{CF } 1,0} = \frac{\sum_{j,k=1}^{n} (X_j + X_k) A(X_j, X_k)}{\sum_{j,k=1}^{n} A(X_j, X_k)}, \tag{4.187}$$

where $A(X_j, X_k)$ equals

$$\left\{ \left[\beta^2 - \frac{\beta}{4a}\left[\frac{(X_j - X_k)^2}{a} - 2\right]\right] + \frac{1}{16a^2}\left[\frac{(X_j - X_k)^4}{4a^2} - \frac{3(X_j - X_k)^2}{a} + 3\right]\right\} \exp\left(-\frac{(X_j - X_k)^2}{4a}\right).$$

As in all the previous sections, to operationalize the estimator we need to estimate the value of $a > 0$ from the sample. Since the distribution in question is similar to the normal distribution, we suppose that a good trade-off between efficiency and robustness occur with $a = k\widehat{S}_n$, that is with a proportional to the sample variance where $k > 0$ remains user-controlled; recall sec. 4.5.1.3. As in the normal distribution case, this choice provides numerical stability of the estimators by preventing from the large exponents.

This case also provides an opportunity to discuss the CF-based estimator computed at single s-point for a two-sided distribution (the one not considered with the normal distribution in sec. 4.5.1). Upon decomposing the equation (4.184) into real and imaginary part, the estimator is defined as the solver of

$$\begin{cases} \text{Re}\Big\{(\beta + s^2/2)\widehat{\phi}_n'(s) - [i\mu(\beta + s^2/2) - \alpha s]\widehat{\phi}_n(s)\Big\} = 0\,, \\ \text{Im}\Big\{(\beta + s^2/2)\widehat{\phi}_n'(s) - [i\mu(\beta + s^2/2) - \alpha s]\widehat{\phi}_n(s)\Big\} = 0\,. \end{cases} \tag{4.188}$$

We obtain then

$$\widehat{\mu}_n(s) = \frac{\sum_{j,k=1}^{n} X_j \sin\Big(s(X_j - X_k)\Big)}{s \sum_{j,k=1}^{n} \cos\Big(s(X_j - X_k)\Big)}, \tag{4.189}$$

$$\widehat{\alpha}_n(s) = \frac{(\beta + s^2/2) \sum_{j,k=1}^{n} X_j \sin\Big(s(X_j - X_k)\Big)}{s \sum_{j,k=1}^{n} \cos\Big(s(X_j - X_k)\Big)}. \tag{4.190}$$

Although this estimator is expected to perform very poorly and be hard to control, it is interesting to note that $\widehat{\mu}_n(s)$ (as estimator of location) is independent of the other two parameters and that

$$\lim_{s \to 0} \widehat{\mu}_n(s) = \frac{1}{n} \sum_{j=1}^{n} X_j = \overline{X}_n\,. \tag{4.191}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

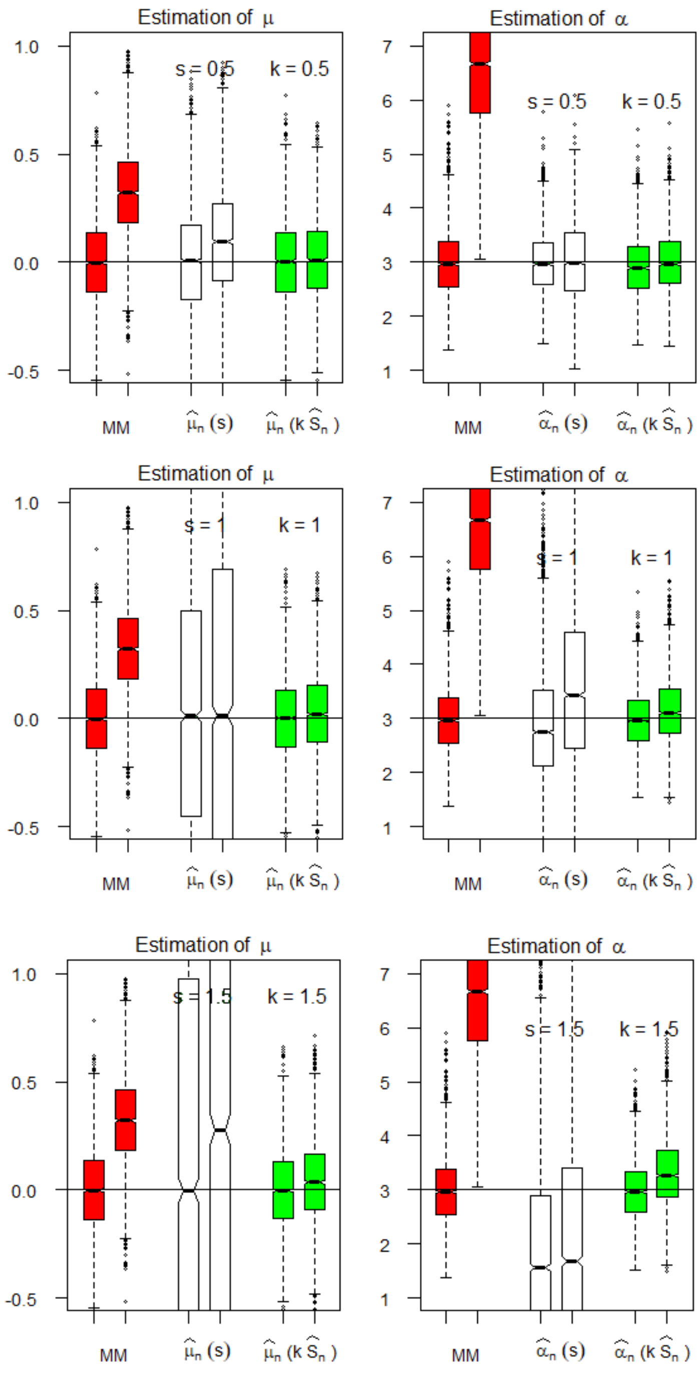


Figure 4.23: Estimation of $\mu = 0$ and $\alpha = 3$ in the normal variance-gamma mixed model for varying k and s. In each case $\beta = 1$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

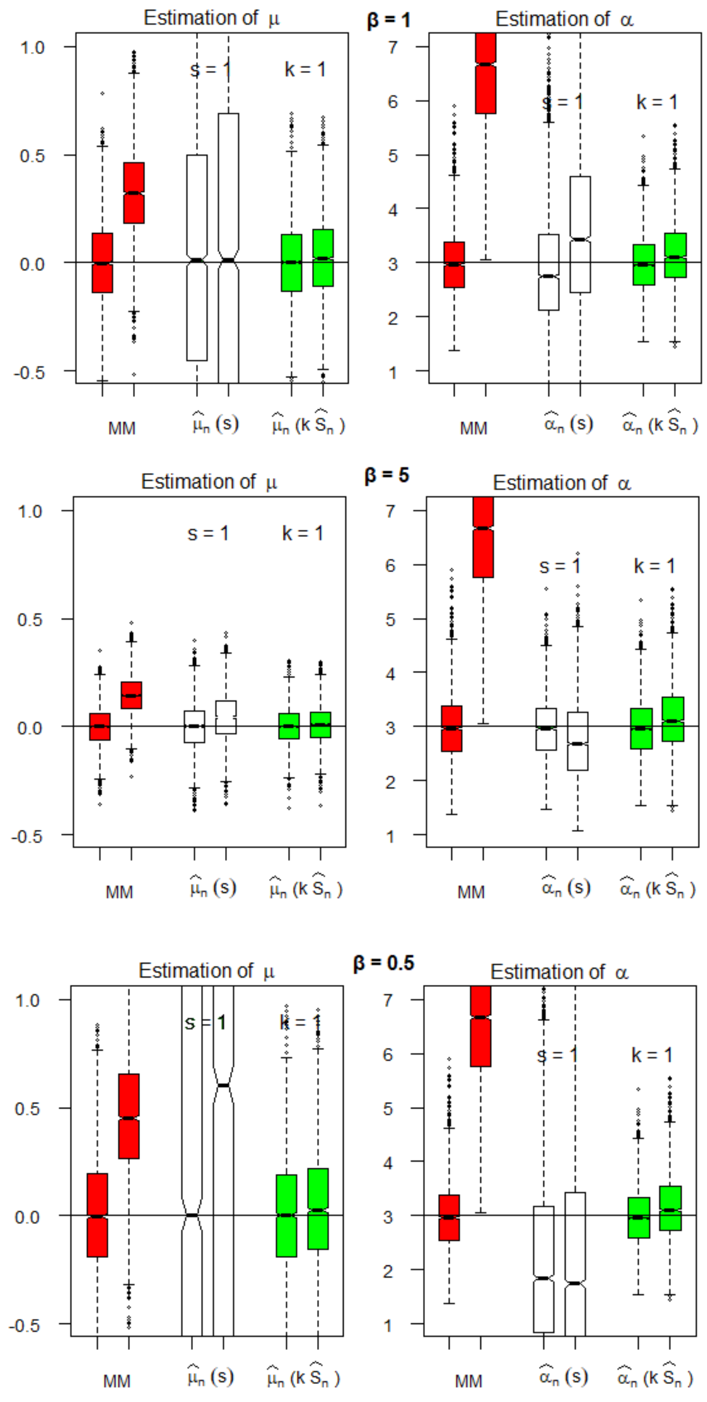


Figure 4.24: Estimation of $\mu = 0$ and $\alpha = 3$ in the normal variance-gamma mixed model for varying β. In each case $k = 1$ and $s = 1$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

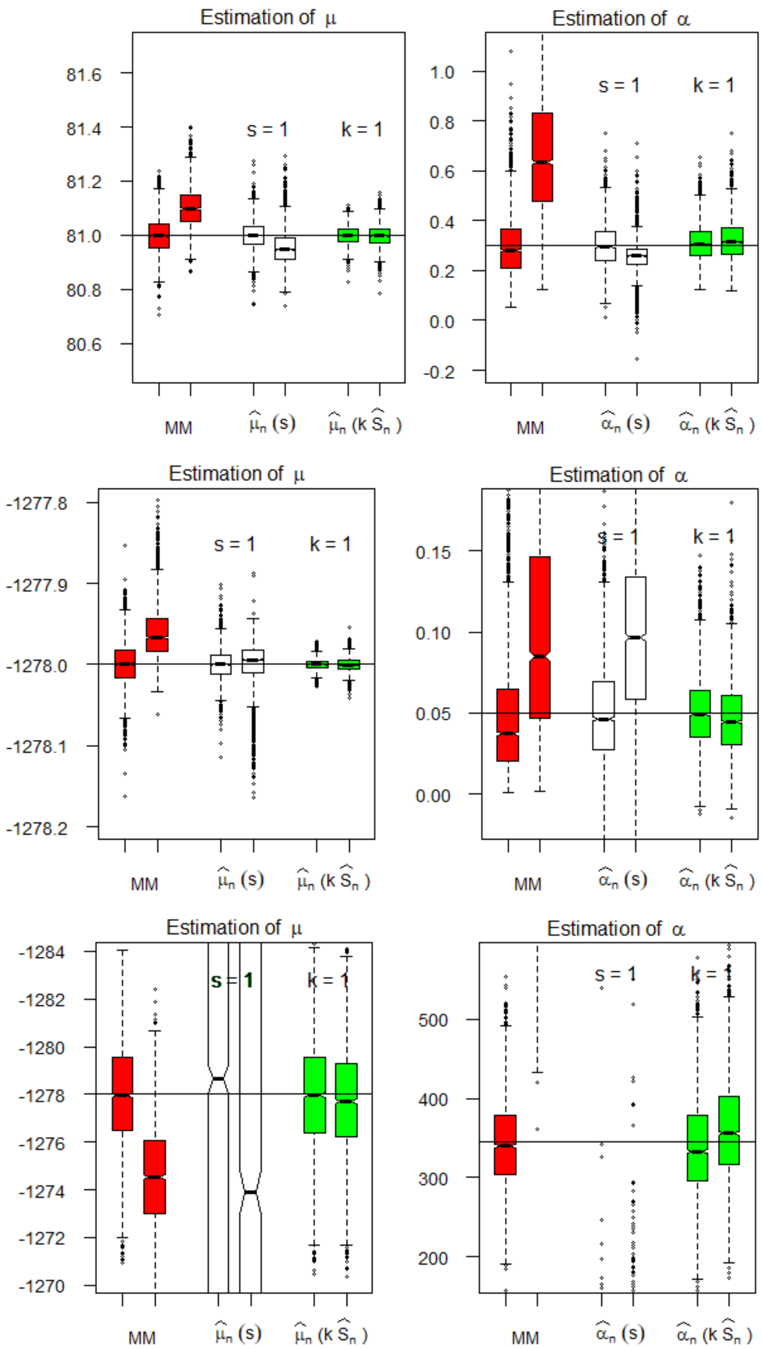


Figure 4.25: Estimation of $\mu = 81$ and $\alpha = 0.3$; $\mu = -1278$ and $\alpha = 0.05$; $\mu = -1278$ and $\alpha = 345$ in the normal variance-gamma mixed model. In each case $\beta = 1$, $k = 1$ and $s = 1$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

4.5.6.2 Finite-sample simulations

In simulations, we compare (4.186), (4.187) and (4.189), (4.190) with the method-of-moments estimators,

$$\widehat{\mu}_n^{\mathrm{MM}} = \overline{X}_n\,, \qquad \widehat{\alpha}_n^{\mathrm{MM}} = \beta \widehat{S}_n\,. \tag{4.192}$$

Figs. 4.23-4.25 show box plots of estimates drawn from 3000 simulated i.i.d. samples of size $n = 70$, for different settings of the parameters μ, α and β and the tuning constants s and k. (In the figures, the $\mathcal{L}^2$ estimators $\widehat{\mu}_n^{\mathrm{CF}\ 1,0}$ and $\widehat{\alpha}_n^{\mathrm{CF}\ 1,0}$ computed at $a = k\widehat{S}_n$ are denoted just by $\widehat{\mu}_n(k\widehat{S}_n)$ and $\widehat{\alpha}_n(k\widehat{S}_n)$).

Each time the left box plot represents estimation in pure samples, whereas the right one refers to samples contaminated by two large outliers. Concretely, the contamination follows by replacing the two first observations, X_1 and X_2, by $X_{0\text{-}1} = \overline{X}_n + 5\sqrt{S_n^2}$ and $X_{0\text{-}2} = \overline{X}_n + 7\sqrt{S_n^2}$.

We see that while the single s-point estimators are virtually useless, the sample-driven $\mathcal{L}^2$ estimators behave excellently. They have smaller or much smaller variance that the MM counterparts while retaining very high robustness against extreme outliers over wide range of parameters. The complicated construction process really pays off in this case.

4.6 Summary

This chapter integrated the theory of the proposed estimators with empirical studies, focusing on six selected distributions. While the examples could be expanded to encompass additional distributions in Tables 3.1 and 3.2, the current set already provides adequate material for drawing some general conclusions about the estimators' properties and performance.

Remark 1: Types of distributions, transforms and parameters

Utilizing both LT and CF, we can construct effective estimators for all types of continuous distributions, excluding only the two-sided (two-tailed) ones with infinite moments, such as the Cauchy or Student-t distributions. In these exceptional instances, the weighted $\mathcal{L}^2$ estimators lack consistency as $n \to \infty$ across the entire parameter space. The transform-based estimation of these distributions may become feasible by employing transforms with alternative kernels that would not impose restrictions on moments.

For two-sided light-tailed distributions (possessing all moments), the CF-based weighted $\mathcal{L}^2$ estimator is the suitable and only choice, as exemplified in cases like the normal and normal variance-gamma distributions. The weighted LT-based counterpart is not appropriate here due to unboundedness of the transform and its derivatives as $s \to \infty$. In the realm of one-sided (life-time) distributions, such as exponential, gamma or Rayleigh, more options are available. These can be successfully estimated using both the CF-

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

and LT estimators in the weighted $\mathcal{L}^2$ version, as well as using the single s-point LT-based estimator, whereby the former option also applies to distributions lacking certain moments, as demonstrated in the Pareto case.

We can also conclude that the proposed estimators are most appropriate for location and scale parameters. Namely, for all the studied location and location-scale families (normal, exponential, Rayleigh and normal variance-gamma) one or more of our proposals were competitive or outperforming the established estimators if the three criteria of relatively low variance, high robustness and explicitness were evaluated jointly. On the other hand, our estimators of shape parameters (α in the gamma and Pareto models), while robust, could be suboptimal in terms of variance in certain ranges.

Remark 2: Orders of differential equations

The factor we expected to play a role, namely the order of the employed differential equation, was indeed proven in simulations. However, the observations are not unambiguous. We initially anticipated that the orders should be kept as low as possible because involving higher derivatives should result in a larger variance of the empirical derivatives in certain regions of the s-domain.

As observed for the exponential distribution, this expectation holds true for the estimators based on CF. On the contrary, for the LT-based counterparts, higher orders need not degrade the estimators but may even increase robustness; recall the comparisons between $\widehat{\lambda}_n^{\text{LT }1,0}$, $\widehat{\lambda}_n^{\text{LT }2,1}$, and $\widehat{\lambda}_n^{\text{LT }2,0}$ in Fig. 4.12 and 4.13 for the exponential model, and between $\widehat{\vartheta^2}_n^{\text{LT }2,1,0}$ and $\widehat{\vartheta^2}_n^{\text{LT }3,2,1}$ in Fig. 4.19 for the Rayleigh case. It appears to be a rule that the LT differential equations not employing zero-order derivatives produce the most robust estimators.

Remark 3: Sample-driven weight functions

When dealing with the $\mathcal{L}^2$ estimators, our choice of the weight function was primarily guided by the simultaneous need to obtain a closed-form expressions and control over estimators properties. Therefore, no comparisons were made across weight functions of various functional forms. Nevertheless, during simulations, the author occasionally experimented by switching between the weights with Gaussian and exponential tail, both in the case of LT and CF-based estimators (with the estimates computed then numerically). Interestingly, the differences in results were marginal, leading to nearly identical box-plot outcomes. Provided that the weight function $w(s)$ was selected to tend to zero as $s \to 0$ and $s \to \infty$, it was rather its scale than shape that conditioned the performance. A crucial finding was that the scale should be pre-estimated by an auxiliary estimator, suggested by analysis of ARE and IFs. Among considered examples, such analyses were tractable for the normal, exponential, Rayleigh models. When these measures were not easily derivable, the auxiliary estimators could be heuristically determined through simulations and a trial-and-error principle (as we did in the gamma and Pareto cases).

However, finding the auxiliary estimator in either way may become increasingly challenging as the dimension of $\boldsymbol{\theta}$ grows. Then, and especially when working with real-world

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

datasets, where the model distribution typically only approximates the true generating mechanism, it may be more effective to determine the scale by assessing the model's predictive power, for example, through cross-validation. This should not be hard if the distribution function is available. If robust estimation is our goal, then, in the author's view, cross-validation should also employ a 'robust' loss criterion. A good candidate would be the Cramér-von Mises distance, chosen also for its relative simplicity. Specifically, if $\widehat{\boldsymbol{\theta}}_n^{\mathcal{T}}(a)$ is the transform-based estimator with the weight function parameter a, the k-fold cross-validatory choice of the latter can be

$$a_{\mathrm{CV}} = \underset{a>0}{\operatorname{argmin}} \sum_{i=1}^{k} \left\{ \frac{1}{12n_i} + \sum_{j_i=1}^{n_i} \left[F\Big(X_{(j_i)} \,|\, \widehat{\boldsymbol{\theta}}^{\mathcal{T}(-i)}(a) \Big) - \frac{2j-1}{2n} \right]^2 \right\}, \tag{4.193}$$

where the outer sum runs over k folds of the original sample, i.e. $(X_1, ..., X_{n_1})$, ..., $(X_1, ..., X_{n_k})$, $n_1 + ... + n_k = n$, j_i is the index within the i-th fold, and for each i the estimator $\widehat{\boldsymbol{\theta}}^{\mathcal{T}(-i)}(a)$ is drawn from all the observations without the i-th fold. The final estimate $\widehat{\boldsymbol{\theta}}_n^{\mathcal{T}}(a_{\mathrm{CV}})$ is drawn from the full sample; see Hastie et al. (2008).

If robustness is not of interest, then other loss criteria based on, for example, likelihood or Kullback-Leibler divergence can be used. The procedure would apply in the same manner for selecting the single s-point in the transform-matching estimators.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

CHAPTER 5

Goodness-of-fit test for the log-normal distribution

5.1 Transform-based testing

This chapter concerns goodness-of-fit (GoF) testing of the hypothesis that a random sample $X_1, ..., X_n$ comes from a specific parametric distribution. We develop a new test for the log-normal distribution based on LT. The test has been initially reported by the author in Hołyński (2019b).

As explained in sec. 1.1, the transform-based GoF tests utilize either

(i) a weighted $\mathcal{L}^2$ norm of the difference between the parametrically and non-parametrically estimated transform; or

(ii) a weighted $\mathcal{L}^2$ norm of empirical differential equation satisfied by the transform.

In context of LT of one-sided distributions, the test statistics corresponding to (i) and (ii) are, respectively:

$$T_n = n\int_0^\infty \left[L(s|\widehat{\boldsymbol{\theta}}_n) - \widehat{L}_n(s)\right]^2 \mathrm{d}W(s)\,, \tag{5.1}$$

and, for example,

$$T_n = n\int_0^\infty \left[\widehat{L}_n(s) + h(\widehat{\boldsymbol{\theta}}_n, s)\widehat{L}_n'(s)\right]^2 \mathrm{d}W(s)\,, \tag{5.2}$$

where $\widehat{\boldsymbol{\theta}}_n$ is some consistent estimator of $\boldsymbol{\theta}$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

A large realization of T_n leads to the rejection of the null-hypothesis $\mathcal{H}_0$. A non-trivial aspect of any GoF test is to determine the distribution of T_n under $\mathcal{H}_0$, say $F_0(t)$, which enables to reject $\mathcal{H}_0$ at a precise level of significance $1-\alpha$ (typically, $\alpha = 0.05$ or 0.1). Specifically, $\mathcal{H}_0$ is rejected if $T_n > F_0^{-1}(1-\alpha)$; otherwise, it is accepted. The value $F_0^{-1}(1-\alpha)$ is referred to as the *critical point*. The *power* $\mathcal{P}$ of the test against an alternative hypothesis (distribution) is the probability of rejecting $\mathcal{H}_0$ when the alternative is true. The test is said to be *consistent* (against any alternative) if $\mathcal{P} \to 1$ as $n \to \infty$ in the presence of any alternative.

5.2 Laplace transform property of log-normality

The log-normal random variable is characterized by the following distribution and density functions:

$$F(x|\mu,\sigma) = \frac{1}{2}\operatorname{erfc}\Big(-\frac{\log x - \mu}{\sigma\sqrt{2}}\Big)\,, \tag{5.3}$$

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right), \tag{5.4}$$

$x, \sigma > 0, \mu \in \mathbb{R}$. If we wish to test for log-normality using (5.1) or (5.2), two problems arise. First, the log-normal LT,

$$L(s|\mu,\sigma) = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)\exp(-sx)\mathrm{d}x\,, \tag{5.5}$$

has no closed-form expression. Although various approximations were proposed for this integral, they are either difficult to implement or not uniformly accurate in s; see Asmussen et al. (2016) and the references therein. Consequently, computation of the integral in (5.1) is troublesome. Secondly, differentiating the density w.r.t. x results in the linear differential equation

$$(\log x - \mu + \sigma^2)f(x|\mu,\sigma) + \sigma^2 x f'(x|\mu,\sigma) = 0, \tag{5.6}$$

whose coefficients are not polynomials in x. Therefore, we cannot apply the method from sec. 3.2 to obtain the differential equation satisfied by the transform and construct the statistic of the type (5.2).

However, we can obtain a *functional differential equation* by taking derivative w.r.t. s in (5.5) under integration sign.

Theorem 18. *Let X be a log-normal random variable with density (5.4). The corresponding Laplace transform $L(s|\mu,\sigma)$ satisfies the following k-th order functional differential equation*

$$L^{(k)}(s|\mu,\sigma) - (-1)^k\exp\Big(k\mu + (k\sigma)^2/2\Big)L(se^{k\sigma^2}|\mu,\sigma) = 0\,. \tag{5.7}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Proof. We know that $X = e^Y$, where $Y \sim \text{Normal}(\mu, \sigma^2)$. It follows then that $X = e^{Y_0+\mu}$, where $Y_0 \sim \text{Normal}(0, \sigma^2)$, and, by the property of expectation,

$$L(s|\mu,\sigma) = \mathbb{E}[e^{-sX}] = \mathbb{E}[\exp(-se^Y)] = \mathbb{E}[\exp(-se^{Y_0+\mu})] = \mathbb{E}[\exp(-se^{\mu}e^{Y_0})]\,. \tag{5.8}$$

Therefore, for any log-normal random variable, we can write

$$L(s|\mu+b,\sigma) = L(se^b|\mu,\sigma)\,, \qquad b \in \mathbb{R}. \tag{5.9}$$

Knowing this, let us substitute $y = \log x$ in (5.5); we obtain

$$L(s|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\Big(-se^y - \frac{(y-\mu)^2}{2\sigma^2}\Big)\mathrm{d}y\,. \tag{5.10}$$

Taking derivative w.r.t. s of both sides gives

$$\begin{aligned} L'(s|\mu,\sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \frac{\partial}{\partial s}\exp\Big(-se^y - \frac{(y-\mu)^2}{2\sigma^2}\Big)\mathrm{d}y \\ &= -\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\Big(-se^y + y - \frac{(y-\mu)^2}{2\sigma^2}\Big)\mathrm{d}y\,. \end{aligned} \tag{5.11}$$

The expression in the exponent of the integrand can be written as

$$y - (y-\mu)^2/2\sigma^2 = -(y-\mu-\sigma^2)^2/2\sigma^2 + [(\mu+\sigma^2)^2 - \mu^2]/2\sigma^2\,, \tag{5.12}$$

and this gives

$$L'(s|\mu,\sigma) = -\exp(\mu+\sigma^2/2)\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\Big(-se^y - \frac{(y-\mu-\sigma^2)^2}{2\sigma^2}\Big)\mathrm{d}y\,. \tag{5.13}$$

The integrand above differs from the integrand in (5.10) only by the term $-\sigma^2$. Hence, in consideration of (5.9), (5.7) holds for $k = 1$, namely

$$L'(s|\mu,\sigma) = -\exp\Big(\mu+\sigma^2/2\Big)L(se^{\sigma^2}|\mu,\sigma)\,. \tag{5.14}$$

Taking k-th derivative in (5.10) and repeating the steps yields (5.7) for $k \geq 2$. The proof follows Leipnik (1991), who derived a similar equation satisfied by the log-normal CF. □

Ideally, for the purpose of GoF testing, the converse of Theorem 18 should also hold. In other words, the LT of the log-normal distribution should be the only solution of (5.7) with the boundary condition $L(0|\mu,\sigma) = 1$. This would elevate (5.7) from just being a *property* to a *characterization* of that distribution, leading to a test consistent against any alternative. Unfortunately, it is not the case. Kato and McLeod (1971), who studied

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

general solutions of this type of functional differential equations, proved that (5.7) with $k = 1$ has an infinite number of solutions for $s \in (0, \infty)$. Thus, we cannot be sure that the proposed test is consistent. Nonetheless, we may suppose that within the narrower class of LTs of one-sided probability distributions, this property characterizes a sufficiently rich subclass of functions to make the test consistent against standard distributions encountered in applied statistics.

5.3 The test statistic

Our test will utilize the aforementioned property only with $k = 1$. Notice that to test for the log-normal(μ,σ) distribution, it suffices to test for the log-normal(0,1) distribution after standardizing the sample using (some) estimators of μ and σ. Consequently, the form of (5.7) we use hereafter simplifies to

$$L'(s|0,1) + e^{1/2} L(se|0,1) = 0\,. \tag{5.15}$$

To test $\mathcal{H}_0$ that $X_1, ..., X_n \sim F(x|\mu, \sigma)$ with unspecified parameters, we compute

$$T_n = n \int_0^\infty \left[\widehat{L}'_n(s) + e^{1/2}\widehat{L}_n(se)\right]^2 w(s)\mathrm{d}s\,, \tag{5.16}$$

where $\widehat{L}'_n(s) = \frac{1}{n}\sum_{j=1}^n \exp(-sY_j)$, $\widehat{L}'_n(s) = -\frac{1}{n}\sum_{j=1}^n Y_j \exp(-sY_j)$, $Y_j = \exp\{(\log(X_j) - \widehat{\mu}_n)/\widehat{\sigma}_n\}$, $w(s)$ is a weight function and $\widehat{\mu}_n$ and $\widehat{\sigma}_n$ are the ML estimators,

$$\widehat{\mu}_n = \frac{1}{n}\sum_{j=1}^n \log(X_j)\,, \qquad \widehat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{j=1}^n [\log(X_j) - \widehat{\mu}_n]^2}\,. \tag{5.17}$$

5.3.1 Choice of estimators and weight function

Naturally, we could employ other estimators to standardize the sample. Specifically, we might use the LT-based estimators that minimize the same $\mathcal{L}^2$ distance utilized by the test,

$$\left(\widehat{\mu}_n^{\mathrm{LT}}, \widehat{\sigma}_n^{\mathrm{LT}}\right)^\top = \underset{\substack{\mu \in (-\infty,\infty) \\ \sigma > 0}}{\operatorname{argmin}} \int_0^\infty \left[\widehat{L}'_n(s) + \exp(\mu + \sigma^2/2)\widehat{L}_n(se^{\sigma^2})\right]^2 w(s)\mathrm{d}s\,. \tag{5.18}$$

However, this is not a good idea due to numerical problems associated with large exponents. While these estimators might exhibit robustness against outliers, our current objective is not to create a robust procedure that disregards outliers. Rather, the proposed test should be sensitive to deviations from log-normality, particularly in the tail region. Hence, we decide to use the non-robust but efficient and easily computable ML estimators.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Following the LT-based tests for one-sided distributions, such as Henze (1992), Henze et al. (2012) or Henze and Klar (2002), we equip the test statistic with the exponential weight function,

$$w(s) = e^{-as}, \qquad a > 0. \tag{5.19}$$

Utilizing the Tauberian relations for LTs, the user-chosen rate a controls the power against specific distributional alternatives. A larger a should render the test more powerful against the alternatives that differ from the log-normality in the tail, while a smaller a should sensitize it to the discrepancies near the origin. Additionally, this weight function provides a closed-form expression for (5.16), which we denote by $T_n(a)$ to stress its dependence on the constant a. After some algebra, we obtain

$$T_n(a) = \frac{1}{n}\sum_{j,k=1}^{n}\left(\frac{e}{e(Y_j+Y_k)+a} - \frac{e^{1/2}Y_k}{eY_j+Y_k+a} - \frac{e^{1/2}Y_j}{eY_k+Y_j+a} + \frac{Y_jY_k}{Y_j+Y_k+a}\right). \tag{5.20}$$

5.3.2 Distribution of the test statistic

We shall discuss now $F_0(t)$, the distribution of T_n under null-hypothesis $\mathcal{H}_0$. Both the finite-sample and asymptotic distribution of the statistic are highly non-trivial. It is evident that T_n represents a (normalized) degree-2 V-statistic with the kernel

$$\psi(Y_j, Y_k) = \int_0^\infty \left(e^{1/2-seY_j} - Y_j e^{-sY_j}\right)\left(e^{1/2-seY_k} - Y_k e^{-sY_k}\right) w(s)\mathrm{d}s\,, \tag{5.21}$$

and, as a consequence, its asymptotic distribution is the same as that of the corresponding U-statistic; see Appendix A. Since $\mathbb{E}[\psi(y_j, Y_k)] = 0$, the first variance component of T_n computed by (A.5) is $\sigma_1^2 = 0$. However, the second component $\sigma_2^2 > 0$. Hence, this U-statistic has first-order degeneracy and

$$nT_n \xrightarrow{\mathrm{d}} \sum_{j=1}^{\infty} \lambda_j Z_j^2\,, \tag{5.22}$$

where $Z_1, Z_2, ...$ are independent unit normal variables, and the λ_j are the eigenvalues of the integral equation (A.17). Unfortunately, solution of this eigenproblem in our case seems impossible. Even if λ_j could be found, obtaining $F_0(t)$ would necessitate a numerical inversion of CF or LT of the infinite mixture of chi-squared variables. Finally, there is no guarantee that this asymptotic distribution would be accurate enough to give a good approximation of the critical points for small samples.

Another approach would be to approximate the finite sample version of $F_0(t)$ using some standard skewed distribution through the moment-matching technique. However, computing, the mean and variance of T_n (let alone the higher moments) involves intricate integrals that depend on the unknown log-normal LT (5.5).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

5.4 Simulations of the power

We observe that the challenge of determining an exact form of $F_0(t)$ remains unsolved for most of the transform-based GoF tests in the literature. A common solution is to either simulate $F_0(t)$ or apply parametric bootstrap to estimate critical points; see, for instance, Henze and Klar (2002), Meintanis (2005), Zghoul (2010), Henze et al. (2012) and Batsidis et al. (2020). Given that we are working with a skewed distribution, we opt for the latter approach. Below, we describe how the bootstrap is used in simulating the power of the test.

5.4.1 Parametric bootstrap

We generate r independent samples $\boldsymbol{X} = (X_1, ..., X_n)$ from an alternative distribution such as gamma, exponential, inverse Gaussian and so on. To estimate the power of our test against this alternative (for given n and a significance level of $1-\alpha$), we perform these steps:

1. For each of the r samples $\boldsymbol{X}$, compute the ML estimators $\widehat{\mu}_n$ and $\widehat{\sigma}_n$ using (5.17). Standardize the sample to obtain $\boldsymbol{Y} = \{Y_1, ..., Y_n\}$, where $Y_i = \exp\{(\log(X_i) - \widehat{\mu}_n)/\widehat{\sigma}_n\}$.

2. For each of the r standardized samples $\boldsymbol{Y}$, compute $T_n(a)$ using (5.20).

3. For each $\boldsymbol{Y}$, generate B bootstrap samples $\boldsymbol{X}^{*(1)}, ..., \boldsymbol{X}^{*(j)}, ..., \boldsymbol{X}^{*(B)}$ of size n from the log-normal distribution with parameters equal to the corresponding estimates $\widehat{\mu}_n$ and $\widehat{\sigma}_n$.

4. For each j-th bootstrap sample $\boldsymbol{X}^{*(j)}$ compute the ML estimators $\widehat{\mu}_n^{*(j)}$, $\widehat{\sigma}_n^{*(j)}$ using (5.17) and standardize them taking $Y_i^{*(j)} = \exp\{(\log(X_i^{*(j)}) - \widehat{\mu}_n^{*(j)})/\widehat{\sigma}_n^{*(j)}\}$. In this way, obtain B standardized bootstrap samples $\{\boldsymbol{Y}^{*(1)}, ..., \boldsymbol{Y}^{*(j)}, ..., \boldsymbol{Y}^{*(B)}\}$ and for each j-th of them compute the test statistic $T_n^{*(j)}(a)$ using (5.20). The order statistic $T^*_{[(1-\alpha)B]}(a)$ of so obtained vector estimates the critical value of $T_n(a)$ at significance level of $1-\alpha$. If $T_n(a) > T^*_{[(1-\alpha)B]}(a)$, reject $\mathcal{H}_0$; otherwise, retain it.

The power $\mathcal{P}$ of the test against an alternative for sample size n is then estimated by

$$\widehat{\mathcal{P}}_n = \frac{\#\{\mathcal{H}_0 \ \text{ rejected}\}}{r}. \tag{5.23}$$

5.4.2 Competitive tests

In simulations, our test is compared in terms of the power with the four classical GoF procedures:

- Cramér-von Mises (CM) test,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

- Anderson-Darling (AD) test,
- Shapiro-Wilk (SW) test,
- Jarque-Bera (JB) test.

The CM and AD tests utilize distribution function and are consistent against any alternative. The SW and JB tests are famous tests for normality that can be applied to our problem after log-transformation of the sample. They are inconsistent but powerful. The CM and AD test statistics, mentioned already in the introduction chapter, can also be expressed as

$$T_n^{\mathrm{CM}} = \frac{1}{12n} + \sum_{j=1}^{n} \left(F(X_{(j)}|\widehat{\mu}_n, \widehat{\sigma}_n) - \frac{2j-1}{2n} \right)^2, \tag{5.24}$$

and

$$T_n^{\mathrm{AD}} = -n - \frac{1}{n} \sum_{j=1}^{n} \Big((2j-1) \log F(X_{(j)}|\widehat{\mu}_n, \widehat{\sigma}_n) + [2(n-j)+1] \log[1 - F(X_{(j)}|\widehat{\mu}_n, \widehat{\sigma}_n)] \Big), \tag{5.25}$$

where $\widehat{\mu}_n$ and $\widehat{\sigma}_n$ are, for example, the ML estimators (5.17).

The SW statistic, used to test for log-normality, is given by

$$T_n^{\mathrm{SW}} = \frac{\left(\sum\limits_{j=1}^{n} c_{j,n} Y_{(j)} \right)^2}{\sum\limits_{j=1}^{n} (Y_j - \overline{Y}_n)^2}, \tag{5.26}$$

where $Y_j = \log X_j$ and the coefficients $c_{j,n}$ can be found in D'Agostino and Stephens (1986).

Finally, the JB statistic is

$$T_n^{\mathrm{JB}} = \frac{n}{6} \left(\widehat{Sk}_n^2 + \frac{1}{4} (\widehat{K}_n - 3)^2 \right), \tag{5.27}$$

where $\widehat{Sk}_n = \dfrac{\frac{1}{n} \sum_{j=1}^{n} (Y_j - \overline{Y}_n)^3}{\left(\frac{1}{n} \sum_{j=1}^{n} (Y_j - \overline{Y}_n)^2 \right)^{3/2}}$ and $\widehat{K}_n = \dfrac{\frac{1}{n} \sum_{j=1}^{n} (Y_j - \overline{Y}_n)^4}{\left(\frac{1}{n} \sum_{j=1}^{n} (Y_j - \overline{Y}_n)^2 \right)^2}$

are the sample skewness and sample kurtosis, respectively, and $Y_j = \log X_j$.

For more details on all the four tests see, e.g., D'Agostino and Stephens (1986), Thas (2010), Shapiro and Wilk (1965), and Jarque and Bera (1987). In simulations, the critical points for the CM and AD tests were also determined using bootstrap. The SW and JB tests were conducted by using the shapiro.test(·) and jb.norm.test(·) functions from the R-packages stats and normtest, respectively.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

5.4.3 Alternative distributions

In the simulations, we examine the power of the four tests against several two-parameter life-time distributions[1]. These alternatives include Weibull(λ, c), gamma(α, β), Pareto(α, β) and inverse Gaussian(λ, μ) with densities given in Table 2.1; the other alternatives are:

- Fréchet(λ, c): $\quad f(x|\lambda, c) = \dfrac{c\lambda^c}{x^{c+1}} \exp\{-(\lambda x)^{-c}\}, \quad x, c, \lambda > 0;$
- Burr(α, c): $\quad f(x|\alpha, c) = \dfrac{\alpha c x^{c-1}}{(1+x^c)^{\alpha+1}}, \quad x, \alpha, c > 0;$
- Inv. gamma(α, β): $\quad f(x|\alpha, \beta) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\{-\beta/x\}, \quad x, \alpha, \beta > 0;$
- Gompertz(α, β): $\quad f(x|\alpha, \beta) = \alpha\beta \exp\{\alpha + \beta x - \alpha e^{\beta x}\}, \quad x \geq 0, \;\; \alpha, \beta > 0.$

To verify whether the tests accurately recover the nominal level of significance $1-\alpha$, we conduct simulations of power for log-normal samples, expecting $\widehat{\mathcal{P}}_n \approx 5\%$ for any sample size n.

5.4.4 Results

The simulations were conducted for sample sizes $n = 20, 30, 50, 70, 100$, with $\alpha = 0.05$, for all five tests. All the alternative distributions were considered in four distinct settings of their parameters. The power of our test was simulated for six values of the weight function parameter a, specifically $a = 0.01, 0.1, 0.5, 1, 2, 5, 10$. In the bootstrap procedure, we used $r = 3000$ and $B = 200$. The results are reported by the graphs placed in Tables 5.1-5.5, where the estimated power $\widehat{\mathcal{P}}_n$ is plotted as function of n. The power of our test is plotted for the value of a that provided the highest average power across sample sizes from $n = 20$ to 100. The powers estimated for the log-normal samples are approximately 0.05 ($= \alpha$) for any n, indicating that all the tests are well calibrated and implemented correctly.

We observe that the performance of our test is very satisfactory in comparison with all the competitors. It usually has higher or slightly higher power than the SW and JB tests and much higher than AD and CM (even by around 20% for $n = 100$). Exceptions do occur in the case of the Burr and Pareto alternatives where JB appears more powerful for two combinations of parameters (by about 5% in large samples). Of course, this performance is so good with the optimal choice of a (which may be unknown when testing against any alternative). However, notice that in nearly all cases, the optimal a is very low, i.e. $a = 0.01$ and 0.1. This is not surprising, as most of the alternatives

[1]The parameter α of these distributions should not be confused with α denoting here also the test level significance.

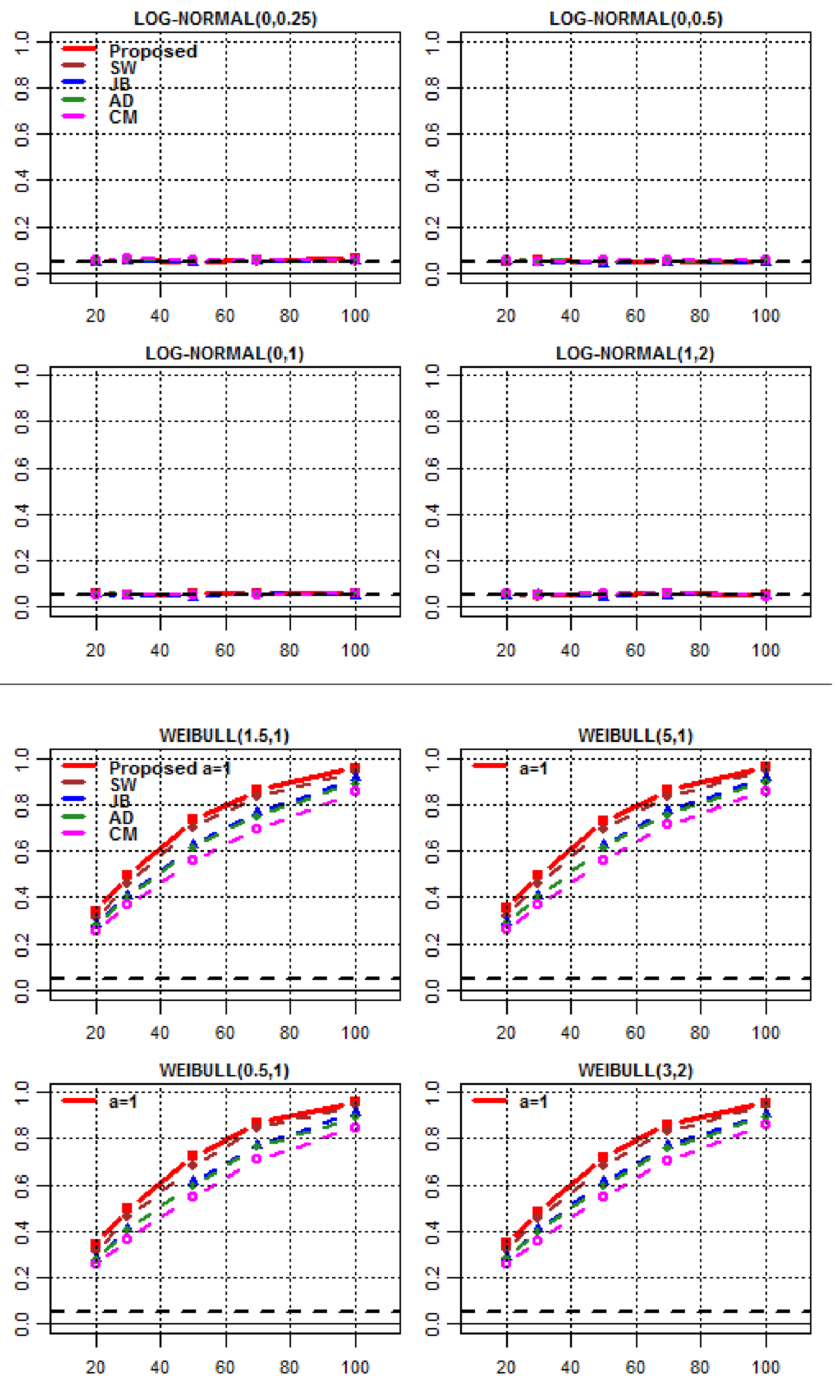
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

here differ from log-normality in the tail area (rather than in shape as $x \to 0$), and, as discussed earlier, the choice of a small a should enhance the test's power in this regard.

We should still check if and how the power drops when a suboptimal a is chosen. Table 5.6 displays powers (rounded to second decimal place) for all a in the case of $n = 50$. We see that the powers do not drop by more than 5% for the worst choices (the Fréchet and inverse gamma cases). For most alternatives, they remain at a fairly stable level. This feature distinguishes our proposal from some of the existing transform-based tests where the choice of a is rather crucial and typically changes the power by 10 to 40%; see the power plots in Baringhaus et al. (2000) and Tenreiro (2009) demonstrating this fact for several LT- and CF-based tests.

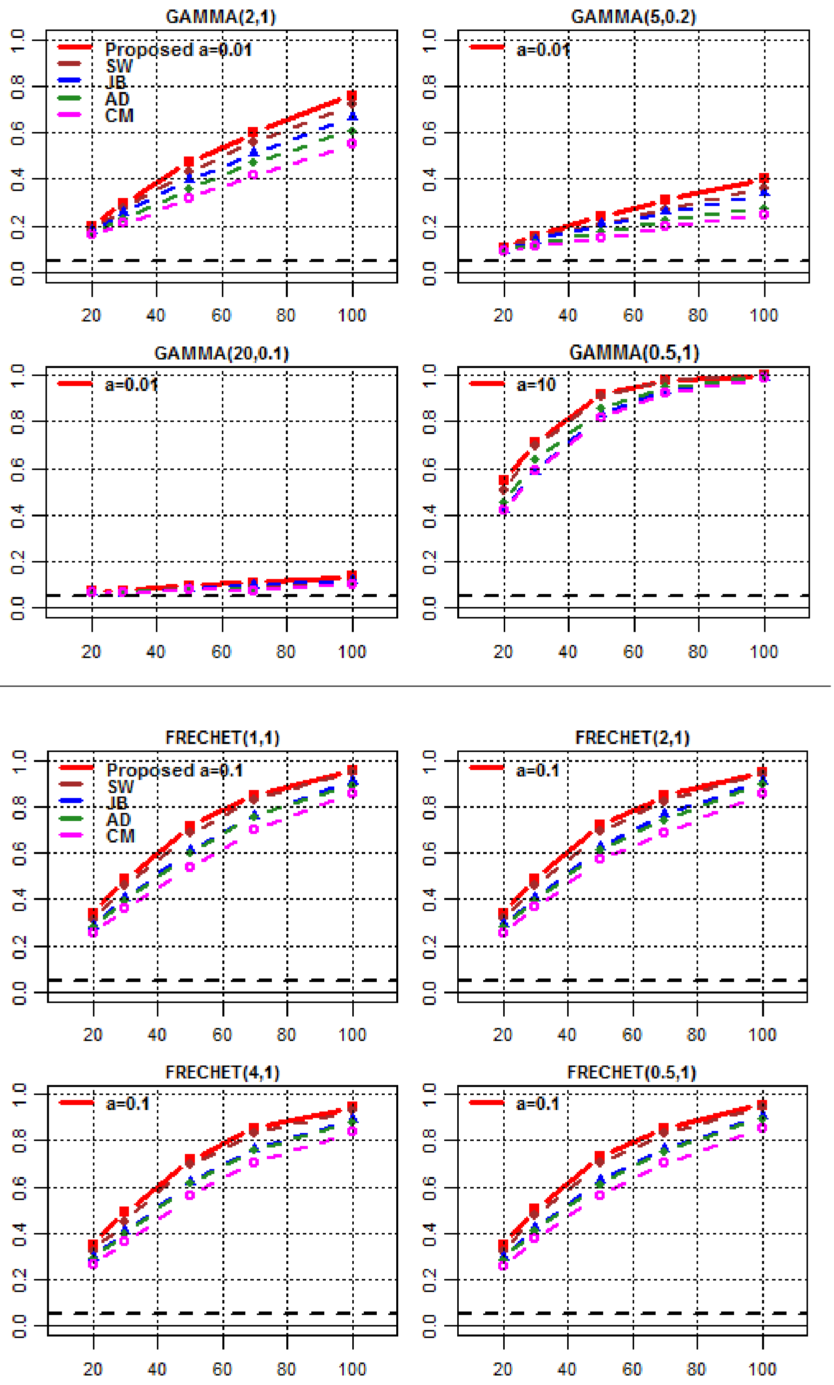
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Table 5.1: Estimated power $\widehat{\mathcal{P}}_n$ against the log-normal(μ, σ) and Weibull(λ, c) alternatives.



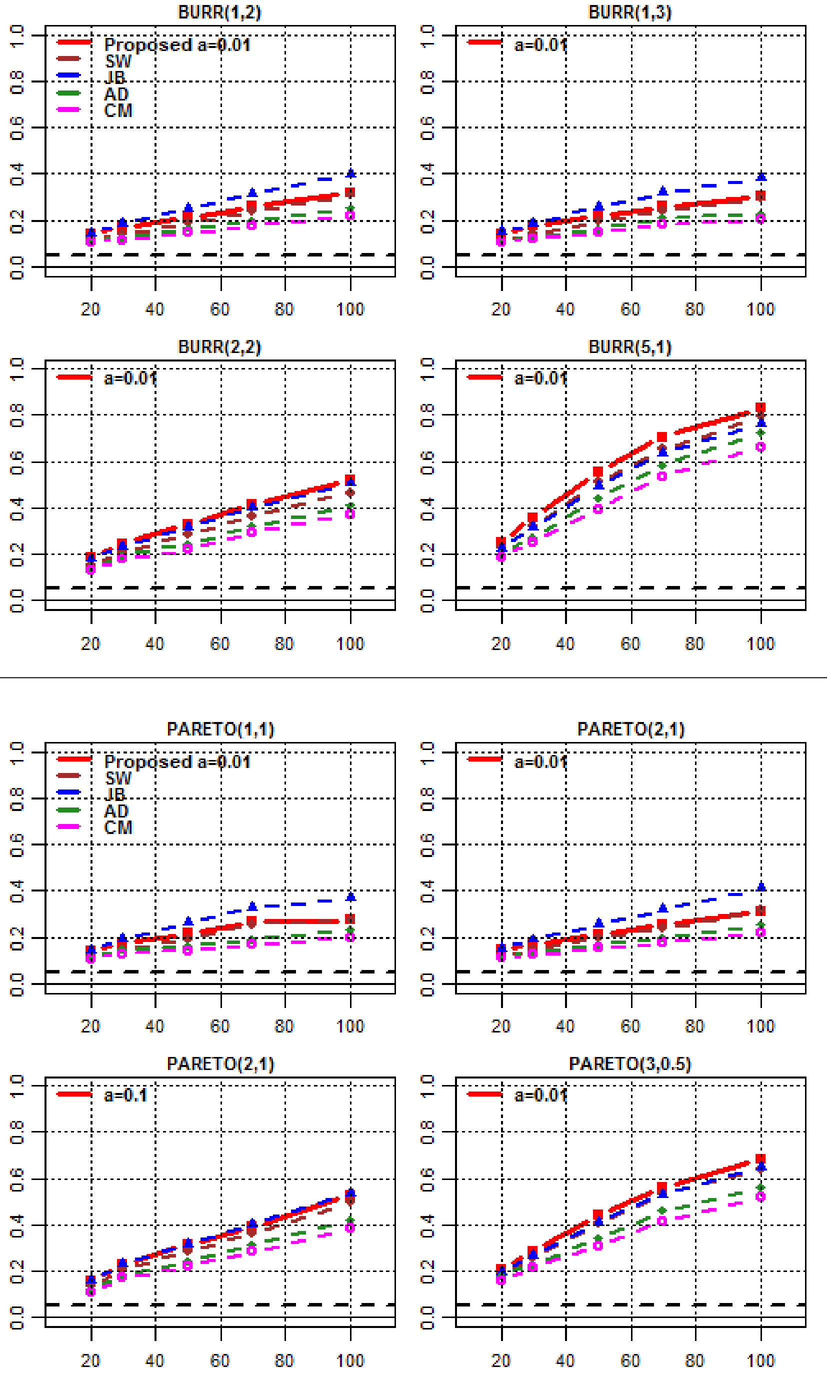
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Table 5.2: Estimated power $\widehat{\mathcal{P}}_n$ against the gamma(α, β) and Fréchet(λ, c) alternatives.



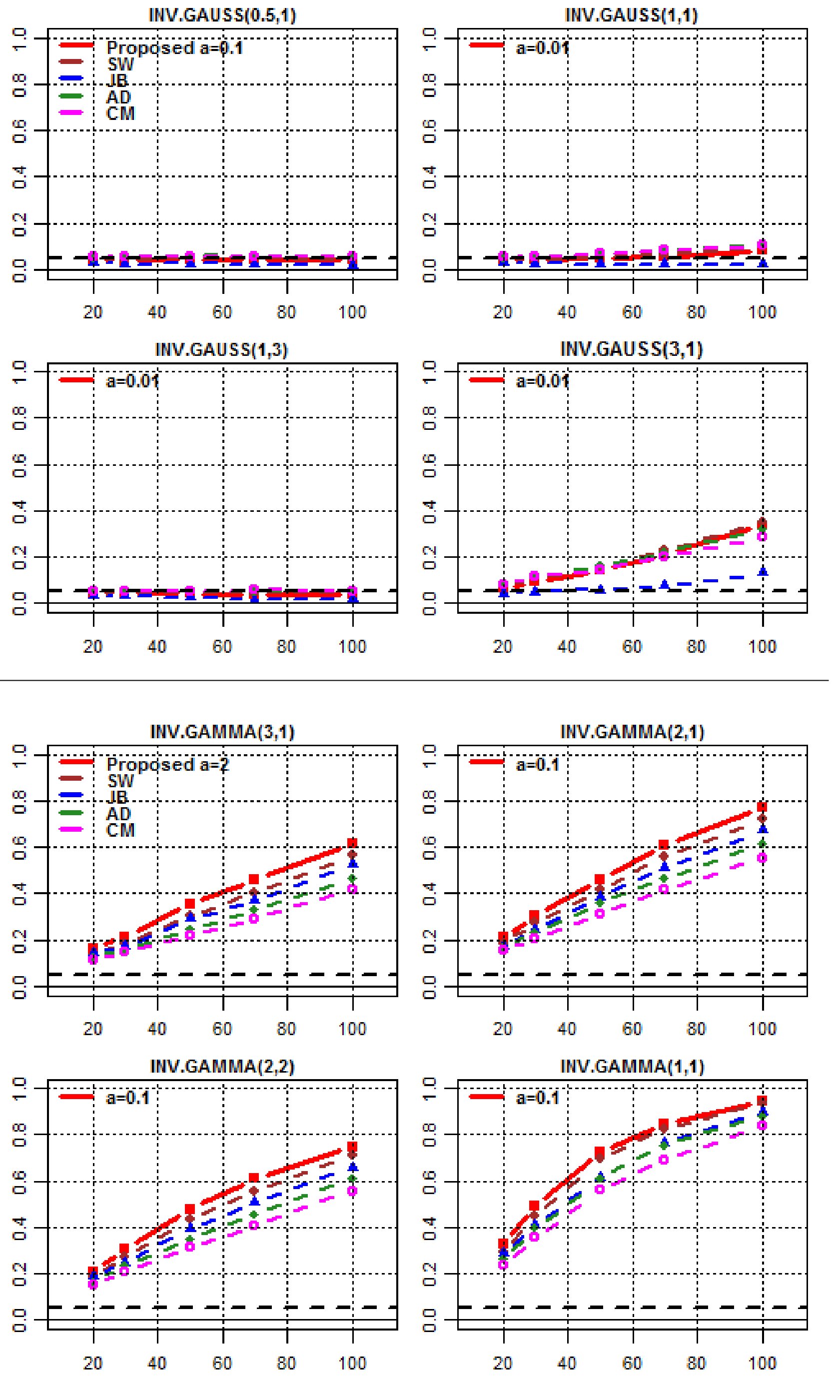
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Table 5.3: Estimated power $\widehat{\mathcal{P}}_n$ against the Burr(α, c) and Pareto(α, β) alternatives.



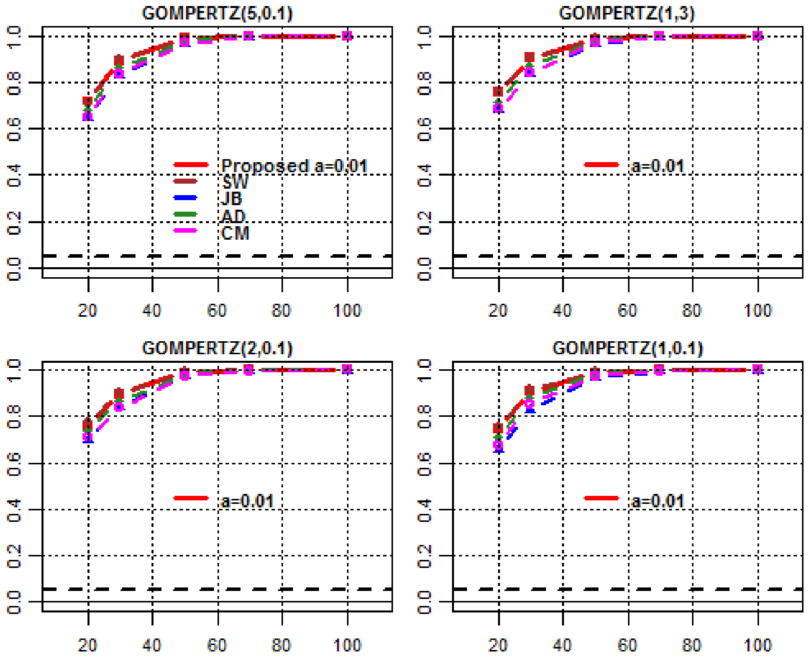
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Table 5.4: $\widehat{\mathcal{P}}_n$ against the inv. Gauss.(λ, μ) and inv. gamma(α, β) alternatives.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Table 5.5: Estimated power $\widehat{\mathcal{P}}_n$ against the Gompertz(α, β) alternative.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Table 5.6: Estimated power $\widehat{\mathcal{P}}_n$ of the proposed test in function of the weight function parameter a for $n = 50$. Comparison with the Shapiro-Wilk and Jarque-Bera tests.

alternative	a value 0.01	0.1	0.5	1.0	2.0	5.0	10	SW	JB
log-normal(0,0.25)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
log-normal(0,0.5)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04
log-normal(0,1)	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.04	0.04
log-normal(1,2)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04
Weibull(1.5,1)	**0.73**	0.72	**0.73**	**0.73**	**0.73**	0.72	0.70	0.70	0.62
Weibull(5,1)	0.72	0.72	**0.73**	**0.73**	0.72	0.72	0.70	0.69	0.62
Weibull(0.5,1)	0.71	0.71	**0.72**	**0.72**	0.71	0.70	0.70	0.68	0.61
Weibull(3,2)	0.71	0.71	0.71	**0.72**	**0.72**	0.70	0.70	0.68	0.61
gamma(2,1)	**0.47**	**0.47**	0.45	0.46	0.46	0.44	0.44	0.43	0.40
gamma(5,0.2)	**0.24**	0.23	0.23	0.23	0.22	0.20	0.21	0.21	0.20
gamma(20,0.1)	**0.10**	0.09	0.09	0.09	0.08	0.08	0.08	0.08	0.08
gamma(0.5,1)	0.91	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	0.91	0.83
Fréchet(1,1)	0.71	**0.72**	0.70	0.69	0.68	0.67	0.65	0.68	0.61
Fréchet(2,1)	**0.73**	0.72	0.71	0.70	0.70	0.68	0.67	0.69	0.63
Fréchet(4,1)	**0.72**	**0.72**	0.71	0.70	0.70	0.68	0.66	0.70	0.62
Fréchet(0.5,1)	0.72	**0.73**	0.71	0.70	0.70	0.68	0.67	0.70	0.63
Burr(1,2)	0.21	0.21	0.21	0.20	0.19	0.19	0.20	0.19	**0.25**
Burr(1,3)	0.22	0.22	0.21	0.21	0.20	0.20	0.21	0.19	**0.26**
Burr(2,2)	**0.33**	0.31	0.29	0.30	0.28	0.26	0.27	0.28	0.32
Burr(5,1)	**0.55**	0.54	0.53	0.53	0.52	0.51	0.52	0.51	0.49
Pareto(1,1)	0.22	0.22	0.21	0.21	0.20	0.21	0.20	0.19	**0.26**
Pareto(1,3)	0.21	0.21	0.20	0.21	0.19	0.19	0.19	0.20	**0.26**
Pareto(2,1)	**0.32**	**0.32**	0.30	0.29	0.28	0.26	0.27	0.29	**0.32**
Pareto(3,0.5)	**0.45**	0.44	0.43	0.42	0.41	0.39	0.40	0.40	0.41
inv. gamma(3,1)	0.34	**0.35**	**0.35**	**0.35**	**0.35**	**0.35**	0.34	0.30	0.29
inv. gamma(2,1)	**0.46**	**0.46**	0.45	0.45	0.45	0.45	0.43	0.41	0.39
inv. gamma(2,2)	**0.48**	**0.48**	0.47	0.46	0.47	0.46	0.45	0.43	0.39
inv. gamma(1,1)	**0.72**	**0.72**	0.70	0.70	0.69	0.68	0.66	0.69	0.61
inv. Gaussian(0.5,1)	0.04	**0.05**	0.04	**0.05**	0.04	0.04	0.04	**0.05**	0.02
inv. Gaussian(1,1)	0.05	0.05	0.04	0.04	0.04	0.04	0.04	**0.06**	0.02
inv. Gaussian(1,3)	**0.04**	**0.04**	**0.04**	**0.04**	**0.04**	**0.04**	**0.04**	**0.04**	0.03
inv. Gaussian(3,1)	0.14	0.15	0.13	0.12	0.11	0.09	0.08	**0.16**	0.06
Gompertz(5,0.1)	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	0.97
Gompertz(1,3)	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	0.97
Gompertz(2,0.1)	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	0.98
Gompertz(1,0.1)	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	0.97

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Bibliothek, Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

CHAPTER 6

Comments and further research

The thesis aimed to advance statistical inference based on empirical integral transforms, specifically by utilizing their differential equations. This goal appears to be accomplished, particularly in the area of parameter estimation. The pursued differential equation approach has, to a great extent, alleviated computational burdens, which, in the author's opinion, impede the more widespread use of the existing transform-based paradigm. A broad family of estimators has been developed that are explicit, robust, and, at the same time, can maintain relatively low variances. While detailed observations and conclusions have already been given in Chapters 4 and 5, some general remarks are in order.

First, it must be stressed that the author's overall intention was to strike a balance between theory and empirical evidence of the estimators, at least in the form of extensive simulations, along with an account of computational aspects. This motivation arose from the fact that quite often literature of the subject emphasizes theoretical aspects (mostly in asymptotic settings), rather than providing sufficient, comprehensive empirical evidence, and addressing practical and methodological considerations. For instance, the monograph *Statistical Inference: The Minimum Distance Approach*, see Basu et al. (2011), by far the most extensive source on various minimum distance estimators to date, presents nearly no empirical results. Of course, many small-scale simulations exist in paper contributions. Nevertheless, these are often selective in the choice of distributions and conducted for parameter settings that do not pose numerical problems or reveal out-of-bound estimates. Contamination or misspecification scenarios, if considered at all, are very arbitrary too. Consequently, the scattered partial results make it challenging to compare techniques and draw confident conclusions regarding their (average) superiority over one another.

In this context, and to avoid overly optimistic statements about the proposed estimators, it must be admitted that the constructions presented in the thesis involve at least five 'degrees of freedom': 1) the type of transform (kernel), 2) the order of the employed differential equation, 3) the degree of the $\mathcal{L}^q$-norm, 4) the type of the weight function

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

and its parametrization, and 5) an auxiliary estimator of the weight function scale. In the parlance of statistical learning, these can be considered as 'hyperparameters' of the estimation algorithm. And, as we have seen, despite general hints provided by the theory of the transforms, the performance relies strongly on the settings and interplay of these hyperparameters.

Therefore, instead of making generalizations about the entire class of estimators, the author's main recommendation for further research would be a deeper exploration of the optimal settings for particular distributions. This involves conducting more extensive simulations and making comparisons with other types of estimators. Yet other aspects, such as the impact of dependent observations or breakdown behavior, can be investigated. In practical applications, where modeling assumptions are rarely met, cross-validation could be helpful in tuning the hyperparameters and is purposeful especially if repeated estimation in similar situations is planned. This suggestion was already discussed in sec. 4.6 for the parameters of the weight functions. Ultimately, a valuable estimator should demonstrate its usefulness over the long term when applied to real data sets. Here, feedback from practitioners in specific domains would be very valuable.

The proposed estimators and the related GoF tests, certainly require and deserve further explorations. Even if they may not be fully trusted at this stage of research as the primary tool of inference, they can certainly play important roles in complementing other techniques. For instance, being explicitly computable, the estimators offer good starting points for optimization procedures needed for the non-explicit estimators. Additionally, some of them can function as rapidly computable, robust alternatives. A significant discrepancy between these estimators and, for example, the ML or MM estimators, may indicate the presence of outliers.

6.1 Further research

The remainder of this chapter provides insights into statistical models and problems not addressed in this thesis, where the principles of differential equations of transforms can be applied or show potential applications. We explore suggestions, ideas, and open questions for further research, drawing connections with past and recent literature.

6.1.1 Discrete distributions

As explained in the introduction, our work focused solely on inference for the continuous distributions. However, there is no obstacle to applying the estimation methodology to the discrete probability laws. The transform traditionally considered as the most appropriate in discrete settings is the PGF. For a discrete variable K with probability mass function $p_{\boldsymbol{\theta}}(k)$, defined on some subset of integers $\mathcal{K}$, the PGF is given by

$$G_{\boldsymbol{\theta}}(s) = \sum_{\mathcal{K}} p_{\boldsymbol{\theta}}(k) s^k \,, \tag{6.1}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

while its non-parametric counterpart, based on the sample $(K_1, ..., K_n)$, is

$$\widehat{G}_n(s) = \sum_{j=1}^{n} s^{K_j} \,. \tag{6.2}$$

In fact, as evident in many sources on discrete distributions, PGF plays a pivotal role in characterizing these type of models, often providing much simpler representation than the mass function. In this regard, readers are directed to distinguished monographs such as *Univariate Discrete Distributions*, see Johnson et al. (1993) or Johnson et al. (2005), or *Lagrangian Probability Distributions*, see Consul and Famoye (2006). Interestingly, while long passages on various distributions are devoted there to parameter estimation problems, the authors do not mention the possibility of obtaining estimates through PGFs, although it was shown already by Kemp and Kemp (1988) and Dowling and Nakamura (1997), and then continued in more recent papers like Sim and Ong (2010), Sharifdoust et al. (2016) and Jiménez-Gamero and Batsidis (2017). In particular, estimation employed in the two monographs is consistently based on moments and likelihood. And quite often, these do not yield explicit estimates, whether for the mass function expressed through a special function or for the parameters trapped inside binomial coefficients.

Examining Tables 2.1-2.3 in Consul and Famoye (2006), where PGFs of over forty different, overlapping, or nested discrete families are compiled, one can check that most of them satisfy certain differential equations. Therefore, the approach proposed in this work should apply in many cases and prove especially useful for less popular models such as Borel, Haight, Felix, Sunil or Teja. But, in order not to consider such niche examples, let us briefly exemplify the benefits of our approach in the case of the standard binomial distribution. Its probability mass function, given by

$$p_{\boldsymbol{\theta}}(k) = \binom{m}{k} q^k (1-q)^{m-k} \,, \qquad k = 0, 1, ..., m \,, \tag{6.3}$$

expresses the probability of k successes in m independent trials with probability of a single success q. If m is known, estimation of q from the sample is straightforward: the ML estimator, coinciding with the MM one, is given by $\widehat{q}_n^{\text{ML}} = \widehat{q}_n^{\text{MM}} = (1/nm) \sum_{j=1}^{n} K_j$. However, when both parameters are to be estimated, then, as discussed in the examples 7.2.2, 7.2.9 and 7.2.13 in Casella and Berger (2002), both the explicit MM estimators and non-explicit ML estimators may be unstable. Additionally, the minimum distance estimators involving $p_{\boldsymbol{\theta}}(k)$, such as the CM or Scott's estimators, are difficult to construct as the support of the distribution is not fixed.

However, the binomial PGF,

$$G_{\boldsymbol{\theta}}(s) = (1 - q + qs)^m, \tag{6.4}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

$s \in \mathbb{R}$, opens the way for an alternative approach. Specifically, explicit estimates become feasible with help of the associated differential equation. By differentiating (6.4) w.r.t. s and replacing the transform and its derivative by their empirical counterparts, we obtain

$$(1 - q + qs)\widehat{G}'_n(s) - mq\widehat{G}_n(s) = 0. \tag{6.5}$$

From this, the estimators of m and $1/q$ can be explicitly derived either in form of the $\mathcal{L}^2$ estimators or the transform matching estimators operating at a single s-point.

Equivalently, we can work with the associated LT or CF and so stay in line with the theory elaborated in this thesis. For example, the LT would be $L_{\boldsymbol{\theta}}(s) = (1 - q + qe^{-s})^m$ with the following differential equation:

$$(1 - q + qe^{-s})L'_{\boldsymbol{\theta}}(s) + mqe^{-s}L_{\boldsymbol{\theta}}(s) = 0. \tag{6.6}$$

To apply Theorem 1 or 4 to the resulting $\mathcal{L}^2$ estimators, additional considerations on the weight function are needed. Moreover, an appropriate data-driven tuning of that function must be designed.

Of course, a PGF- or LT-based estimator can also be used to estimate q when m known. The following estimator,

$$\widehat{q}_n(s) = \frac{\sqrt[m]{\widehat{G}_n(s)} - 1}{s - 1}, \tag{6.7}$$

which stems directly from (6.4), was studied in Hołyński (2019a). The estimator was found to be highly robust for small success (or failure) probabilities. Hence, it may be of interest in certain applications where, for example, a 'true' error rate needs to be estimated in presence of batch errors (interpreted as outliers in the sense of binomial model).

6.1.2 Non-normalized distributions

An interesting estimation problem to which our approach can be applied arises with *non-normalized distributions*. These distributions are analytically known up to a normalization constant. That is, their density with parameter $\boldsymbol{\theta}$ can be represented as

$$f_{\boldsymbol{\theta}}(x) = C(\boldsymbol{\theta})g(x, \boldsymbol{\theta})\,, \tag{6.8}$$

where $C(\boldsymbol{\theta})$ is the constant (independent of x) such that

$$C(\boldsymbol{\theta})^{-1} = \int g(x, \boldsymbol{\theta})\mathrm{d}x\,. \tag{6.9}$$

Since the constant can only be approximated numerically, the maximum likelihood method requires an intensive numerical approach, both for computing the likelihood and

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

optimizing it. The same obstacle pertains, of course, to other density or distribution-based approaches, making the non-normalized distributions not really usable in statistics.

However, if a suitable differential equation satisfied by $f_{\boldsymbol{\theta}}(x)$ can be found (in the way described in sec. 3.2), the problematic constant can be eliminated from the derivation of estimators. As an example, consider the family of the polynomial-exponential models, with p-dimensional parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^\top$, with the density

$$f_{\boldsymbol{\theta}}(x) = C(\boldsymbol{\theta}) \exp\Big(\theta_1 x + \theta_2 x^2 + ... + \theta_p x^p\Big), \tag{6.10}$$

where $x > 0$ or $x \in \mathbb{R}$. This family includes, as the simplest cases, the exponential and normal distributions for $p = 1$ and $p = 2$, respectively. However, when higher powers of x are involved, the analytical forms of $C(\boldsymbol{\theta})$ remain unknown.

By applying the method from sec. 3.2, we obtain

$$f'_{\boldsymbol{\theta}}(x) = f_{\boldsymbol{\theta}}(x)(\theta_1 + 2\theta_2 x ... + \theta_p p x^{p-1}), \tag{6.11}$$

so that the corresponding differential equations satisfied by LT and CF are respectively

$$L_{\boldsymbol{\theta}}(s) + (s - \theta_1)L'_{\boldsymbol{\theta}}(s) + \sum_{k=2}^{p} (-1)^k k\theta_k L^{(k)}_{\boldsymbol{\theta}}(s) = 0, \tag{6.12}$$

and

$$\phi_{\boldsymbol{\theta}}(s) + (s - \theta_1)\phi'_{\boldsymbol{\theta}}(s) + \sum_{k=2}^{p} (-i)^k k\theta_k \phi^{(k)}_{\boldsymbol{\theta}}(s) = 0. \tag{6.13}$$

We note that, by using the empirical versions of the equations, the derivation of explicit estimators for these models for any p is possible thanks to linearity in parameter.

The non-normalized distributions has been recently addressed by Betsch et al. (2021), who devised yet other estimation technique coping with the problem of normalization. Specifically, the authors build on the so-called *Stein characterization*, to show that, under certain regularity conditions, the cumulative distribution function $F_{\boldsymbol{\theta}}(x)$ and density $f_{\boldsymbol{\theta}}(x)$ of a positive random variable X satisfy the following relation:

$$F_{\boldsymbol{\theta}}(z) = \mathbb{E}\Big[-\frac{f'_{\boldsymbol{\theta}}(X)}{f_{\boldsymbol{\theta}}(X)} \min(X, z)\Big], \quad z > 0. \tag{6.14}$$

With that, the estimator of $\boldsymbol{\theta}$ can be constructed as a minimizer of an integrated discrepancy between the sample counterparts of the two sides of the equality. For example, the weighted $\mathcal{L}^2$ estimator of this type can be defined as

$$\widehat{\boldsymbol{\theta}}_n^{\text{Stein}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \int_0^\infty \Big[\frac{1}{n} \sum_{j=1}^{n} \frac{f'_{\boldsymbol{\theta}}(X_j)}{f_{\boldsymbol{\theta}}(X_j)} \min(X_j, z) + \widehat{F}_n(z)\Big]^2 e^{-az} \text{d}z, \tag{6.15}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

where the exponential weight function controls the properties by user's setting of $a > 0$. As the parametric information is present here through the ratio $f'_{\boldsymbol{\theta}}(x)/f_{\boldsymbol{\theta}}(x)$, the unwanted normalization constants always cancel out. As shown in Betsch et al. (2021), these estimators admit complicated yet closed expressions involving double sums of order statistics.

It would be of interest to compare the performance of this method with estimators that can be derived from transforms using equation like (6.12) and (6.13). In the context of the polynomial exponential models, the author derived his own $\mathcal{L}^2$ LT-based estimators and conducted comparative simulations for the two-parameter case

$$f_{\theta_1,\theta_3}(x) = C(\theta_1, \theta_3) \exp\left(\theta_1 x + \theta_3 x^3\right), \tag{6.16}$$

$x > 0$, $\theta_1 \in \mathbb{R}$ and $\theta_3 \in (-\infty, 0)$, estimated in sec. 9 of Betsch et al. (2021) by their method. It was found that the own derived estimator is simpler and tends to be more robust against large outliers. Also, we note the estimators built on the Stein's characterization apply only to distributions on positive axis while the transform-based ones do not suffer from this limitation when based on CFs.

In connection with non-normalized distributions, it is important to mention the *saddlepoint approximation*; see Daniels (1954), Butler (2007). This technique is a universal method for approximating the density or cumulative distribution function based on the corresponding Laplace transform $L(s)$, or more precisely, the cumulant generating function (CGF) $\Psi_{\boldsymbol{\theta}}(s) = \log L_{\boldsymbol{\theta}}(-s)$. In particular, the saddlepoint approximation of a density is given by the following formula:

$$\tilde{f}_{\boldsymbol{\theta}}(x) = \frac{1}{\sqrt{2\pi \Psi''_{\boldsymbol{\theta}}(s_{\mathrm{o}})}} \exp\{\Psi_{\boldsymbol{\theta}}(s_{\mathrm{o}}) - s_{\mathrm{o}} x\}, \tag{6.17}$$

where $s_{\mathrm{o}} = s_{\mathrm{o}}(x)$ is called the *saddlepoint* and is the unique solution of the equation

$$\Psi'_{\boldsymbol{\theta}}(s_{\mathrm{o}}) = x, \tag{6.18}$$

on the interval of existence of the transform. Thus, it can be viewed as either an analytical or numerical method for transform inversion, depending on whether the equation can be solved analytically or requires a numerical approach.

The key idea behind this technique is to provide simpler expressions compared to the original (possibly unknown) density. However, the saddlepoint density is usually not a proper density because $\int \tilde{f}_{\boldsymbol{\theta}}(x)\mathrm{d}x \neq 1$. Therefore, it requires a normalization constant, computed using (6.9) with $g(x, \boldsymbol{\theta}) = \tilde{f}_{\boldsymbol{\theta}}(x)$. If, for some reason, we wish to use the approximation as the basis for statistical inference, the need for normalization can be a prohibitive obstacle. In such cases, our proposed method of differential equations, applied to the approximated density (6.17), may enable this line of approach.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

6.1.3 System of Pearson's distributions

Researchers with an interest in the historical foundations of mathematical statistics may observe that the technical aspects of our differential equation approach share similarities with the early parametric inference methodology developed by Karl Pearson. In Pearson (1895), it was demonstrated that numerous standard probability densities satisfy the following differential equation:

$$\frac{f'(x)}{f(x)} = \frac{x-a}{b_0 + b_1 x + b_2 x^2} \,. \tag{6.19}$$

They constitute the so-called *Pearson's system* of 'frequency curves'. Depending on the values of the four parameters, the system generates, for example, the normal, gamma, the Gamma, Laplace, or Pareto distributions, which, according to existence and number of the solutions of $b_0 + b_1 x + b_2 x^2 = 0$ are categorized into one of the twelve Pearson's types and subtypes. The associated statistical idea was to use (6.19) to find the recurrence relationship between the raw moments of the distributions. Rearranging the equation, multiplying both sides by x^k, and integrating by parts over the range of x, say (x_1, x_2), leads to

$$\begin{aligned}\Big[(b_0 + b_1 x + b_2 x^2)x^k f(x)\Big]_{x=x_1}^{x=x_2} - \int_{x_1}^{x_2} [k b_0 x^{k-1} + (k+1)b_1 x^k + (k+2)b_2 x^{k+1}] f(x)\mathrm{d}x \\ = \int_{x_1}^{x_2} x^{k+1} f(x)\mathrm{d}x - a \int_{x_1}^{x_2} x^k f(x)\mathrm{d}x \,,\end{aligned} \tag{6.20}$$

and, provided that the first term is zero, we have

$$-k b_0 \mathbb{E}[X^{k-1}] - (k+1)b_1 \mathbb{E}[X^k] - (k+2)b_2 \mathbb{E}[X^k] = \mathbb{E}[X^{k+1}] - a\mathbb{E}[X^k] \,. \tag{6.21}$$

Putting $k = 1, 2, 3$, we obtain the system that can be solved for a, b_0, b_1, and b_2. Finally, given a sample $X_1, ..., X_n$ we replace the theoretical moments with their sample counterparts, $\frac{1}{n}\sum_{j=1}^{n} X_j^k$, and estimate the four parameters; for the explicit though complicated solution see e.g. Thompson and Tapia (1990). This is the Pearson's method of moments in its very original form.

The distinction between these and the 'modern' MM estimators, utilized throughout this work, lies in the underlying inferential paradigm. Specifically, the MM estimators introduced in eq. (1.3) align with the *inductive (Fisherian)* paradigm originated by R.A. Fisher (see also the footnote on p. 1). They are derived under a narrowly postulated distributional model, in the same way as the ML or MD estimators. Optionally, they can be employed later in a GoF test to verify the adequacy of the initially hypothesized model. In contrast, Pearson's concept directly utilizes the realized values of the moment estimators of a, b_0, b_1, and b_2 to specify the model from a broad spectrum of possibilities that satisfy his equation, without assuming any concrete one a priori. For example, the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

realization $\widehat{b}_0 < 0$, $\widehat{b}_1 \approx 0$, and $\widehat{b}_2 \approx 0$ suggest that the sample is generated by the normal distribution; see sec. 1.3 of Thompson and Tapia (1990) or sec. 13.4.1 of Spanos (2019) for detailed and interesting expositions.

Although the Pearson's approach is rarely used nowadays, we should note that its core derivation relies on the same principle as the alternative method of differential equations presented in sec. 3.2. With the difference that there we took an integral transform of both sides instead of computing moments. Because the members of the Pearson's system can be two-tailed, CF becomes the appropriate transform to treat the problem in our way and, hopefully, to define consistent estimators. So, beginning with

$$(b_0 + b_1 x + b_2 x^2) f'(x) = (x - a) f(x)\,, \tag{6.22}$$

we compute CF of both sides. Next, utilizing (3.18), we derive the differential equation

$$b_2 is\phi''(s) + [(2b_2+1)i - b_1 s]\phi'(s) + (a - isb_0 - b_1)\phi(s) - b_0 f(x_1) = 0, \tag{6.23}$$

where x_1 is the lower end-point of the distribution's support. The challenge in working with this equation is in necessity of determining $f(x_1)$, as the system encompasses distributions with diverse shapes and supports. One approach could involve anticipating and, if necessary, estimating $f(x_1)$ from a histogram or kernel-density estimator. Of course, we may get rid of the term by taking the one order higher equation:

$$b_2 is\phi'''(s) + [(3b_2+1)i - b_1 s]\phi''(s) + (a - isb_0 - 2b_1)\phi'(s) - ib_0\phi(s) = 0. \tag{6.24}$$

The weighted $\mathcal{L}^2$ estimators of $\boldsymbol{\theta} = (a, b_0, b_1, b_2)^\top$ based on these equations will be

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,2,1,0} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int_{-\infty}^{\infty} \Big| & b_2 is\widehat{\phi}_n''(s) + [(2b_2+1)i - b_1 s]\widehat{\phi}_n'(s) \\ & + (a - isb_0 - b_1)\widehat{\phi}_n(s) - b_0 \widehat{f}_n(x_1)\Big|^2 w(s)\mathrm{d}s,\end{aligned} \tag{6.25}$$

and

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,3,2,1,0} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int_{-\infty}^{\infty} \Big| & b_2 is\widehat{\phi}_n'''(s) + [(3b_2+1)i - b_1 s]\widehat{\phi}_n''(s) \\ & + (a - isb_0 - 2b_1)\widehat{\phi}_n'(s) - ib_0\widehat{\phi}_n(s)\Big|^2 w(s)\mathrm{d}s,\end{aligned} \tag{6.26}$$

with the weight function, for example, $w(s) = s^\gamma e^{-as^2}$, $a > 0$, $\gamma = 0, 2, \ldots$. The derivation of four explicit estimators in both cases should follow a similar, albeit more tedious, process to that outlined for the two-parameter normal distribution in Appendix B.3. With such expressions in hand, it would be of interest to investigate how the CF-based estimators perform in comparison to Pearson's MM estimators, particularly in terms of robustness.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Finally, it is worth mentioning that the Pearson family does not exhaust the families which encompass a number of popular distributions as special cases, and which are characterized by some sort of differential (or difference) equations in the space of distribution and/or the related transform. In the realm of discrete distributions, analogues has been developed such as the Katz-, Sund and Jewell-, Ord- and Kemp families; see ch. 2 of Johnson et al. (1993). For example, the Katz family is described by the following recursion for its probability mass function

$$\frac{p(k+1)}{p(k)} = \frac{\alpha + \beta k}{1+k}, \qquad k = 0, 1, 2\dots. \tag{6.27}$$

The standard binomial-, Poisson- and negative binomial distributions are members of this family for $\beta < 0$, $\beta = 0$, and $\beta > 0$, respectively. Therefore, discrimination between the members based on a sample, can be accomplished, in the spirit of Pearson's approach, through parameter estimation.

The PGF of the family

$$G(s) = \left(\frac{1-\beta s}{1-\beta}\right)^{-\alpha/\beta}, \tag{6.28}$$

is not convenient for constructing the estimators. On contrary, the associated differential equation

$$(1-\beta s)G'(s) - \alpha G(s) = 0, \tag{6.29}$$

or the differential equation of the corresponding LT

$$(1-\beta e^{-s})L'(s) - \alpha e^{-s} L(s) = 0, \tag{6.30}$$

where the parameters appear in a linear relationship, make the task much easier and lead to closed-form expressions.

Thanks to its flexibility, the Katz family finds important applications in actuarial science, where the model has to accommodate under- and overdispersion of certain discrete quantities with respect to the Poisson model. Therefore, a specialized alternative and robust estimation technique for this family would be valuable.

6.1.4 Regression models

While empirical transform methods have primarily been applied to the estimation of distribution parameters, some researchers have investigated their applicability to parametric regression models. Estimators in these settings can be formulated in at least two ways:

1. as minimum distance estimators based on the distributional assumptions about residuals, see Paulson and Nicklin (1983);

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

2. as minimum distance estimators minimizing the distance between the model and the empirical transform of the regression function (and disregarding the assumption about residuals).

The second approach is more flexible and better suits the methodology of our work. It was proposed by Leedow and Tweedie (1983), Feigin et al. (1983), and Laurence and Morgan (1987) with application to specific biological experiments. Subsequently, a more comprehensive theoretical foundation was provided by Yao and Morgan (1999). All these authors worked with the Laplace transform.

To this end, consider a classical univariate regression model

$$Y = \eta(x|\boldsymbol{\theta}) + \epsilon\,, \tag{6.31}$$

with zero-mean observation error, $\mathbb{E}[\epsilon] = 0$. Given a sample of n observations (x_1, Y_1), ..., (x_n, Y_n), forming

$$Y_j = \eta(x_j|\boldsymbol{\theta}) + \epsilon_j\,, \tag{6.32}$$

$j = 1, ..., n$, our goal is to estimate the vector of parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^\top$, and with the estimate $\widehat{\boldsymbol{\theta}}_n$ at hand, fit the regression line

$$\widehat{\mathbb{E}}[Y|x] = \eta(x|\widehat{\boldsymbol{\theta}}_n)\,, \tag{6.33}$$

which predicts the response for all the values of the predictor x.

Parameter estimation requires computation of the model LT,[1]

$$\mathscr{L}(s|\boldsymbol{\theta}) = \int \eta(x|\boldsymbol{\theta}) e^{-sx} \mathrm{d}x\,, \tag{6.34}$$

and construction of the empirical transform from the sample. The latter process is by far not unique, but the most obvious option is to use the Riemann-sum approximation of (6.34). In particular, we sort the sample, (x_1, Y_1), ..., (x_n, Y_n), so that $x_1 < x_2 < ... < x_n$, and partition the x-axis with $n + 1$ points c_j, computed via the mid-point rule,

$$c_j = \frac{x_j + x_{j+1}}{2}\,, \quad j = 2, 3, ..., n-1\,. \tag{6.35}$$

For the end-points, we take $c_1 = x_1$ and $c_{n+1} = x_n$.

Using this partition, we can form a discrete approximation of the function

$$\eta(x|\boldsymbol{\theta}) \approx \begin{cases} \eta(x_1|\boldsymbol{\theta}) & \text{for } c_1 \le x < c_2 \\ \eta(x_2|\boldsymbol{\theta}) & \text{for } c_2 \le x < c_3 \\ \vdots & \\ \eta(x_n|\boldsymbol{\theta}) & \text{for } c_n \le x < c_{n+1}\,, \end{cases} \tag{6.36}$$

[1]To avoid confusion with LTs of probability distribution, we use here the stylish letter $\mathscr{L}$ to denote LTs of the regression functions.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

as shown in Fig. 6.1, and approximate (6.34) by the Riemann sum,

$$\begin{aligned}\mathscr{L}(s|\boldsymbol{\theta}) &\approx \sum_{j=1}^{n}\eta(x_j|\boldsymbol{\theta})\int_{c_j}^{c_{j+1}} e^{-sx}\mathrm{d}x \\ &= \sum_{j=1}^{n}\eta(x_j|\boldsymbol{\theta})\frac{e^{-sc_j}-e^{-sc_{j+1}}}{s}. \qquad (6.37)\end{aligned}$$

Finally, the empirical LT is obtained by replacing $\eta(x_j|\boldsymbol{\theta})$ by its unbiased estimate Y_j, that is

$$\widehat{\mathscr{L}_n}(s) = \sum_{j=1}^{n} Y_j \frac{e^{-sc_j}-e^{-sc_{j+1}}}{s}. \qquad (6.38)$$

Now, in spirit of the transform-matching approach proposed for distributions, the estimator of $\boldsymbol{\theta}$ may be defined as the p-dimensional solver of

$$\begin{cases}\mathscr{L}(s_1|\boldsymbol{\theta}) = \widehat{\mathscr{L}_n}(s_1) \\ \mathscr{L}(s_2|\boldsymbol{\theta}) = \widehat{\mathscr{L}_n}(s_2) \\ \vdots \\ \mathscr{L}(s_p|\boldsymbol{\theta}) = \widehat{\mathscr{L}_n}(s_p)\,,\end{cases} \qquad (6.39)$$

computed at set of p s-points inside the transform's region of convergence. Alternatively, as advocated by Yao and Morgan (1999), the limiting estimator as $s_1 \to \ldots \to s_p \to s$ is preferred for easier controllability; obtaining this version of the estimator amounts to solving

$$\begin{cases}\mathscr{L}(s|\boldsymbol{\theta}) = \widehat{\mathscr{L}_n}(s) \\ \mathscr{L}'(s|\boldsymbol{\theta}) = \widehat{\mathscr{L}'_n}(s) \\ \vdots \\ \mathscr{L}^{(p-1)}(s|\boldsymbol{\theta}) = \widehat{\mathscr{L}_n^{(p-1)}}(s).\end{cases} \qquad (6.40)$$

The mentioned authors also pointed out that the so-defined estimator can be strongly biased when sample data is available only for a part of the region over which $\eta(x|\boldsymbol{\theta})$ is non-zero. To overcome this problem, the model LT should be computed only in that range, just as the empirical version is. That is, instead of (6.34), one should use the *end-corrected transform*,

$$\mathscr{L}_{\mathrm{ec}}(s|\boldsymbol{\theta}) = \int_{x_1}^{x_n} \eta(x|\boldsymbol{\theta})e^{-sx}\mathrm{d}x\,, \qquad (6.41)$$

(and its derivatives), where x_1 and x_n are the smallest and largest x-coordinates in the sorted sample. Although not explicitly mentioned by the authors, the correction is crucial for another important reason: it liberates us from the issues related to transform convergence and boundedness - problems we consistently encountered when working with

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

distribution functions. In fact, over a finite interval of x, transforms with any kernel can be computed. Of course, the end-corrected transforms will have more complicated expressions than those without corrections.

The author of this thesis has explored the approach, as outlined above, for several popular regression models. Explicit and effective estimators have been obtained for linear, polynomial, exponential, and sinusoidal regression functions. This encourages further work because explicit regression estimators are known only in case of the least-squares approach applied to linear models. Nevertheless, certain aspects of the methodology published until now merit reconsideration or modification.

First, despite reducing the dimension of the transform variable from p to 1, the issue of choosing a single s-point on the real line remains essentially unresolved. The data-driven selection of this point, so successful in distribution estimation, seems unfeasible in the regression scenario, possibly due to the different properties of the Laplace transform. In these circumstances, we propose a different selection strategy that aims to maintain the (asymptotic) variance of the estimator at a low level. Note that each element of the estimator $\widehat{\boldsymbol{\theta}}_n(s) = \{\widehat{\theta}_n^{(1)}(s), ..., \widehat{\theta}_n^{(p)}(s)\}^\top$ that solves (6.40) can expressed as

$$\widehat{\theta}_n^{(i)}(s) = g_i\Big(\widehat{\mathscr{L}_n}(s), \widehat{\mathscr{L}_n'}(s), ..., \widehat{\mathscr{L}_n}^{(p-1)}(s)\Big), \quad i = 1, ..., p. \tag{6.42}$$

for some function g_i. The Delta Method, Casella and Berger (2002), indicates that its asymptotic variance is linearly proportional to the variances of the empirical LT and its derivatives at s. Although achieving low variance by putting s close to zero would be desirable, an examination of (6.38) reveals the opposite effect: the variance is maximized as $s \to 0$. Therefore, we recommend working with the transform with a modified kernel, namely

$$\mathscr{L}_{\text{ec}}(s|\boldsymbol{\theta}) = \int_{x_1}^{x_n} \eta(x|\boldsymbol{\theta}) s^k e^{-sx} \mathrm{d}x\,, \tag{6.43}$$

$k = 1, 2, ...$, so that its empirical counterpart is

$$\widehat{\mathscr{L}_n}(s) = \sum_{j=1}^{n} Y_j\, s^{k-1}[e^{-sc_j} - e^{-sc_{j+1}}]\,. \tag{6.44}$$

The variances of (6.44) and its derivatives exhibit now the desired behavior as $s \to 0$. Importantly, the modification of the kernel does not complicate the computation of the end-corrected model LT in any way.

It is the author's consistent experience that the precision (bias and variance) of estimators constructed in this modified manner is very close to that of the non-linear least squares in the respective model. Of course, the question arises: How small should s be? The answer is that it should be the smallest possible within the constraints that ensure the numerical stability of computing the terms e^{-sc_1} and $e^{-sc_{n+1}}$. Alternatively, the search for an appropriate s in the vicinity of zero may be supported by cross-validation.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Yet another option is to employ $\mathcal{L}^2$ estimator of the form

$$\widehat{\boldsymbol{\theta}}_n \;=\; \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int_0^\infty \left[\mathscr{L}_{\mathrm{ec}}(s|\boldsymbol{\theta}) - \widehat{\mathscr{L}_n}(s)\right]^2 e^{-as}\mathrm{d}s\,, \qquad a>0, \tag{6.45}$$

which does not require the derivatives of transforms and avoids the numerical problems connected with exponentials. Also, this type of estimator is going to be robust against outliers in the Y variable, as opposed to (6.40), which, remarkably, does not show this property in regression settings. However, the explicitness of the estimates is not guaranteed unless $\mathscr{L}_{\mathrm{ec}}(s|\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$.

Of course, as with probability distributions, we may encounter cases where the regression model has no explicit or sufficiently tractable transform computed by (6.34), (6.41) or (6.43), for example, if $\eta(x|\boldsymbol{\theta})$ is a rational function of x. In such instances, however, we may resort to our methodology and find a differential equation satisfied by $\eta(x|\boldsymbol{\theta})$ to derive the one satisfied by $\mathscr{L}_{\mathrm{ec}}(s|\boldsymbol{\theta})$. The relations between the derivatives are slightly more involved than for the probability distributions because of end-corrections. Nevertheless, as examined already by the author, the approach described in sec. 3.2 generally applies to regression problems as well.

The presented methodology, along with all its modifications and potential $\mathcal{L}^2$ variants of the estimators, require separate asymptotic distribution- and robustness theories. Both offer fruitful topics for future research.

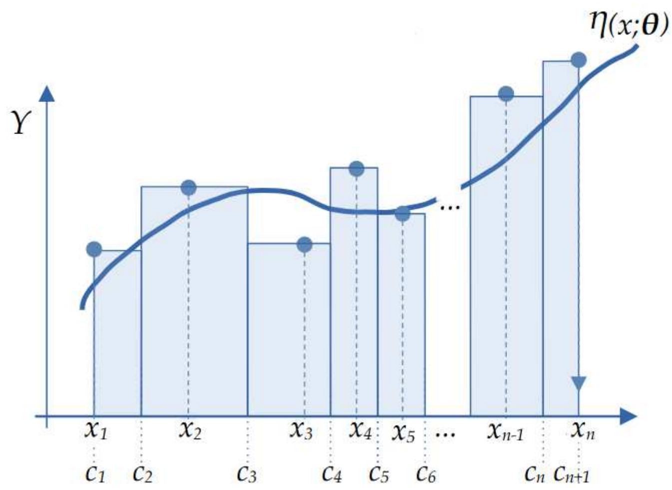


Figure 6.1: Riemann-sum approximation of the regression function $\eta(x|\boldsymbol{\theta})$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

6.1.5 Multivariate and non-parametric problems

While this thesis addressed univariate distributions, it is crucial to emphasize that empirical transform methods are also applicable to vectors of random variables. In fact, recent developments in this area of statistics are primarily driven by multivariate settings. However, when it comes to strictly parametric problems, the transform approach encounter here limitations, as very few multivariate distributions admit tractable expressions for CFs, LTs and PGFs (yet even fewer than the univariate ones do); consider, for instance, the table of standard bivariate distributions in Spanos (2019) or consult various chapters on the multivariate distributions in Kotz et al. (2000).

For this discussion, consider a d-dimensional continuous random vector $\boldsymbol{X} = (X_1, ..., X_d)^\top$, described by the joint cumulative distribution $F(\boldsymbol{x}|\boldsymbol{\theta}) = P(X_1 \leq x_1, ..., X_d \leq x_d)$ with parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^\top$. The corresponding multivariate LT and CF are defined as

$$L(\boldsymbol{s}|\boldsymbol{\theta}) = \mathbb{E}[e^{-\boldsymbol{s}^\top \boldsymbol{X}}] = \int_{\mathbb{R}^d} e^{-\boldsymbol{s}^\top \boldsymbol{x}} \mathrm{d}F_{\boldsymbol{\theta}}(\boldsymbol{x}) \,, \tag{6.46}$$

and

$$\phi(\boldsymbol{s}|\boldsymbol{\theta}) = \mathbb{E}[e^{i\boldsymbol{s}^\top \boldsymbol{X}}] = \int_{\mathbb{R}^d} e^{i\boldsymbol{s}^\top \boldsymbol{x}} \mathrm{d}F_{\boldsymbol{\theta}}(\boldsymbol{x}) \,, \tag{6.47}$$

where $\boldsymbol{s} = (s_1, ..., s_d)^\top$ is the vector of transform variables.

Given a random sample of n realizations, $(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)$, where $\boldsymbol{X}_j = (X_{j1}, X_{j2}, ..., X_{jd})$, with corresponding empirical distribution $\widehat{F}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^n \prod_{k=1}^d I\{X_{jk} \leq x_k\}$, the empirical non-parametric counterparts of (6.46) and (6.47) are expressed as

$$\widehat{L}_n(\boldsymbol{s}) = \frac{1}{n}\sum_{j=1}^n e^{-\boldsymbol{s}^\top \boldsymbol{X}_j} \,. \tag{6.48}$$

and

$$\widehat{\phi}_n(\boldsymbol{s}) = \frac{1}{n}\sum_{j=1}^n e^{i\boldsymbol{s}^\top \boldsymbol{X}_j} \,, \tag{6.49}$$

Whereas the multivariate ELT rarely occurs in literature, extensive treatments on the properties and applications of multivariate ECF are available in several monographs and overviews such as Csörgő (1981), Kankainen (1995), Ushakov (1999) or Meintanis (2016).

Multivariate distributions that possess explicit and tractable model transforms are nearly limited the elliptical families. Among these, the CFs of the d-variate normal, stable and Laplace distributions are expressed, respectively, by

$$\phi(\boldsymbol{s}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(i\boldsymbol{s}^\top \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s}\right) , \tag{6.50}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

$$\phi(\boldsymbol{s}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \exp\left(i\boldsymbol{s}^\top\boldsymbol{\mu} - \frac{1}{2}(\boldsymbol{s}^\top\boldsymbol{\Sigma}\boldsymbol{s})^{\alpha/2}\right), \tag{6.51}$$

and

$$\phi(\boldsymbol{s}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{i\boldsymbol{s}^\top\boldsymbol{\mu}}{1+\frac{1}{2}\boldsymbol{s}^\top\boldsymbol{\Sigma}\boldsymbol{s}}, \tag{6.52}$$

where $i^2 = -1$, $0 < \alpha \leq 2$, $\boldsymbol{\mu} \in \mathbb{R}^d$ is the $d \times 1$ vector of location parameters and $\boldsymbol{\Sigma}$ is a $d \times d$ symmetric, positive definite scale matrix.

But even in these favorable cases, parameter estimation directly from the transforms is often difficult. Rather, the available expressions have been used to construct goodness-of-fit statistics computed for standardized data (by means of alternative estimators of location and scale), thereby rendering the parameters irrelevant; see the tests proposed by Henze and Wagner (1997), Fragiadakis and Meintanis (2011) or Meintanis et al. (2015).

Explicit estimators for the stable distribution family, derived directly from (6.51), were proposed by already by Press (1972) and, later by Zolotarev (1981); see also sec. 3.4.1. of Ushakov (1999) for an overview. However, these estimators were computed at a single d-dimensional $\boldsymbol{s}$ point. Given our unsatisfactory experiences with so-defined CF estimators in univariate scenarios, such as (4.83) or (4.189) and (4.190), their multivariate counterparts are also anticipated to be weak. Notably, neither of the two mentioned papers provide empirical results in this respect.

More reliable estimation is expected with multidimensional integrated $\mathcal{L}^2$-type estimators. To our best knowledge, the only such estimator was developed for the normal distribution by Paulson and Lawrence (1982). Using (6.50), it was defined as

$$\left(\widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right) = \underset{\boldsymbol{\mu},\boldsymbol{\Sigma}}{\operatorname{argmin}} \sum_{j=1}^{n} \int_{\mathbb{R}^d} \left|\exp\left(i\boldsymbol{s}^\top\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{s}^\top\boldsymbol{\Sigma}\boldsymbol{s}\right) - \exp\left(i\boldsymbol{s}^\top\boldsymbol{X}_j\right)\right|^2 \mathrm{d}W(\boldsymbol{s}), \tag{6.53}$$

that is, as the minimizer of sums the weighted $\mathcal{L}^2$ distances between CF and ECF of each *single* observation $\boldsymbol{X}_j$ (rather than as the minimizer of the weighted $\mathcal{L}^2$ distance between CF and ECF of *all* n observations). Upon applying a d-variate Gaussian weight function, the definition resulted in a set of estimating equations where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are coupled implicitly. Consequently, the estimator necessitates a numerical approach and, despite having proven robustness, has not gained popularity.

However, explicit counterparts of (6.53) appear to be derivable by resorting to *partial* differential equations satisfied by (6.50). For simplicity, let us consider the bivariate normal case,

$$\phi(s_1, s_2) = \exp\left(i(s_1\mu_1 + s_2\mu_2) - \tfrac{1}{2}(\sigma_1^2 s_1^2 + \sigma_2^2 s_2^2 + 2\sigma_{12}s_1s_2)\right), \tag{6.54}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where σ_1^2, σ_2^2, and σ_{12} stand for the variances and the covariance. By taking the partial derivatives, we obtain

$$\begin{cases} \dfrac{\partial \phi(s_1, s_2)}{\partial s_1} = \phi(s_1, s_2)(i\mu_1 - \sigma_1^2 s_1 - 2\sigma_{12} s_2) \\ \dfrac{\partial \phi(s_1, s_2)}{\partial s_2} = \phi(s_1, s_2)(i\mu_2 - \sigma_2^2 s_2 - 2\sigma_{12} s_1)\,. \end{cases} \tag{6.55}$$

Then, eliminating $\phi(s_1, s_2)$ and replacing the derivatives by their estimates, we arrive at the empirical equality on which we may base the estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\sigma}_{12})^\top$. That is, our candidate for the weighted $\mathcal{L}^2$ estimator, computable in closed form, is

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_n = \underset{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12}}{\operatorname{argmin}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Bigg| & \frac{\partial \widehat{\phi}_n(s_1, s_2)}{\partial s_1}(i\mu_2 - \sigma_2^2 s_2 - 2\sigma_{12} s_1) \\ & - \frac{\partial \widehat{\phi}_n(s_1, s_2)}{\partial s_2}(i\mu_1 - \sigma_1^2 s_1 - 2\sigma_{12} s_2)\Bigg|^2 w(s_1, s_2)\mathrm{d}s_1\mathrm{d}s_2\,, \end{aligned} \tag{6.56}$$

where $w(s_1, s_2) = s_1^{\gamma_1} s_2^{\gamma_2} \exp(-a_1 s_1^2 - a_2 s_2^2)$, $\gamma_1, \gamma_2 = 0, 2, ...$, and $a_1, a_2 > 0$.

An important question for further research is whether the differential approach could be extended to models whose multivariate transform is unknown or too complicated, but a partial differential equation satisfied by it is derivable from the corresponding equation of the density. This would require a generalization of our alternative method from sec. 3.2 to multivariate settings. In view of complexity of the formulas for Laplace transforms of partial derivatives of a just two-variable function, as presented in Cohen (2007), extending these derivations to an arbitrary dimension d appears to be a laborious task.

While multivariate parametric models present certain challenges for empirical transform methods, these methods flourish in non-parametric multivariate procedures. One application area involves tests for *total independence* among a group of random variables. Such tests rely on the premise that, in the case of total independence, the joint CF of the vector, say X_1, X_2 ..., X_d, equals the product of the CFs of the marginals:

$$\phi_{X_1,\dots,X_d}(\boldsymbol{s}) = \prod_{k=1}^{d} \phi_{X_k}(s_k). \tag{6.57}$$

Therefore, upon replacing the CFs by their empirical versions, i.e. $\phi_{X_1,\dots,X_d}(\boldsymbol{s})$ by $\widehat{\phi}_n(\boldsymbol{s})$ and $\phi_{X_k}(s_k)$ by $\widehat{\phi}_{k,n}(s_k)$, the independence can be tested via realization of the following statistic:

$$T_n = n \int_{\mathbb{R}^d} \Big|\widehat{\phi}_n(\boldsymbol{s}) - \prod_{k=1}^{d} \widehat{\phi}_{k,n}(s_k)\Big|^2 \mathrm{d}W(\boldsymbol{s})\,. \tag{6.58}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Statistics of this type were studied by Kankainen (1995). More recently, they have been applied to non-parametric regression and time-series models to test for independence between the residuals (innovations) and the regressors (past observations); see Hlávka et al. (2011) and Hušková et al. (2019), respectively.

The second group of non-parametric procedures deals with *change-point* detection; see, e.g., Tan et al. (2016) and Hlávka et al. (2017). Consider a time series of d-dimensional vector observations $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$, each having distribution function $F_j(\mathbf{x})$, $j = 1, ..., n$. It is assumed that the distribution changes at some point, i.e.

$$F_1 = ... = F_k \neq F_{k+1} = ... = F_{n-1} = F_n \tag{6.59}$$

where k is called the *change point*.

Taking the d-variate ECFs,

$$\widehat{\phi}_k(\boldsymbol{s}) = \frac{1}{k} \sum_{j=1}^{k} e^{i\boldsymbol{s}^\top \boldsymbol{X}_j}\,, \tag{6.60}$$

and

$$\widehat{\phi}_k^0(\boldsymbol{s}) = \frac{1}{n-k} \sum_{j=k+1}^{n} e^{i\boldsymbol{s}^\top \boldsymbol{X}_j}\,, \tag{6.61}$$

we can adopt the following statistic

$$T_n(k) = \left(\frac{k(n-k)}{n^2}\right)^{\gamma} \frac{k(n-k)}{n} \int_{\mathbb{R}^d} \left|\widehat{\phi}_k(\boldsymbol{s}) - \widehat{\phi}_k^0(\boldsymbol{s})\right|^2 \mathrm{d}W(\boldsymbol{s})\,, \tag{6.62}$$

$\gamma \in (0, 1]$, that measures the divergence between successive subsamples $(\boldsymbol{X}_1, ..., \boldsymbol{X}_k)$ and $(\boldsymbol{X}_{k+1}, ..., \boldsymbol{X}_n)$. The estimator of the change-point is then

$$\widehat{k} \;=\; \underset{1 \leq k < n}{\operatorname{argmax}}\; T_n(k)\,. \tag{6.63}$$

Notice that such distance statistics can also be used for *cluster analysis*. For example, assuming that each of the observations $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ belongs to one of two clusters, one calculates (6.62) for every division of the sample into two sets. The division that maximizes the statistic is considered as the one indicating the strongest separation between the two populations, and determines the cluster assignment. This approach, used to multiclass clustering in the context of the support vector machines, has been presented by Cubiles-de-la Vega et al. (2012).

Given a couple of fundamental applications of the non-parametric statistics such as (6.58) or (6.62), it raises an interesting question whether basing them on derivatives of the transforms would be beneficial in some way, as it turned out to be in the parametric settings of this thesis.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

APPENDIX A

Results from theory of V- and U-statistics

We review the properties of V- and U-statistics exploited in this thesis to analyze asymptotic distribution of $\mathcal{L}^2$ estimators and statistics based on transforms. For more details the reader is referred to Serfling (1980), Lee (1990) or van der Vaart (2000).

In many statistical models, quantities of interest can be represented by an m-degree expectation functional acting on a distribution function $F = F(x) = P(X \le x)$, namely

$$T(F) \;=\; \mathbb{E}_F[\,\psi(X_1, ..., X_m)]\,. \tag{A.1}$$

The function $\psi : \mathbb{R}^m \to \mathbb{R}$, called *kernel*, is permutation symmetric in its arguments and integrable w.r.t to F. Such functional usually refers to a parameter of F, or to some non-parametric characteristic (mean, median, quantile).

A natural way of estimating $T(F)$ from random sample of $n \ge m$ i.i.d. observations, $\{X_1, ..., X_n\}$, is by the plug-in estimator

$$T(\widehat{F}_n) \;=\; V_n \;=\; \frac{1}{n^m}\sum_{i_1=1}^{n} ... \sum_{i_m=1}^{n} \psi(X_{i_1}, ..., X_{i_m})\,, \tag{A.2}$$

called the *V-statistic* of degree m.

Since for $m > 1$ the sum contains some terms $i_1, ..., i_m$ which are not distinct, V_n is a biased estimator of $T(F)$, that is $\mathbb{E}[V_n] \ne T(F)$. Moreover, the asymptotic normality of V_n cannot be shown by the Central Limit Theorem which assumes independence of the summands. However, the asymptotic distribution of V_n can be determined by the corresponding *U-statistic*

$$U_n \;=\; \binom{n}{m}^{-1} \sum\sum_{1\le i_1 < i_2 ... < i_m \le n} ... \sum \psi(X_{i_1}, ..., X_{i_m})\,, \tag{A.3}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

in which no duplicates occur. U_n is an unbiased estimator of $T(F)$, $\mathbb{E}[U_n] = T(F)$, since it is an average of $\binom{n}{m}$ terms, each with expectation $\mathbb{E}_F\left[\psi(X_1, ..., X_m)\right]$.

Variance of U-statistics

Combinatorial arguments lead to the following exact expression for variance of U-statistic:

$$\mathbb{V}\mathrm{ar}[U_n] \;=\; \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c}\sigma_c^2\,, \tag{A.4}$$

with

$$\sigma_c^2 = \mathbb{V}\mathrm{ar}[\psi^{(c)}(X_1, ..., X_c)]\,, \tag{A.5}$$

where

$$\psi^{(c)}(x_1, ..., x_c) = \begin{cases} \mathbb{E}[\psi(x_1, ..., x_c, X_{c+1}, ..., X_m)]\,, & 1 \le c \le m-1\,, \\ \psi(x_1, ..., x_m)\,, & c = m\,. \end{cases} \tag{A.6}$$

For proof see Lee (1990), sec. 1.3.

Asymptotic normality of V- and U-statistics

By the method of projection due to Hoeffding (1948) and Hájek (1968), it can be shown that if $\mathbb{E}^2[\psi(X_1, ..., X_m)] < \infty$ and $\sigma_1^2 > 0$, then

$$\sqrt{n}(U_n - T(F)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, m^2\sigma_1^2)\,, \tag{A.7}$$

for proof see van der Vaart (2000), sec. 12.1.

Also, if $\mathbb{E}^r[\psi(X_1, ..., X_m)] < \infty$, $r = 1, 2, ...$, it can be proved that

$$\mathbb{E}|U_n - V_n|^r = O(n^{-r})\,, \tag{A.8}$$

as shown in sec. 5.7.3 of Serfling (1980). Therefore, with $r = 2$ it follows that the corresponding U- and V-statistic have the same asymptotic distribution, namely

$$\sqrt{n}(V_n - T(F)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, m^2\sigma_1^2)\,. \tag{A.9}$$

In this thesis, we work only with U- and V- statistics of degree $m = 2$. According to (A.5) and (A.6), we can write compactly

$$\sigma_1^2 = \mathbb{V}\mathrm{ar}\Big[\mathbb{E}[\psi(x_1, X_2)]\Big|_{x_1 = X_1}\Big]\,. \tag{A.10}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

The result of main importance to us is the asymptotic multivariate normality of a vector of p degree-2 V-statistics, each with different kernel $\psi_i(X_1, X_2)$, $i = 1, ..., p$, but based on the same i.i.d sample $(X_1, ..., X_n)$, namely

$$\boldsymbol{V}_n = \left(V_n^1, V_n^2, ..., V_n^p\right)^\top , \tag{A.11}$$

where

$$V_n^i = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \psi_i(X_j, X_k). \tag{A.12}$$

$\boldsymbol{V}_n$ is considered as an estimator of a vector of functionals

$$\boldsymbol{T}(F) = \left(T^{(1)}, T^{(2)}, ..., T^{(p)}\right)^\top , \tag{A.13}$$

acting on the same distribution function F. The joint asymptotic normality of $\boldsymbol{V}_n$ is established on the ground of the Cramér-Wold theorem; see Sen and Singer (1993). We have then

$$\sqrt{n}(\boldsymbol{V}_n - \boldsymbol{T}(F)) \xrightarrow{\mathrm{d}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}) , \tag{A.14}$$

where $\boldsymbol{\Sigma}$ is the $p \times p$ symmetric covariance matrix whose (i, j)-th entry equals

$$\sigma_{i,j} = 4\mathbb{C}\mathrm{ov}\left[\mathbb{E}[\psi_i(x_1, X_2)]\Big|_{x_1 = X_1}, \mathbb{E}[\psi_j(x_1, X_2)]\Big|_{x_1 = X_1}\right] . \tag{A.15}$$

First-order degeneracy

If the formula (A.5) yields $\sigma_1^2 = 0$ but $\sigma_2^2 > 0$, then the U-statistic is said to possess the *first order degeneracy*. The asymptotic distribution of such U-statistic is not normal. In particular, for degree-2 U-statistic based on kernel $\psi(x_1, x_2) = f(x_1)f(x_2)$ with $\mathbb{E}[\psi(x_1, X_2)] = 0$ (hence $\sigma_1^2 = 0$), $\mathbb{E}^2[\psi(X_1, X_2)] < \infty$ and $\mathbb{E}[\psi(X_1, X_2)] = 0$ it can be shown that

$$n(U_n - T(F)) \xrightarrow{\mathrm{d}} \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1) , \tag{A.16}$$

where $Z_1, Z_2, ...$ are independent standard normal random variables, and λ_j are the eigenvalues of the integral equation

$$\int \psi(x_1, x_2) f(x_2) \mathrm{d}F(x_2) = \lambda f(x_1) . \tag{A.17}$$

For proof see Lee (1990), sec. 3.2.

This means that nU_n is asymptotically distributed as an infinite linear combination of chi-squared variables with one degree of freedom.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

APPENDIX B

Derivations of selected results

B.1 Covariance function of the ELT process

The covariance function (2.22) of the empirical Laplace transform of i.i.d. sample computed at two points s_1 and s_2 is derived as follows

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}[\widehat{L}_n(s_1), \widehat{L}_n(s_2)] &= \mathbb{E}[\widehat{L}_n(s_1)\widehat{L}_n(s_2)] - \mathbb{E}[\widehat{L}_n(s_1)]\mathbb{E}[\widehat{L}_n(s_2)] \\
&= \tfrac{1}{n^2}\mathbb{E}\Big[\textstyle\sum_{j=1}^n e^{-s_1 X_j} \sum_{k=1}^n e^{-s_2 X_k}\Big] - L(s_1)L(s_2) \\
&= \tfrac{1}{n^2}\mathbb{E}\Big[\textstyle\sum_{j=1}^n \sum_{k=1}^n e^{-s_1 X_j} e^{-s_2 X_k}\Big] - L(s_1)L(s_2) \\
&= \tfrac{1}{n^2}\textstyle\sum_{j=1}^n \mathbb{E}[e^{-(s_1+s_2)X_j}] + \tfrac{1}{n^2}\sum_{j\neq k}\mathbb{E}[e^{-s_1 X_j} e^{-s_2 X_k}] - L(s_1)L(s_2) \\
&= \tfrac{1}{n}\, L(s_1+s_2) + \tfrac{1}{n^2}\textstyle\sum_{j\neq k}\mathbb{E}[e^{-s_1 X_j}]\mathbb{E}[e^{-s_2 X_k}] - L(s_1)L(s_2)\,, \\
&= \tfrac{1}{n}\,[L(s_1+s_2) - L(s_1)L(s_2)]\,,
\end{aligned}
$$

where we use independence of X_j and X_k for $i \neq j$. □

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

B.2 Mean of the squared modulus of ECF

The expected value of the squared modulus of ECF (2.33) is derived as follows

$$
\begin{aligned}
\mathbb{E}[|\widehat{\phi}_n(s)|^2] &= \mathbb{E}\Big[\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\cos(s(X_j-X_k))\Big] \\
&= \mathbb{E}\Big[\frac{1}{n^2}\sum_{j=1}^{n}\cos(0)\Big]+\mathbb{E}\Big[\frac{2}{n^2}\sum\sum_{1\le j<k\le n}\cos(s(X_j-X_k))\Big] \\
&= \frac{1}{n}+\frac{n-1}{n}\mathbb{E}\Big[\cos(sX_j)\cos(sX_k)+\sin(sX_j)\sin(sX_k)\Big] \\
&= \frac{1}{n}+\frac{n-1}{n}[\widehat{u}_n^2(s)+\widehat{v}_n^2(s)] \\
&= |\phi(s)|^2+\frac{1-|\phi(s)|^2}{n}. \quad \square
\end{aligned}
$$

B.3 Derivation of the $\mathcal{L}^2$ estimators for the normal distribution

We obtain the expressions for the estimators (4.38) and (4.39) as follows.

By $J_n(\mu,\sigma^2)$ denote the integral in (4.37) to be minimized. Using the definition of the complex modulus we can write

$$
J_n(\mu,\sigma^2)=\int_{-\infty}^{\infty}\Big\{\mathrm{Re}^2\Big[(s\sigma^2-i\mu)\widehat{\phi}_n(s)-\widehat{\phi}_n'(s)\Big]+\mathrm{Im}^2\Big[(s\sigma^2-i\mu)\widehat{\phi}_n(s)-\widehat{\phi}_n'(s)\Big]\Big\}s^{\gamma}e^{-as^2}\mathrm{d}s.
$$

After some algebra we find that

$$
\mathrm{Re}^2\Big[(s\sigma^2-i\mu)\widehat{\phi}_n(s)+\widehat{\phi}_n'(s)\Big]=\Big[\frac{1}{n}\sum_{j=1}^{n}s\sigma^2\cos(sX_j)+(\mu-X_j)\sin(sX_j)\Big]^2,
$$

$$
\mathrm{Im}^2\Big[(s\sigma^2-i\mu)\widehat{\phi}_n(s)+\widehat{\phi}_n'(s)\Big]=\Big[\frac{1}{n}\sum_{j=1}^{n}s\sigma^2\sin(sX_j)-(\mu-X_j)\cos(sX_j)\Big]^2.
$$

Taking both squares, we obtain two double sums. Putting them back into the first equation and applying the "product-to-sum" trigonometric identities, we find that several terms cancel, and we obtain

$$
\begin{aligned}
J_n(\mu,\sigma^2) = \frac{1}{n^2}\int_{-\infty}^{\infty}\Big\{\sum_{j,k=1}^{n}s^2\sigma^4\cos\big(s(X_j-X_k)\big)-s\sigma^2(X_j-X_k)\sin\big(s(X_j-X_k)\big) \\
+(\mu-X_j)(\mu-X_k)\cos\big(s(X_j-X_k)\big)\Big\}s^{\gamma}e^{-as^2}\mathrm{d}s.
\end{aligned}
$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Now, we integrate analytically both for $\gamma = 0$ and $\gamma = 2$ using the formulas (C.8), (C.9), (C.11), (C.10) and (C.12) in Appendix C; we obtain

$$n^2 J_n(\mu, \sigma^2) = A_n\mu^2 + B_n\mu + C_n\sigma^4 + D_n\sigma^2 + E_n\,,$$

whereby for $\gamma = 0$ we have

$$\begin{aligned}
A_n &= \sum_{j,k=1}^{n} \sqrt{\frac{\pi}{a}} \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),\\
B_n &= -\sum_{j,k=1}^{n} \sqrt{\frac{\pi}{a}} (X_j+X_k) \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),\\
C_n &= -\sum_{j,k=1}^{n} \frac{1}{4}\sqrt{\frac{\pi}{a^3}} \Big(\frac{(X_j-X_k)^2}{a} - 2\Big) \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),\\
D_n &= -\sum_{j,k=1}^{n} \frac{1}{2}\sqrt{\frac{\pi}{a^3}} (X_j-X_k)^2 \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),
\end{aligned}$$

while for $\gamma = 2$ we have

$$\begin{aligned}
A_n &= -\sum_{j,k=1}^{n} \frac{1}{4}\sqrt{\frac{\pi}{a^3}} \Big(\frac{(X_j-X_k)^2}{a} - 2\Big) \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),\\
B_n &= \sum_{j,k=1}^{n} \frac{1}{4}\sqrt{\frac{\pi}{a^3}} (X_j+X_k) \Big(\frac{(X_j-X_k)^2}{a} - 2\Big) \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),\\
C_n &= \sum_{j,k=1}^{n} \frac{1}{4}\sqrt{\frac{\pi}{a^5}} \Big(\frac{(X_j-X_k)^4}{4a^2} - 3\frac{(X_j-X_k)^2}{a} + 3\Big) \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big),\\
D_n &= \sum_{j,k=1}^{n} \frac{1}{8}\sqrt{\frac{\pi}{a^7}} (X_j-X_k)^2 \exp\Big(-\frac{(X_j-X_k)^2}{4a}\Big) \Big[(X_j-X_k)^2 - 6a\Big].
\end{aligned}$$

The term E_n is in both cases independent of μ and σ^2.

$J_n(\mu, \sigma^2)$ has a unique minimizer being the solver of

$$\begin{cases}
\dfrac{\partial J_n(\mu, \sigma^2)}{\partial \sigma^2} = 2A_n\mu + B_n = 0,\\[2ex]
\dfrac{\partial J_n(\mu, \sigma^2)}{\partial \mu} = 2C_n\sigma^2 + D_n = 0,
\end{cases}$$

so that the estimator (4.37) equals

$$\widehat{\mu}_n^{\,\mathrm{CF}\,0,1} = -\frac{B_n}{2A_n}, \qquad \widehat{\sigma_n^2}^{\,\mathrm{CF}\,0,1} = -\frac{D_n}{2C_n}\,,$$

yielding the final forms (4.38) and (4.39). □

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

B.4 Derivation of the $\mathcal{L}^2$ estimators for the exponential distribution

First, we show how to derive the estimator $\widehat{\lambda}_n^{\mathrm{LT}\,2,1}$ in (4.70) from (4.61). By $J_n(\lambda)$ let us denote the integral in (4.61) to be minimized. Write it as follows

$$\begin{aligned} J_n(\lambda) &= \int_0^\infty \Big[(\lambda+s)\frac{1}{n}\sum_{j=1}^n X_j^2 e^{-sX_j} - \frac{2}{n}\sum_{j=1}^n X_j e^{-sX_j}\Big]^2 s^\gamma e^{-as}\mathrm{d}s \\ &= \int_0^\infty \frac{1}{n^2}\Big[\sum_{j=1}^n (\lambda+s)X_j^2 e^{-sX_j} - 2X_j e^{-sX_j}\Big]^2 s^\gamma e^{-as}\mathrm{d}s \\ &= \int_0^\infty \frac{1}{n^2}\sum_{j,k=1}^n \Big[(\lambda+s)X_j^2 e^{-sX_j} - 2X_j e^{-sX_j}\Big]\Big[(\lambda+s)X_k^2 e^{-sX_k} - 2X_k e^{-sX_k}\Big] s^\gamma e^{-as}\mathrm{d}s\,. \end{aligned}$$

Multiplying and collecting terms according to powers of s gives

$$\begin{aligned} J_n(\lambda) &= \frac{1}{n^2}\sum_{j,k=1}^n [\lambda^2 X_j^2 X_k^2 - 2\lambda(X_j X_k^2 + X_j^2 X_k)] \int_0^\infty s^\gamma e^{-as}\mathrm{d}s \\ &+ \frac{1}{n^2}\sum_{j,k=1}^n [2\lambda X_j^2 X_k^2 - 2\lambda(X_j X_k^2 + X_j^2 X_k)] \int_0^\infty s^{\gamma+1} e^{-as}\mathrm{d}s \\ &+ \frac{1}{n^2}\sum_{j,k=1}^n [X_j^2 X_k^2 - 4X_j X_k] \int_0^\infty s^{\gamma+2} e^{-as}\mathrm{d}s\,. \end{aligned} \tag{B.1}$$

By using the integral $\int_0^\infty s^k e^{-cs}\mathrm{d}s = \frac{k!}{c^{k+1}}$, $k = 0,1,2,\ldots$, $c > 0$, rearranging and dropping the terms which do not depend on λ, we find that

$$\widehat{\lambda}_n^{\mathrm{LT}\,2,1} = \operatorname*{argmin}_{\lambda>0}\Bigg[\lambda^2 \underbrace{\sum_{j,k=1}^n \frac{\gamma! X_j^2 X_k^2}{(X_j+X_k+a)^{\gamma+1}}}_{A_n} + \lambda \underbrace{\sum_{j,k=1}^n \Bigg(\frac{2(\gamma+1)! X_j^2 X_k^2}{(X_j+X_k+a)^{\gamma+2}} - \frac{2\gamma!(X_j^2 X_k + X_j X_k^2)}{(X_j+X_k+a)^{\gamma+1}}\Bigg)}_{B_n}\Bigg].$$

The unique minimizer is $-B_n/2A_n$ and is equivalent to final form (4.70).

Next, we sketch how $\widehat{\lambda}_n^{\mathrm{CF}\,1,0}$ in (4.72) is derived from (4.64). With

$$\begin{aligned} \widehat{\phi}_n(s) &= \frac{1}{n}\sum_{j=1}^n \cos(sX_j) + i\frac{1}{n}\sum_{j=1}^n \sin(sX_j)\,, \\ \widehat{\phi}_n'(s) &= -\frac{1}{n}\sum_{j=1}^n X_j\sin(sX_j) + i\frac{1}{n}\sum_{j=1}^n X_j\cos(sX_j)\,, \end{aligned}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

the squared modulus in the integrand of (4.64) can be written as

$$\Big|\ \cdot\ \Big|^2 = \Big[\frac{1}{n}\sum_{j=1}^{n}(1-\lambda X_j)\cos(sX_j) - sX_j\sin(sX_j)\Big]^2 + \Big[\frac{1}{n}\sum_{j=1}^{n}(1-\lambda X_j)\sin(sX_j) + sX_j\cos(sX_j)\Big]^2 .$$

Taking squares, grouping the terms, and applying the trigonometric "product-to-sum" identities leads to convenient cancellations after which we have

$$\Big|\ \cdot\ \Big|^2 = \frac{1}{n^2}\sum_{j,k=1}^{n}(1-\lambda X_j)(1-\lambda X_k)\cos\Big(s(X_j - X_k)\Big) + s^2 X_j X_k \cos(sX_j)\cos(sX_k)\,.$$

We drop the second term in the above sum (which is not a function of λ) and we find that (4.64) equals

$$\widehat{\lambda}_n^{\mathrm{CF}\,1,0} = \operatorname*{argmin}_{\lambda>0}\sum_{j,k=1}^{n}(1-\lambda X_j)(1-\lambda X_k)\underbrace{\int_{-\infty}^{\infty}\cos\Big(s(X_j - X_k)\Big)s^{\gamma}e^{-as^2}\mathrm{d}s}_{A_n}\,.$$

The function to be minimized is quadratic in λ. Hence

$$\widehat{\lambda}_n^{\mathrm{CF}\,1,0} = \frac{\sum_{j,k=1}^{n}(X_j+X_k)A_n}{2\sum_{j,k=1}^{n}X_jX_kA_n}. \tag{B.2}$$

To compute A_n for $\gamma = 0$ and $\gamma = 2$, we apply integrals (C.8) and (C.10) from Appendix C. They yield, respectively,

$$A_n = \sqrt{\frac{\pi}{a}}\exp\Big(-(X_j-X_k)^2/4a\Big), \quad \text{for } \gamma = 0\,,$$
$$A_n = -\frac{1}{4}\sqrt{\frac{\pi}{a^3}}\exp\Big(-(X_j-X_k)^2/4a\Big)\Big[(X_j-X_k)^2/a - 2\Big], \quad \text{for } \gamma = 2\,.$$

Putting these expressions into (B.2) gives the final expressions for the estimator in (4.72). □

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

APPENDIX C

Definite integrals used in the thesis

For any $a, \gamma > 0,\ \ b \in \mathbb{R},\ \ k = 0, 1, 2, ...$

$$\int_0^\infty x^k e^{-ax} \mathrm{d}x = \frac{k!}{a^{k+1}} . \tag{C.1}$$

$$\int_{-\infty}^\infty e^{-ax^2} \mathrm{d}x = \sqrt{\frac{\pi}{a}} . \tag{C.2}$$

$$\int_{-\infty}^\infty e^{-ax^2+bx} \mathrm{d}x = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a}\right) . \tag{C.3}$$

$$\int_{-\infty}^\infty x e^{-ax^2+bx} \mathrm{d}x = \frac{b}{2}\sqrt{\frac{\pi}{a^3}} \exp\left(\frac{b^2}{4a}\right) . \tag{C.4}$$

$$\int_{-\infty}^\infty x^2 e^{-ax^2} \mathrm{d}x = \frac{1}{2}\sqrt{\frac{\pi}{a^3}} . \tag{C.5}$$

$$\int_{-\infty}^\infty x^2 e^{-ax^2+bx} \mathrm{d}x = \frac{(2a+b^2)}{4}\sqrt{\frac{\pi}{a^5}} \exp\left(\frac{b^2}{4a}\right) . \tag{C.6}$$

$$\int_0^\infty x^k e^{-ax^\gamma} \mathrm{d}x = \frac{1}{\gamma} a^{-(k+1)/\gamma}\, \Gamma\left(\frac{k+1}{\gamma}\right) . \tag{C.7}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

$$\int_{-\infty}^{\infty} \cos(bx) e^{-ax^2} \mathrm{d}x = \sqrt{\frac{\pi}{a}} \exp\left(-\frac{b^2}{4a}\right). \tag{C.8}$$

$$\int_{-\infty}^{\infty} x \sin(bx) e^{-ax^2} \mathrm{d}x = \frac{b}{2}\sqrt{\frac{\pi}{a^3}} \exp\left(-\frac{b^2}{4a}\right). \tag{C.9}$$

$$\int_{-\infty}^{\infty} x^2 \cos(bx) e^{-ax^2} \mathrm{d}x = -\frac{1}{4}\sqrt{\frac{\pi}{a^3}} \exp\left(-\frac{b^2}{4a}\right)\left(\frac{b^2}{a} - 2\right). \tag{C.10}$$

$$\int_{-\infty}^{\infty} x^3 \sin(bx) e^{-ax^2} \mathrm{d}x = -\frac{1}{8}\sqrt{\frac{\pi}{a^7}} \exp\left(-\frac{b^2}{4a}\right) b(b^2 - 6a). \tag{C.11}$$

$$\int_{-\infty}^{\infty} x^4 \cos(bx) e^{-ax^2} \mathrm{d}x = \frac{1}{4}\sqrt{\frac{\pi}{a^5}} \exp\left(-\frac{b^2}{4a}\right)\left(\frac{b^4}{4a^2} - \frac{3b^2}{a} + 3\right). \tag{C.12}$$

The formulas can be found in Gradshteyn and Ryzhik (2015) and Poularikas (2010).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

APPENDIX D

Mathematical symbols and abbreviations

D.1 Mathematical symbols

Ω	sample space
$\mathcal{F}$	σ-algebra of events
$P(A)$	probability of event A
$I\{\cdot\}$	indicator function
X, Y, Z, K	random variables
n	sample size
X_j	j-th element of the random sample $(X_1, ..., X_n)$
$X_{(j)}$	j-th order statistic
$\boldsymbol{X}$	random vector
$\boldsymbol{X}_j$	j-th element of the random sample $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$
$\boldsymbol{\theta}$	vector of parameters
$\widehat{\boldsymbol{\theta}}_n$	estimator of $\boldsymbol{\theta}$
$F(x), F_{\boldsymbol{\theta}}(x), F(x\|\boldsymbol{\theta})$	cumulative distribution function (cdf) with parameter $\boldsymbol{\theta}$
$f(x), f_{\boldsymbol{\theta}}(x), f(x\|\boldsymbol{\theta})$	probability density function (pdf) with parameter $\boldsymbol{\theta}$
$f_X(x), f_X(x\|\boldsymbol{\theta}), f_Y(y), f_Y(y\|\boldsymbol{\theta})$	probability density functions of specific variables X, Y
$p(k)$, $p_{\boldsymbol{\theta}}(k)$	probability mass function (pmf) of a discrete variable K
$\widehat{F}_n(x)$	empirical distribution function
$\widehat{f}_n(x)$	estimated density function
$F(x\|\widehat{\boldsymbol{\theta}}_n)$	parametrically estimated cumulative distribution function
$f(x\|\widehat{\boldsymbol{\theta}}_n)$	parametrically estimated density function

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

$F_{\boldsymbol{\theta}}(\boldsymbol{x})$	multivariate cumulative distribution function of a vector $\boldsymbol{X}$
$\widehat{F}_n(\boldsymbol{x})$	multivariate empirical distribution function
$\mathbb{E}[X], \mathbb{E}_{\boldsymbol{\theta}}[X]$	expectation of a random variable X
$\mathbb{E}[X^k], \mathbb{E}_{\boldsymbol{\theta}}[X^k]$	k-th raw moment of a random variable X
$\mathbb{E}[Y\|x]$	conditional expectation of Y given $X = x$
$\mathrm{Med}[X]$	median of a random variable X
$\mathbb{V}[X]$	variance of a random variable X
$\mathbb{V}_{\mathrm{as}}[\,\cdot\,]$	asymptotic variance (of a statistic or estimator)
$\mathbb{C}\mathrm{ov}[\,\cdot\,,\,\cdot\,]$	covariance
$\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_n$	covariance matrix, estimated covariance matrix
$T(F)$	functional of a distribution function F
$\mathrm{IF}[x;T]$	influence function of a functional T at point x
$\mathcal{T}(s)$	integral transform
$K(s,x)$	kernel of an integral transform
s, t	transform variable
$\boldsymbol{s}$	vector of transform variables
$\mathcal{D}$	interval of convergence of a transform
$\mathcal{D}'$, $\mathcal{D}^I$	subinterval of $\mathcal{D}$, interior of $\mathcal{D}$
$\widehat{\mathcal{T}}_n(s)$	empirical integral transform
$\mathcal{T}(s\|\widehat{\boldsymbol{\theta}}_n)$	parametrically estimated integral transform
$L(s), L_{\boldsymbol{\theta}}(s), L(s\|\boldsymbol{\theta})$,	Laplace transform (LT) of $F_{\boldsymbol{\theta}}(x)$
σ_-, σ_+	abscissa of convergence of Laplace transform
$\widehat{L}_n(s)$	empirical Laplace transform (ELT)
$L(s\|\widehat{\boldsymbol{\theta}}_n)$	parametrically estimated Laplace transform
$\phi(s), \phi_{\boldsymbol{\theta}}(s), \phi(s\|\boldsymbol{\theta})$	characteristic function (CF) of $F_{\boldsymbol{\theta}}(x)$
$u(s), u_{\boldsymbol{\theta}}(s)$	real part of characteristic function
$v(s), v_{\boldsymbol{\theta}}(s)$	imaginary part of characteristic function
$\|\phi(s)\|^2$	squared modulus of characteristic function
$\widehat{\phi}_n(s)$	empirical characteristic function (ECF)
$\widehat{u}_n(s)$	real part of empirical characteristic function
$\widehat{v}_n(s)$	imaginary part of empirical characteristic function
$\|\widehat{\phi}_n(s)\|^2$	squared modulus of empirical characteristic function
$\phi(\boldsymbol{s})$	multivariate characteristic function
$\widehat{\phi}_n(\boldsymbol{s})$	multivariate empirical characteristic function
$G(s)$, $G_{\boldsymbol{\theta}}(s)$	probability generating function (PGF) of pmf $p_{\boldsymbol{\theta}}(k)$
$\widehat{G}_n(s)$	empirical probability generating function (EPGF)
$\Psi_{\boldsymbol{\theta}}(s)$	cumulant generating function (CGF)
$\eta(x\|\boldsymbol{\theta})$	regression function
$\mathscr{L}(s\|\boldsymbol{\theta})$	Laplace transform of a regression function
$\mathscr{L}_{\mathrm{ec}}(s\|\boldsymbol{\theta})$	end-corrected Laplace transform of a regression function

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

$\widehat{\mathscr{L}}_n(s)$	empirical Laplace transform of a regression function
T_n	statistic, test statistic
$\mathrm{Re}\{z\}$	real part of a complex number z
$\mathrm{Im}\{z\}$	imaginary part of a complex number z
$\overline{z}$	complex conjugate of z
i	imaginary unit
$\rho(\,\cdot\,,\,\cdot\,)$	distance or divergence measure
$\xrightarrow{\mathrm{d}}$	convergence in distribution
$\xrightarrow{\mathrm{a.s.}}$	strong convergence
$O(\cdot)$	big 'Oh', $f(x) = O(g(x))$ if $\lim_{x\to 0} f(x)/g(x) = \text{const.}$
$\mathcal{N}$	normal distribution
B	number of bootstrap replications
$\dim(\,\cdot\,)$	dimension of a vector
$\overline{X}_n$, $\overline{Z}_n$, $\overline{K}_n$,	sample means
$\widehat{\mathrm{med}}_n$	sample median
S_n^2, $\widehat{\mathbb{V}}_n[Z]$, $\widehat{\mathbb{V}}_n[K]$	sample variances
$\widehat{Sk}_n$	sample skewness
$\widehat{K}_n$	sample kurtosis
MAD_n	median absolute deviation
Q_n	the Q_n estimator
V_n, U_n	V-statistic, U-statistic
$A_n, B_n, C_n, D_n, E_n, J_n$	auxilary V-statistics
$\psi(X_1, ..., X_m)$	kernel of a degree-m V- or U-statistic
$\mu, \sigma, \lambda, \alpha, \beta, \vartheta, \gamma, a, b, c, q$	parameters of probability distributions
$\boldsymbol{\mu}$	vector of means or location parameters
$\sigma_1^2, \sigma_2^2, ...$	auxiliary variances / variance components
$\boldsymbol{\Omega}$, $\boldsymbol{K}$, $\boldsymbol{\nu}$	auxiliary matrices and vectors
$i, j, k, \ell, r, m, c, p, d$	integer constants, indices, dimensions
$\gamma, \delta, \eta, a, b, k, q, C$	real constants
$W(s)$	weight function of $\mathcal{L}^2$ distances in transform domain
a, γ	scale and shape parameters of $W(s)$
$h(\,\cdot\,), g(\,\cdot\,), w(\,\cdot\,), \xi(\,\cdot\,)$	auxiliary functions
$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m}$	estimator based on a differential equation of LT involving its r-th and m-th derivative
$\widehat{\boldsymbol{\theta}}_n^{\mathrm{CF}\,r,m}$	estimator based on a differential equation of CF involving its r-th and m-th derivative
$\widehat{\boldsymbol{\theta}}_n^{\mathrm{LT}\,r,m,\ell}$	estimator based on a differential equation of LT involving its r-th, m-th and ℓ-th derivative
$\mathcal{H}_0$	null-hypothesis

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

$\mathcal{P}$	power of a test
$\widehat{\mathcal{P}}_n$	estimated power for sample size n
$\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_+^0$	real numbers, positive and non-negative real numbers
$\mathbb{N}_+^0$, $\mathcal{K}$	non-negative integers, subset of integers
$\mathrm{erfc}(\,\cdot\,)$	complementary error function
$\mathrm{B}(\,\cdot\,,\,\cdot\,)$	beta function
$\Gamma(\,\cdot\,)$, $\Gamma(\,\cdot\,,\,\cdot\,)$	gamma function, lower incomplete gamma function
$\mathrm{K}_p(\cdot)$	modified Bessel function of 2nd kind
$\psi(\cdot)$	digamma function
$E_1(\cdot)$	exponential integral
$[f(x)]^{k\circledast}$	k-fold self-convolution a function
$\tilde{f}(x)$	saddlepoint approximation of density function $f(x)$
$s_\mathrm{o}, s_\mathrm{o}(x)$	saddlepoint

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

D.2 Abbreviations

AD	Anderson-Darling (test)
ARE	asymptotic relative efficiency
CF	characteristic function
CGF	cumulant generating function
ch.	chapter
CLT	Central Limit Theorem
CM	Crámer-von Mises (test, estimator)
CV	cross-validation
ECF	empirical characteristic function
EI	empirical influence
ELT	empirical Laplace transform
EPGF	empirical probability generating function
GoF	goodness-of-fit (test)
i.i.d.	independent identically distributed
i.o.c.	interval of convergence
IF	influence function
JB	Jarque-Bera (test)
KLD	Kullback-Leibler divergence
KS	Kolmogorov-Smirnov (test)
LT	Laplace transform
MAD	median absolute deviation
MGF	moment generating function
MISE	mean integrated squared error
MIAE	mean integrated absolute error
MKLD	mean Kullback-Leibler divergence
ML	maximum likelihood (estimator)
MD	minimum distance (estimator)
MM	method-of-moments (estimator)
MSE	mean squared error
ODE	ordinary differential equation
PGF	probability generating function
sec.	section, subsection
SLLN	Strong Law of Large Numbers
s.t.	such that
SW	Shapiro-Wilk (test)
w.r.t.	with respect to

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Bibliography

J. Abate and W. Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10:5–87, 1992.

J. Abate and W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, 7(1):36–43, 1995.

J. Abate and W. Whitt. An operational calculus for probability distributions via Laplace transforms. *Advances in Applied Probability*, 28(1):75–113, 1996.

B. C. Arnold. *Pareto Distributions*. 2nd Ed. Monographs on Statistics and Applied Probability 140. CRC Press, 2015.

S. Asmussen, J. Jensen, and L. Rojas-Nandayapa. On the Laplace transform of the lognormal distribution. *Methodology and Computing in Applied Probability*, 18(2): 441–458, 2016.

L. Baringhaus and N. Henze. A class of consistent tests for exponentiality based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, 43(3): 551–564, 1991.

L. Baringhaus and N. Henze. A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statistics and Probability Letters*, 13(4):269–274, 1992.

L. Baringhaus, N. Gürtler, and N. Henze. Weighted integral test statistics and components of smooth tests of fit. *Australian and New Zealand Journal of Statistics*, 42(2): 179–192, 2000.

V. Barnett. *Comparative Statistical Inference*. 3rd Ed. Wiley, 1999.

V. Barnett and T. Lewis. *Outliers in Statistical Data*. 3rd Ed. Wiley, 1994.

A. Basu, S. Shioya, and C. Park. *Statistical Inference. The Minimum Distance Approach*. Chapman & Hall/CRC, 2011.

A. Batsidis, M. D. Jiménez-Gamero, and A. J. Lemonte. On goodness-of-fit tests for the Bell distribution. *Metrika*, 83:297–319, 2020.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

P. T. Besbeas and B. J. T. Morgan. Integrated squared error estimation of normal mixtures. *Computational Statistics and Data Analysis*, 43:517–526, 2003.

S. Betsch, B. Ebner, and B. Klar. Minimum L^q-distance estimators for non-normalized parametric models. *Canadian Journal of Statistics*, 49(2):514–554, 2021.

K. O. Bowman and L. R. Shenton. *Properties of Estimators for the Gamma Distribution*. Marcel Dekker, 1988.

G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.

M. P. Braun, S. G. Meintanis, and V. B. Melas. Optimal design approach to GMM estimation of parameters based on empirical transforms. *International Statistical Review*, 76(3):387–400, 2008.

R. W. Butler. *Saddlepoint Approximations with Applications*. Cambridge University Press, 2007.

E. P. Campbell. *Robustness of Estimation Based on Empirical Transforms*. PhD Thesis. University of Kent at Canterbury, United Kingdom, 1992.

E. P. Campbell. Influence for empirical transforms. *Communications in Statistics - Theory and Methods*, 22(9):730–752, 1993.

R. Cao, A. Cuevas, and R. Fraiman. Minimum distance density-based estimation. *Computational Statistics and Data Analysis*, 20(6):611–631, 1995.

G. Casella and R. Berger. *Statistical Inference*. 2nd Ed. Duxbury, 2002.

J. D. Chergui. The integrated squared error estimation of parameters. *Extracta Mathematicae*, 11(3):435–442, 1996.

B. R. Clarke, P. L. McKinnon, and G. Riley. A fast robust method for fitting gamma distributions. *Statistical Papers*, 53(4):1001–1014, 2012.

A. M. Cohen. *Numerical Methods for Laplace Transform Inversion*. Springer, 2007.

P. C. Consul and F. Famoye. *Lagrangian Probability Distributions*. Birkhäuser, 2006.

D. R. Cox. *Renewal Theory*. Methuen, 1970.

S. Csörgő. Multivariate empirical characteristic functions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 55:203–229, 1981.

S. Csörgő. The empirical moment generating function. In *Nonparametric Statistical Inference. Colloquia Mathematica Societatis Janos Bolyai 32*, pages 139–150. North-Holland, Amsterdam, 1982.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

S. Csörgő. Adaptive estimation of the parameters of stable laws. In *Limit Theorems in Probability and Statistics. Colloquia Mathematica Societatis Janos Bolyai 36*, pages 305–368. North-Holland, Amsterdam, 1984.

M. D. Cubiles-de-la Vega, R. Pino-Mejías, and E. L. Silva-Ramírez. Clustering multivariate empiric characteristic functions for multi-class SVM classification. *International Journal of Electrical and Computer Engineering*, 64:380–384, 2012.

H. Daniels. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25:631–650, 1954.

V. A. Ditkin and A. P. Prudnikov. *Integral Transforms and Operational Calculus*. Pergamon Press, 1965.

M. M. Dowling and M. Nakamura. Estimating parameters for discrete distributions via the empirical probability generating function. *Communications in Statistics - Simulation and Computation*, 26:301–313, 1997.

T. Duchesne, J. Rioux, and A. Luong. Minimum Cramér-von Mises distance methods for complete and grouped data. *Communications in Statistics - Theory and Methods*, 26(2):401–420, 1997.

R. D'Agostino and M. Stephens. *Goodness-of-fit techniques*. Marcel Dekker, 1986.

B. Efron and T. Hastie. *Computer Age Statistical Inference*. Cambridge University Press, 2016.

T. W. Epps. Characteristic functions and their empirical counterparts: geometrical interpretations and applications to statistical inference. *The American Statistician*, 47(1):33–38, 1993.

T. W. Epps. Tests for location-scale families based on the empirical characteristic function. *Metrika*, 62(1):99–114, 2005.

T. W. Epps and L. B. Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70(3):723–726, 1983.

T. W. Epps and L. B. Pulley. Parameter estimates and tests of fit for infinite mixture distributions. *Communications in Statistics - Theory and Methods*, 14(12):3125–3145, 1985.

Y. Fang. GMM tests for the Katz family of distributions. *Journal of Statistical Planning and Inference*, 110:55–73, 2003.

S. J. Farlow. *An Introduction to Differential Equations and Their Applications*. McGraw-Hill, 1994.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

P. D. Feigin, R. L. Tweedie, and C. Belyea. Weighted area techniques for explicit parameter estimation in multi-stage models. *Australian Journal of Statistics*, 25:1–16, 1983.

W. Feller. *An Introduction to Probability Theory and Its Applications*. Vol. I. Wiley, 1968.

W. Feller. *An Introduction to Probability Theory and Its Applications*. Vol. II. Wiley, 1971.

K. Fergusson and P. Platen. On the distributional characterization of log-returns of a world stock index. *Applied Mathematical Finance*, 13(1):19–38, 2006.

A. Feuerverger and P. McDunnough. On statistical transform methods and their efficiency. *The Canadian Journal of Statistics*, 12(4):307–317, 1984.

A. Feuerverger and R. A. Mureika. The empirical characteristic function and its applications. *The Annals of Statistics*, 5(1):88–97, 1977.

C. Field and E. Ronchetti. *Small Sample Asymptotics. Lecture Notes-Monograph Series. Vol. 13.* Institute of Mathematical Statistics. Hayward, California, 1990.

K. Fragiadakis and S. G. Meintanis. Goodness-of-fit tests for multivariate Laplace distributions. *Mathematical and Computer Modelling*, 53(5-6):769–779, 2011.

R. E. Gaunt. A simple proof of the characteristic function of Student's t-distribution. *Communications in Statistics - Theory and Methods*, 50(14):3380–3383, 2021.

P. O. Goffard, S. Rao Jammalamadaka, and S. G. Meintanis. Goodness-of-fit tests for compound distributions with applications in insurance. *Journal of Statistical Theory and Practice*, 16(3), 2022.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. 8th Ed. Elsevier & Academic Press, 2015.

J. Grandell. *Mixed Poisson Processes*. CRC Press, 1997.

N. Gürtler and N. Henze. Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Annals of the Institute of Statistical Mathematics*, 52(2):267–286, 2000.

J. Hájek. Asymptotic normality of simple linear rank statistics under alternatives. *Annals of the Institute of Statistical Mathematics*, 39:325–346, 1968.

F. R. Hampel, P. J. Rousseeuw, E. Ronchetti, and W. Stahel. *Robust Statistics. The Approach Based on Influence Functions*. Wiley, 1986.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Annals of the Institute of Statistical Mathematics*, 50(4):1029–1054, 1982.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning*. Springer, 2008.

C. R. Heathcote. The integrated squared error estimation of parameters. *Biometrika*, 64(2):255–264, 1977.

N. Henze. Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Communications in Statistics - Theory and Methods*, 22(1): 115–133, 1992.

N. Henze and B. Klar. Goodness-of-fit tests for the inverse gaussian distribution based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, 54(2):425–444, 2002.

N. Henze and S. Koch. On a test of normality based on the empirical moment generating function. *Stat Papers*, 61:17–29, 2020.

N. Henze and T. Wagner. A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, 62:1–23, 1997.

N. Henze, S. G. Meintanis, and B. Ebner. Goodness-of-fit tests for the gamma distribution based on the empirical Laplace transform. *Communications in Statistics - Theory and Methods*, 41(9):1543–1556, 2012.

Z. Hlávka, M. Hušková, and S. G. Meintanis. Tests for independence in non-parametric heteroscedastic regression models. *Journal of Multivariate Analysis*, 102(4):816–827, 2011.

Z. Hlávka, M. Hušková, and S. G. Meintanis. Change point detection with multivariate observations based on characteristic functions. In *From Statistics to Mathematical Finance. Festschrift in Honour of Winfried Stute*, pages 273–290. Springer, 2017.

J. L. Hodges and E. L. Lehmann. Estimates of location based on rank tests. *Annals of Mathematical Statistics*, 34(2):598–611, 1963.

W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19:293–325, 1948.

J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29:339–349, 1987.

J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme-value distribution by the method of probability–weighted moments. *Technometrics*, 27:251–261, 1985.

T. Hołyński. Robust estimation of block-error ratio under excessive noise based on empirical probability generating function. In *42nd IEEE International Conference on Telecommunications and Signal Processing (TSP'19), Budapest, Hungary*, pages 705–711, 2019a.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

T. Hołyński. Goodness-of-fit test for (log-) normality based on a derivative property of the log-normal Laplace transform. In *4th Workshop on Goodness-of-Fit, Change Point and Related Problems (GOFCP 2019), Trento, Italy. Workshop booklet*, pages 33–34, 2019b.

T. Hołyński and J. Haneczok. Robustness-efficiency trade-offs of minimum distance estimators for the Poisson-lognormal distribution. In *4th Conference of the International Society for Nonparametric Statistics, Salerno, Italy. Conference book. ISBN 978-88-61970-00-7*, pages 76–77, 2018.

D. R. Hunter. *Notes for a Graduate-level Course in Asymptotics for Statisticians*. Penn State University, https://sites.math.rutgers.edu/ sg1108/asymp1.pdf, 2014.

A. V. Huzurbazar. *Flowgraph Models for Multistate Time-to-Event Data*. Wiley, 2005.

M. Hušková, N. Neumeyer, T. Niebuhr, and L. Selk. Specification testing in nonparametric AR-ARCH models. *Scandinavian Journal of Statistics*, 46(1):26–58, 2019.

T. Y. Hwang and P. H. Huang. On new moment estimation of parameters of the gamma distribution using its characterization. *Annals of the Institute of Statistical Mathematics*, 54(4):840–847, 2002.

C. M. Jarque and A. K. Bera. A test for normality of observations and regression residuals. *International Statistical Review*, 55(2):163–172, 1987.

M. D. Jiménez-Gamero and M. V. Alba-Fernández. Testing for the Poisson–Tweedie distribution. *Mathematics and Computers in Simulation*, 164:146–162, 2019.

M. D. Jiménez-Gamero and A. Batsidis. Minimum distance estimators for count data based on the probability generating function with applications. *Metrika*, 80:503–545, 2017.

M. D. Jiménez-Gamero, M. V. Alba-Fernández, J. Muñoz-García, and Y. Chalco-Cano. Goodness-of-fit tests based on empirical characteristic functions. *Computational Statistics and Data Analysis*, 53(12):3957–3971, 2009.

M. D. Jiménez-Gamero, A. Batsidis, and M. V. Alba-Fernández. Fourier methods for model selection. *Annals of the Institute of Statistical Mathematics*, 68(1):105–133, 2016.

N. L. Johnson, S. Kotz, and A. Kemp. *Univariate Discrete Distributions*. 2nd. Ed. Wiley, 1993.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*. Vol. 1. Wiley, 1994.

N. L. Johnson, A. Kemp, and S. Kotz. *Univariate Discrete Distributions*. 3rd. Ed. Wiley, 2005.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

B. Jørgensen. *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Lecture Notes in Statistics. Vol. 9. Springer, 1982.

J. Jurečková, J. Picek, and M. Schindler. *Robust Statistical Methods with R.* 2nd Ed. CRC Press, 2019.

A. Kankainen. *Consistent Testing of Total Inpendence Based on the Empirical Characteristic Function*. Ph.D. dissertation. University of Jyväkylä, Finnland. Published as: Jyväkylä Studies in Computer Science, Economics and Statistics. Vol. 29, 1995.

T. Kato and L. B. McLeod. The functional-differential equation $y'(x) = ay(\lambda x) + by(x)$. *Bulletin of the American Mathematical Society*, 77:891–937, 1971.

T. Kawata. *Fourier Analysis in Probability Theory*. Academic Press, 1972.

C. D. Kemp and A. W. Kemp. Rapid estimation for discrete distributions. *The Statistician*, 37(3):213–255, 1988.

L. Kleinrock. *Queueing Systems: Volume I - Theory*. Wiley Interscience, 1975.

L. Kleinrock. *Queueing Systems: Volume II - Computer Applications*. Wiley Interscience, 1976.

S. Kotz, N. Balakrishnan, and N. L. Johnson. *Continuous Multivariate Distributions*. Vol. 1. Wiley, 2000.

I. A. Koutrouvelis. Regression-type estimation of parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918–928, 1980.

I. A. Koutrouvelis and G. C. Canavos. Estimation in the three-parameter gamma distribution based on the empirical moment generation function. *Journal of Statistical Computation and Simulation*, 59(1):47–62, 1997.

I. A. Koutrouvelis and S. G. Meintanis. Estimating the parameters of Poisson-exponential models. *Australian and New Zealand Journal of Statistics*, 44(2):233–245, 2002.

I. A. Koutrouvelis, G. C. Canavos, and S. G. Meintanis. Estimation in the three-parameter inverse Gaussian distribution. *Computational Statistics and Data Analysis*, 49:1132–1147, 2005.

A. F. Laurence and B. J. T. Morgan. Selection of the transformation variable in the Laplace transfom method of estimation. *Australian Journal of Statistics*, 29:113–127, 1987.

A. J. Lee. *U-Statistics: Theory and Practice*. CRC Press, 1990.

M. I. Leedow and R. L. Tweedie. Weighted area techniques for the estimation of the parameters of a growth curve. *Australian Journal of Statistics*, 25:310–320, 1983.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

R. B. Leipnik. On lognormal random variables: I. The characteristic function. *Journal of the Australian Mathematical Society. Series B*, 32:327–347, 1991.

E. Lukacs. *Characteristic Functions*. 2nd ed. Hafner Pub. Co., 1970.

A. Luong and M. E. Thompson. Minimum-distance methods based on quadratic distances for transforms. *The Canadian Journal of Statistics*, 15(3):239–251, 1987.

H. Madsen. *Time Series Analysis*. Chapman & Hall/CRC, 2007.

A. Marcheselli, A. Baccini, and L. Barabesi. Minimum-distance methods based on quadratic distances for transforms. *Communications in Statistics - Theory and Methods*, 37(6):815–830, 2008.

M. B. Marcus. Weak convergence of the empirical characteristic function. *The Annals of Probability*, 9(2):194–201, 1981.

M. Matsui and A. Takemura. Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Annals of the Institute of Statistical Mathematics*, 57(1):183–199, 2005.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Ed. Princeton University Press, 2015.

S. G. Meintanis. Large-sample sample tests in negative binomial distributions with both parameters unknown. *InterStat*, June:13–19, 2003.

S. G. Meintanis. Transform methods for testing the negative binomial hypothesis. *Statistica*, 65(3):293–300, 2005.

S. G. Meintanis. New inference procedures for generalized Poisson distributions. *Journal of Applied Statistics*, 35(7):751–762, 2008.

S. G. Meintanis. A review of testing procedures based on the empirical characteristic function. *South African Statistical Journal*, 50(1):1–14, 2016.

S. G. Meintanis and G. S. Donatos. On robustness and efficiency of certain statistics involving the empirical characteristic function. *Journal of the Italian Statistical Society*, 5(1):149–161, 1996.

S. G. Meintanis and G. Iliopoulos. Characterizations of the exponential distribution based on certain properties of its characteristic function. *Kybernetika*, 39(3):295–298, 2003a.

S. G. Meintanis and G. Iliopoulos. Tests of fit for the Rayleigh distribution based on the empirical laplace transform. *Annals of the Institute of Statistical Mathematics*, 55(1): 137–151, 2003b.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

S. G. Meintanis, J. Ngatchou-Wandji, and E. Taufer. Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function. *Journal of Multivariate Analysis*, 140:171–192, 2015.

S. Nadarajah and S. Kotz. On the Laplace transform of the Pareto distribution. *Queueing Systems*, 54:243–244, 2006.

M. Nasser and M. Alam. Estimators of influence function. *Communications in Statistics - Theory and Methods*, 35(1):21–32, 2006.

T. Nombebe, J. S. Allison, L. Santana, and I. J. H. Visagie. On fitting the Lomax distribution: a comparison between minimum distance estimators and other estimation techniques. *https://arxiv.org/abs/2207.06086*, 2022.

A. S. Paulson and C. E. Lawrence. *Some modified integrated squared error procedures for multivariate normal data*. U.S. Army Research Office, 1982.

A. S. Paulson and E. H. Nicklin. Integrated distance estimators for linear models applied to some published data sets. *Applied Statistics*, 32(1):32–50, 1983.

A. S. Paulson, E. W. Holcomb, and R. A. Leitch. The estimation of the parameters of the stable law. *Biometrika*, 62:163–169, 1975.

K. Pearson. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. Series A*, 186:343–414, 1895.

K. A. Penson and K. Górska. On the Laplace transform of the Fréchet distribution. *Journal of Mathematical Physics*, 55(9), 2014.

E. J. G. Pitman. On the behaviour of the characteristic function of a probability distribution in the neighbourhood of the origin. *Journal of the Australian Mathematical Society*, 8(3):423–443, 1968.

A. D. Poularikas. *Transforms and Applications Handbook*. 3rd Ed. CRC Press, 2010.

N. U. Prabhu. *Stochastic Storage Processes: Queues, Insurance Risk and Dams*. Springer, 1980.

S. J. Press. Estimation in univariate and multivariate stable distribution. *Journal of the American Statistical Association*, 67(340):842–846, 1972.

J. Pudełko. A family of goodness-of-fit tests for the Cauchy distribution. *Technical Transactions of the Cracow University of Technology*, 109(18):13–26, 2012.

R. E. Quandt and J. B. Ramsey. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73:730–738, 1978.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

B. Ranneby. The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, 11(2):93–112, 1984.

T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels. *Stochastic Processes for Insurance and Finance*. Wiley, 1999.

P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 82(424):1273–1284, 1993.

R. Rueda, V. Pérez-Abreu, and F. O'Reilly. Goodness of fit for the Poisson distribution based on the probability generating function. *Communications in Statistics - Theory and Methods*, 20(10):3093–3110, 1991.

M. Rytgaard. Estimation in the Pareto distribution. *ASTIN Bulletin*, 20(2):201–216, 1990.

P. Schmidt. An improved version of the Quandt-Ramsey m.g.f. estimator for mixtures of normal distributions and switching regressions. *Econometrica*, 50:501–516, 1982.

D. W. Scott. Parametric statistical modeling by minimum integrated square error. *Technometrics*, 43:274–285, 2001.

P. K. Sen and J. M. Singer. *Large Sample Methods in Statistics*. Chapman & Hall Inc., 1993.

R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(34):591–611, 1965.

M. Sharifdoust, C. M. Ng, and S. H. Ong. Probability generating function based Jeffrey's divergence for statistical inference. *Communications in Statistics - Simulation and Computation*, 45(7):2445–2458, 2016.

S. Z. Sim and S. H. Ong. Parameter estimation for discrete distributions by generalized Hellinger-type divergence based on probability generating function. *Communications in Statistics - Simulation and Computation*, 39(2):305–314, 2010.

A. Spanos. *Probability Theory and Statistical Inference. Empirical Modeling with Observational Data*. 2nd Ed. Cambridge University Press, 2019.

M. D. Springer. *The Algebra of Random Variables*. Wiley, 1979.

R. G. Staudte and S. J. Sheather. *Robust Estimation and Testing*. Wiley, 1990.

U. Sumita and M. Kijima. Theory and algorithms of the Laguerre transform. Part I: Theory. *Journal of the Operations Research Society of Japan*, 31(4):467–494, 1988.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

C. Tan, X. Shi, and Y. Wu. On nonparametric change point estimator based on empirical characteristic. *Science China Mathematics*, 59(12):2463–2484, 2016.

C. Tenreiro. On the choice of the smoothing parameter for the BHEP goodness-of-fit test. *Computational Statistics and Data Analysis*, 53:1038–1053, 2009.

H. Thaler. *Nonparametric Density Estimation and the Empirical Characteristic Function*. Ph.D. dissertation. State University of New York at Buffalo, Department of Statistics, 1974.

O. Thas. *Comparing Distributions*. Springer, 2010.

J. R. Thompson and R. A. Tapia. *Nonparametric Function Estimation, Modeling, and Simulation*. SIAM, 1990.

D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite of Mixture Distributions*. Wiley, 1985.

K. S. Trivedi. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. Wiley, 2016.

N. G. Ushakov. *Selected Topics in Characteristic Functions*. De Gruyter, 1999.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.

J. M. Van Zyl. Regression with an infinite number of observations applied to estimating the parameters of the stable distribution using the empirical characteristic function. *Communications in Statistics - Theory and Methods*, 45(11):3323–3331, 2016.

D. V. Widder. *The Laplace Transform*. 5th Ed. Princeton University Press, 1959.

D. Xu and J. Knight. Continuous empirical characteristic function estimation of mixtures of normal parameters. *Econometric Reviews*, 30(1):25–50, 2011.

Q. Yao and B. J. T. Morgan. Empirical transform estimation for indexed stochastic models. *Journal of the Royal Statistical Society. Series B*, 61(1):127–141, 1999.

Z. S. Ye and N. Chen. Closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician*, 71(2):195–199, 2017.

A. A. Zghoul. A goodness of fit test for normality based on the empirical moment generating function. *Communications in Statistics - Simulation and Computation*, 39 (6):1292–1304, 2010.

V. M. Zolotarev. Integral transformations of distributions and estimates of parameters of multidimensional spherically symmetric stable laws. In *Contributions to Probability. A Collection of Papers Dedicated to Eugene Lukacs*, pages 283–305. Ed. J. Gani, V. K. Rohatgi. Academic Press, 1981.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub