A Mathematical Model for Grouped Extreme Values with an Application in Automotive Engineering

Sebastian Seifen

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern
zur Erlangung des akademischen Grades Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

Gutachter

1. Gutachter: Prof. Dr. Jürgen Franke (TU Kaiserslautern)
2. Gutachter: Prof. Dr. Wolfgang Schmid (EUV Frankfurt (Oder))

Ort und Datum der Disputation

Kaiserslautern, 22. Oktober 2013

– D 386 –

Zusammenfassung

In der vorliegenden Dissertation wird ein mathematisches Modell für die statistische Analyse gruppierter Extremwerte präsentiert. Als gruppiert bezeichnet man Datensätze, die nicht den exakten Ausgang eines Experimentes angeben, sondern lediglich die Häufigkeiten widergeben, mit denen festgelegte Wertebereiche oder Intervalle die Ergebnisse von unabhängigen Wiederholungen desselben Experimentes beinhalten. Insbesondere zeichnet das der Arbeit zugrunde liegende Experiment extreme Werte auf. Datensätze aus verschiedenen Quellen basieren dabei auf unterschiedlichen Beobachtungszeiträumen.

Ausgehend von der Extremwerttheorie und der üblichen Behandlung von Zähldaten werden parametrische Modelle für die Anzahl an Ereignissen pro Zeiteinheit und Wertebereich entwickelt. Ein Hypothesentest wird präsentiert, der die Annahme überprüft, diese Anzahl folge einer Poisson-Verteilung, und dabei mit den unterschiedlichen Beobachtungszeiträumen umgehen kann. Die Überprüfung der Genauigkeit und der Güte dieses Tests ist Teil der Arbeit.

Die Modellparameter werden mit Hilfe der Maximum-Likelihood-Methode geschätzt. Konsistenz und (asymptotische) Effizienz der Maximum-Likelihood-Schätzer werden (in Teilen) analytisch und per Monte-Carlo-Simulation verifiziert. Die asymptotische Effizienz wird zur Berechnung von Konfidenzintervallen herangezogen. Es wird gezeigt, dass sich die Einteilung der Beobachtungsklassen optimieren lässt, und dass sich ein erheblicher Informationszuwachs erzielen lässt, wenn zumindest der absolute Maximalwert, der sich theoretisch aus dem Experiment ergibt, im Detail bekannt ist.

Das entwickelte Modell wird beispielhaft auf reale Daten aus der Automobilindustrie angewendet.

Abstract

The present doctoral thesis presents a mathematical model for analyzing grouped data based on extreme values. Grouped data means that the exact outcome of the corresponding experiment is not known in detail, but only the occurrence frequency of the outcomes within a particular range or interval is given. In particular, the underlying experiment yields extreme values. In addition, the independent realizations of this experiment are all based on different observation periods.

By dint of extreme value theory and the theory concerning count data, parametric models with regard to the number of events per time unit and domain are developed. A hypothesis test is presented that checks out if this number of events may be Poisson distributed, which cannot be done by standard methods due to the different observation periods. The verification of accuracy and power of this test is part of the thesis.

The model parameters are estimated via maximum likelihood method. It is verified (in part) analytically and by means of Monte Carlo simulations that the maximum likelihood estimators are consistent and (asymptotically) efficient. Based on the asymptotic efficiency, confidence intervals are calculated. It is shown that the partitioning of the observation range can be optimized, and that a huge increase of information can be reached if the absolute maximum value from the experiment is known in detail.

The developed model applies to real data from automotive industry.

Acknowledgement

Foremost, I would like to thank Jürgen Franke for his readiness to supervise me, for his advices and his support during the last years.

Wolfgang Schmid invited me to the *Europa-Universität Viadrina*[1] *EUV* in Frankfurt (Oder) to present my work. I want to thank him for the warm welcome and his readiness to review this thesis.

Many thanks to Klaus Dreßler and the department *Mathematische Methoden in Dynamik und Festigkeit*[2] *MDF* of the *Fraunhofer-Institut für Techno- und Wirtschaftsmathematik*[3] *ITWM* in Kaiserslautern for giving me the opportunity to do a doctorate. Without the organization as well as the technical and financial support, this thesis would not have been possible.

Special thank you within the department *MDF* goes to my advisor Nikolaus Ruf. He was always ready and willing to discuss technical problems and to read my writings. His advices were very important.

Tabea Grebe and Alexander Lemken took over the exhausting task of proof-reading the present thesis. I am very grateful to them for doing so and for any other support during the whole time.

Finally, I want to express my gratitude to the *BMW*[4] *Group*, notably to Mirco Hauke, Kurt Pötter and Christoph Matz. The *BMW Group* invited me to Munich for several months. During this time, I got the chance to get to know the company, to learn a lot about automotive engineering, and to undertake application-oriented research.

I cannot adequately express my gratitude to my parents, because they always supported me unconditionally on my way.

[1] Viadrina European University
[2] Mathematical Methods in Dynamics and Durability
[3] Fraunhofer Institute for Industrial Mathematics
[4] *Bayerische Motoren Werke* (Bavarian Motor Works)

Contents

List of Tables

List of Figures

Table of Abbreviations and Symbols

BMW	Bayerische Motoren Werke
SOLE	supra operating load event
a. s.	almost sure(ly)
$\xrightarrow{d}$	convergence in distribution
$\xrightarrow{P}$	convergence in probability
$\sim$	distributed as
$\overset{a}{\sim}$	asymptotically distributed as
$\forall$	for all
$\forall_{\text{a.s.}}$	almost surely for all
$\hat{\beta}_m$	maximum likelihood estimator of scale β of F_{sev}, page 99
$\text{Bin}(r, q)$	binomial distribution with r trials and success probability q, page 10
$\mathfrak{B}$	Borel σ-algebra on $\mathbb{R}$
$\mathfrak{B}_{>x}$	Borel σ-algebra on $\mathbb{R}_{>x}$
$\mathbb{CV}[X]$	coefficient of variation of X with regard to measure $\mathbb{P}$, page 7
$\mathbb{C}\text{ov}[X, Y]$	covariance of X and Y with regard to measure $\mathbb{P}$, page 7
$\mathbb{D}[X]$	index of dispersion of X with regard to measure $\mathbb{P}$, page 7
e	Euler's number, $\text{e} = 2.71828\ldots$
$\mathcal{E}_{\text{C}}$	counting model, page 57
$\mathcal{E}_{\text{CM}}$	counting-maximum model, page 59
$\mathbb{E}[X]$	expectation of X with regard to measure $\mathbb{P}$, page 7
$\mathbb{E}_{\vartheta}[X]$	expectation of X with regard to measure $\mathbb{P}_{\vartheta}$
exp	exponential function
F_{num}	cumulative distribution function of N_{num}, page 20
F_{sev}	cumulative distribution function of S_{sev}, page 20
$\Gamma(a, b)$	gamma distribution with shape a and scale b, page 13
Γ	gamma function, page 11
$\text{GEV}(\xi)$	generalized extreme value distribution with shape ξ, page 12
$\text{GPar}(\xi, \beta)$	generalized Pareto distribution with shape ξ and scale β, page 12
$\text{GPoi}(\theta, \lambda)$	generalized Poisson distribution with parameters θ and λ, page 11
G_{num}	probability-generating function of N_{num}, page 20
$G_{l,A}$	probability-generating function of $Z_{l,A}$, page 21
$G_{l,A_1\ldots A_d}$	probability-generating function of $(Z_{l,A_1}, \ldots, Z_{l,A_d})$, page 21
$\mathcal{G}_{\text{num}}$	domain of G_{num}, page 20
$\mathcal{G}_{l,A}$	domain of $G_{l,A}$, page 21
$\mathcal{G}_{l,A_1\ldots A_d}$	domain of $G_{l,A_1\ldots A_d}$, page 21
I_{C}	Fisher information of $\mathcal{E}_{\text{C}}$, page 61

I_{num}	Fisher information of $\mathcal{E}_{\text{C}}$ and $\mathcal{E}_{\text{CM}}$ concerning the number of SOLEs, page 62
I_{sev}	Fisher information of $\mathcal{E}_{\text{C}}$ concerning the severity of a SOLE, page 62
$\mathcal{I}_{\text{sev}}$	observed Fisher information of $\mathcal{E}_{\text{CM}}$ concerning the severity of a SOLE, page 114
$\kappa_n[X]$	nth-order cumulant of X with regard to measure $\mathbb{P}$, page 13
Λ_{opt}	optimal class length, page 107
$\mathbb{L}_{\text{C}}$	likelihood function of $\mathcal{E}_{\text{C}}$, page 57
$\mathbb{L}_{\text{CM}}$	likelihood function of $\mathcal{E}_{\text{CM}}$, page 59
log	natural logarithm, logarithm to the base $\mathrm{e} = 2.71828\ldots$
$\text{Log}(r, q)$	logarithmic distribution with parameter q, page 12
ℓ_{C}	log-likelihood function of $\mathcal{E}_{\text{C}}$, page 57
ℓ_{CM}	log-likelihood function of $\mathcal{E}_{\text{CM}}$, page 59
ℓ_{num}	log-likelihood function of $\mathcal{E}_{\text{C}}$ and $\mathcal{E}_{\text{CM}}$ concerning the number of SOLEs, page 60
$\ell_{\text{sev}}^{\text{C}}$	log-likelihood function of $\mathcal{E}_{\text{C}}$ concerning the severity of a SOLE, page 60
$\ell_{\text{sev}}^{\text{CM}}$	log-likelihood function of $\mathcal{E}_{\text{CM}}$ concerning the severity of a SOLE, page 60
M_{sev}	maximum SOLE during one kilometer, page 32
M_{sev}^{*l}	maximum SOLE during l kilometers, page 32
$M_{\text{v}(j)}$	maximum SOLE of vehicle j, page 56
$\hat{\mu}_m$	maximum likelihood estimator of mean μ of F_{num}, page 72
$\mathbb{N}$	set of natural numbers, $\{1, 2, 3, \ldots\}$
$\mathbb{N}_0$	set of natural numbers including zero, $\{0, 1, 2, 3, \ldots\}$
$\mathbb{N}_{\leq x}$	set of natural numbers not greater than x, $\{1, 2, 3, \ldots, x\}$
$\mathbb{N}_{\geq x}$	set of natural numbers not smaller than x, $\{x, x+1, x+2, \ldots\}$
$\text{NBin}(\varrho, \mu)$	negative binomial distribution with exponent ϱ and mean μ, page 11
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2, page 13
N_{num}	number of SOLEs during one kilometer, page 20
N_{num}^{*l}	number of SOLEs during l kilometers, page 21
$N_{\text{v}(j)}$	number of SOLEs of vehicle j, page 56
ψ	digamma function, page 69
π	number pi, $\pi = 3.14159\ldots$
$\text{Poi}(\lambda)$	Poisson distribution with mean λ, page 11
$\mathfrak{P}$	power set of $\mathbb{N}$
$\mathfrak{P}_0$	power set of $\mathbb{N}_0$
p_A	probability that a SOLE lies within A, page 20
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers, $(-\infty, \infty)$
$\mathbb{R}_{<x}$	set of real numbers smaller than x, $(-\infty, x)$
$\mathbb{R}_{\leq x}$	set of real numbers not greater than x, $(-\infty, x]$
$\mathbb{R}_{>x}$	set of real numbers greater than x, (x, ∞)

$\mathbb{R}_{\geq x}$	set of real numbers not smaller than x, $[x, \infty)$
$\hat{\varrho}_m$	maximum likelihood estimator of exponent ϱ of F_{num}, page 81
$\mathfrak{S}$	severity σ-algebra, σ-algebra on $\mathcal{S}$, $\mathfrak{B}_{>u_{\text{sev}}}$, page 20
$\mathcal{S}$	severity space, codomain of S_{sev}, $\mathbb{R}_{>u_{\text{sev}}}$, page 20
S_{sev}	supra operating load event, page 19
Θ	parameter space of $\mathcal{E}_{\text{C}}$ and $\mathcal{E}_{\text{CM}}$, page 57
Θ_{num}	parameter space of $\mathcal{E}_{\text{C}}$ and $\mathcal{E}_{\text{CM}}$ concerning the number of SOLEs, page 57
Θ_{sev}	parameter space of $\mathcal{E}_{\text{C}}$ and $\mathcal{E}_{\text{CM}}$ concerning the severity of a SOLE, page 57
u_{sev}	severity threshold, page 20
$\mathbb{V}\text{ar}[X]$	variance of X with regard to measure $\mathbb{P}$, page 7
$\mathbb{V}\text{ar}_{\vartheta}[X]$	variance of X with regard to measure $\mathbb{P}_{\vartheta}$
$\hat{\xi}_m$	maximum likelihood estimator of shape ξ of F_{sev}, page 99
$Z_{l,A}$	number of SOLEs in A during l kilometers, page 21
$\boldsymbol{Z}_{\text{v}(j)}$	vector of number of SOLEs in all classes of vehicle j, page 56
$Z_{\text{v}(j)_k}$	number of SOLEs in class k of vehicle j, page 56

1. Preface

The present thesis is the result of my work at the *Fraunhofer-Institut für Techno- und Wirtschaftsmathematik*[1] *ITWM* in Kaiserslautern, Rhineland-Palatinate, Germany, which cooperates with the *Technische Universität Kaiserslautern*[2]. The initial motivation for this work came from a collaboration with the *BMW*[3] *Group*. The *BMW Group* is an internationally known German automobile, motorcycle and engine manufacturing company headquartered in Munich, Bavaria, Germany. With BMW, MINI and Rolls-Royce, the *BMW Group* owns three premium brands in the automotive industry.

The task was, generally spoken, to extrapolate from a little image of reality to the big picture. This is a common task for a mathematician and, particularly, a statistician. In the present case, the "little image" were observations from a complex measurement campaign. Due to technical and organizational restrictions, the data were censored, grouped and strongly compacted. Following the quotation *"Por una pequeña muestra podemos juzgar la pieza entera*[4]*"* from Miguel de Cervantes [CS05], these data should be used to learn as much as possible about the powers and forces that conceal behind the observations. However, full knowledge about the "whole piece" cannot be achieved by the observations alone. First of all, a theoretical model has to be established that wants to describe conceptionally the hidden powers until the desired detail is achieved. The data are secondary, even though they can give an idea of how to construct such a model. When an adequate model has been found, the data are used mainly to adjust the model to the reality — or to refute the model.

This thesis is structured exactly in accordance with this procedure. Chapter 2 describes in detail the BMW study on the basis which yields the motivation for this thesis. The example of automotive engineering is used explicitly, but it should be noted that there are many other sectors and situations in which the results of this thesis may be applied. Some of these results can be considered separately from the described experiment since they yield solutions to an abstract issue, e. g. the Poisson hypothesis test established in Section 3.5. After motivation and problem statement, Chapter 2 lists some basic concepts and terms from mathematical statistics and theory of probability which are used throughout the thesis, e. g. estimators, cumulants, maximum likelihood method, distributions and a short overview of extreme value theory.

[1] Fraunhofer Institute for Industrial Mathematics

[2] Technical University Kaiserslautern

[3] *Bayerische Motoren Werke* (Bavarian Motor Works)

[4] By a small sample we may judge the whole piece

Chapter 3 presents the theoretical model which describes the real situation pictured in Chapter 2 mathematically. It lists the assumptions and conditions on which the model is based. In addition, the chapter suggests some parametric approaches and gives criteria for selecting the adequate one. In some parts, the criteria are generalizations of known mathematical concepts (see Section 3.5). In other parts, some known concepts are adapted to the present situation (see Section 3.6).

Since the model is based on a parametric approach, an estimation of these model parameters becomes necessary. At this point, the available data come into play. Chapter 4 prepares the framework for estimating the parameters. It defines the appropriate statistical experiment(s) and specifies some properties like Fisher information and likelihood functions. Thereafter, for all necessary parameters the maximum likelihood estimators are calculated, and the conditions are shown under which these estimators exist.

Chapter 5 studies in detail the found maximum likelihood estimators. Monte Carlo simulations show the behavior and some characteristics of these estimators, e. g. asymptotic efficiency. Moreover, algorithms are presented which can be used in the numerical calculation of the estimators. The accuracy of the Poisson hypothesis test from Section 3.5 under the null hypothesis and the power of this test under several alternative hypotheses is also studied by dint of Monte Carlo simulations. Finally, some real data from the BMW study are used to show how to adjust the model by means of real observations. The presented procedure can be used as standard evaluation when analyzing data which matches form and structure of the BMW data.

The last regular chapter, Chapter 6, summarizes the results and gives a final overview of the concepts of this thesis.

The appendix consists of three parts: Appendix A includes a few technical lemmata. The results of these lemmata are used in some proofs, but they are not relevant furthermore. Appendix B lists all the results from the Monte Carlo studies which are made in connection with the accuracy of the hypothesis test and the maximum likelihood estimators in Chapter 5. And Appendix C shows some plots which illustrate the results of these Monte Carlo studies. Note, that both the tables in Appendix B and the figures in Appendix C refer in their captions to the sections to which they belong. The explanations, descriptions and analyses of the tables and figures can be found in these sections.

2. Motivation, Problem Statement and Methods

This chapter presents a full description of the problem statement that provides the point of departure for this thesis. Section 2.1 introduces supra operating load events (SOLE) and explains their role in automotive development. Section 2.2 describes a measurement campaign initialized by the *BMW Group* the results of which are analyzed in this thesis. The goals of this analysis are explained in Section 2.3. Finally, Section 2.4 provides basics of mathematical statistics and theory of probability which are used throughout this thesis to achieve the goals formulated in Section 2.3.

2.1. SOLE – Supra Operating Load Event

Whenever automotive engineers design and construct a new motorcar they must guarantee a certain durability for all of the vehicle's components. At the same time it is necessary to avoid overdesigning the components, because oversizing would result in increased vehicle weight and higher production costs.

To determine the required strength of a component, according to Zeichfüßl *et al.* [ZGKW08] all loads of a specific type which could act on this component during the vehicle's lifetime are classified into three categories first: operating loads, special event loads and misuse loads. Operating loads are defined as the load level that occur in the vehicle's day-to-day use. These loads must be borne in accordance with the required life of the vehicle. Special events are rare customer-relevant single events. Similar to operating loads they are assigned to the intended use of the vehicle. Special event loads might be rather high, but they must neither reduce the service life nor effect any degradation of performance. Finally, misuse loads are accompanied by impairment, but they must not constitute a security risk to the customer. Therefore, misuse loads require a damage tolerant design.

As an example, consider the load quantity temperature, which is measured on the brake disk. Under normal driving conditions the brake disk heats up and cools down in a characteristic way. These are operating temperatures. A special event could be an emergency braking as a result of an abruptly appearing barrier. Due to the hard braking, the brake disk is heated excessively. When the driver applies the brake and full throttle simultaneously for a longer period of time, he generates an misuse load. This misuse may result in defective brake disks.

Other examples for load quantities are acceleration and the associated force. Here, a special event could be crossing a speed bump or driving through a pothole. In this context, misuse would be passing over a high pavement edge with elevated speed.

According to Zeichfüßl *et al.* [ZGKW08], the boundaries between operating loads, special event loads and misuse loads are not clearly defined and the transition is fluid. Furthermore, the range of possible load magnitudes is not known in every loading case. Environmental factors (e. g. rough roads, mountainous landscape, winding roads, slippery streets, extreme external temperatures), the vehicle parameters (e. g. vehicle mass, level of motorization, set of tires) and the usage patterns of the driver (e. g. stressful and dynamic driving style) determine upper load limits. Not least, the classification as misuse or special event is the driver's subjective decision.

However, both operating loads and special event loads (and, to some extend, misuse loads) are important factors when constructing an automobile. As far as operating loads are concerned, comparatively short measurements under typical driving conditions produce enough data to derive target loads for component testing. On the other hand, very little is known about frequency, severity and other attributes of special events and misuse. Therefore, the resulting target conditions for special event loads and misuse loads are usually worst-case approximations which can lead to a certain level of overdesign.

In this thesis, a model based on data is presented to analyze special events and misuse. These two extreme load situations shall be grouped under the name **supra**[1] **operating load events**, briefly: **SOLE**. This designation illustrates that SOLEs are events creating loads which exceed operating loads.

The statistical model for analyzing SOLEs is based on data provided by a study by the *BMW Group* as mentioned in the preface (see Chapter 1, Section 2.2).

2.2. Experimental Design

The gap of knowledge of SOLE's characteristics as described in Section 2.1 shall be closed by a measurement campaign initialized by the *BMW Group*. All participating test vehicles are used under customer conditions. Extra on-board sensors record loads during day-to-day use which lay above a predefined threshold. This threshold represents the assumed boundary between operating loads and supra

[1] Latin for *above*

operating loads. In irregular intervals the test vehicles are called back to collect the data.

The on-board data acquisition works as follows: When a maneuver in traffic generates a load that is greater than a particular threshold u ($u \in \mathbb{R}_{>0}$), the associated value of the load is temporarily stored. Due to memory restrictions, the load magnitude and the time series structure of a SOLE cannot be recorded and saved exactly. In fact, the detection range $\mathbb{R}_{>u}$ is partitioned into d intervals ($d \in \mathbb{N}_{\geq 2}$),

$$(u, t_1], \quad (t_1, t_2], \quad \ldots, \quad (t_{d-1}, \infty)$$

with class limits $u = t_0 < t_1 < \ldots < t_{d-1} < t_d = \infty$. An algorithm realizes the range the load magnitude lies within, and the counter of this class increases by one. After that the exact value is deleted. The observation resulting from this experiment is often called **grouped data** in data analysis [UC11].

Besides the classified frequency of SOLEs only the magnitude of the maximum load during the whole recording time is saved with an exact value.

The observation period is specified in mileage, because SOLEs are incidents in traffic that only take place during vehicle motion. The number of kilometers travelled since the last readout is also part of the data.

Thus, the observation per vehicle corresponds to the vector

$$(l, z_1, \ldots, z_d, x),$$

where $l \in \mathbb{N}$ is the mileage of the car measured in integer numbers of kilometers, $z_k \in \mathbb{N}_0$ is the number of SOLEs with loads between the magnitudes t_{k-1} and t_k ($k \in \{1, \ldots, d\}$), and $x \in \mathbb{R}_{>u}$ is the maximum value of load. In particular, the total number of SOLEs during the l kilometers, $n = \sum_{k=1}^{d} z_k$, is contained in the observation.

2.3. Main Goals and Approach

The aim of an automotive engineer is to find a construction for the vehicle with an optimal cost-benefit ratio, i. e. the car must resist a certain number of extreme events, but considering the costs the components should not be overdesigned. To find such an optimal cost-benefit ratio an answer to the main question

> What is the probability of observing z events with loads in the range A during l kilometers?

is needed ($z \in \mathbb{N}_0$, $A \subseteq \mathbb{R}_{>u}$, $l \in \mathbb{N}$). If this answer is found and, provided, the capacity of the components is known, the (theoretical) durability of the vehicle is predictable in probabilistic sense.

To illustrate this, let $p(z, A, l)$ be the probability of observing z events in A during l kilometers. Suppose the design limit of a specific component is the load magnitude a_0 ($a_0 \in \mathbb{R}_{>u}$). The probability that this limit is exceeded during the vehicle's life time l_0 is

$$\sum_{z=1}^{\infty} p(z, \mathbb{R}_{\geq a_0}, l_0) = 1 - p(0, \mathbb{R}_{\geq a_0}, l_0).$$

Conversely, the minimum load magnitude that is not exceeded during the vehicle's life time l_0 with at least probability q ($q \in (0,1)$) is

$$\inf\{a \in \mathbb{R}_{>u} \mid p(0, \mathbb{R}_{\geq a}, l_0) \geq q\}.$$

The expected lifetime can be found with the knowledge of $p(z, A, l)$, too. The highest mileage the vehicle can be used such that the design limit a_0 is not exceeded with at least probability q is

$$\max\{l \in \mathbb{N} \mid p(0, \mathbb{R}_{\geq a_0}, l) \geq q\}.$$

In the same way, many other questions can be derived from the main question above.

The collection of data described in Section 2.2 shall help to learn all those things about SOLEs. The analysis of this data shall yield an answer to the main question above. A first naive attempt at an analysis of the collected data could be studying the number of recorded SOLEs class by class. However, this approach brings several difficulties:

- Distinct vehicles could have different class limits and therefore different classes.
- Data of distinct vehicles cannot be compared directly to each other as the mileage is different.
- Approximately d parameters are needed.
- An extrapolation to classes without detected events is not possible.
- Statements can only be made about the given classes.

The last point is the most interesting one. Suppose that the class limits do not depend on the vehicle number, d parameters are not too much to handle, no class is empty, all mileages are the same (more or less), then the first four points above are eliminated. But the last point still reveals that nothing can be said about the expected number of SOLEs in the ranges $\mathbb{R}_{>t_{d-1}+x}$ ($x \in \mathbb{R}_{>0}$) or $\left[t_1 + \frac{t_2-t_1}{3}, t_2 - \frac{t_2-t_1}{3}\right]$, for example. The predictability is limited to the given classes. As mentioned above, the aim should however be to work out a distribution of the number of SOLEs in an arbitrary range or interval.

Thus, the data must be analysed bottom up. A SOLE must be considered as what it is: an occasion with a certain occurrence rate and an exact severity.

The hidden information about these two characteristics must be gathered from the histogram $(z_1, \dots, z_d)$ of the number of SOLEs per class. In this way it is possible to simulate a SOLE in detail. In Chapter 3 the knowledge of the distributions of occurrence rate and severity is shown to be sufficient to determine the probabilities $p(z, A, l)$ from above, i. e. to answer the main question.

2.4. Methods

In this thesis methods of the modern theory of probability and mathematical statistics are used to achieve the goals mentioned in Section 2.3. Some fundamentals of these theories are listed to guarantee a common comprehensibility with respect to nomenclature and notation of mathematical terms. Furthermore, a short overview of frequently used probabilistic methods is given including fundamentals of extreme value theory.

2.4.1. Characteristics of Random Variables

The following statement introduces some basic concepts from probability theory like expectation, variance and index of dispersion of a random variable (definitions are taken from [Eve02, UC11, Bau02, Ahm94, Als05]).

2.4.1 Definition. Let X, Y be integrable random variables from the probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ to the measurable space $(\mathbb{R}, \mathfrak{B})$. The **expectation** and the **variance** of X as well as the **covariance** of X and Y are

$$\begin{aligned} \mathbb{E}[X] &:= \int_{\mathbb{R}} x \; \mathbb{P}(X \in \mathrm{d}x) && \textit{(expectation)} \\ \mathbb{V}\mathrm{ar}[X] &:= \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 && \textit{(variance)} \\ \mathbb{C}\mathrm{ov}[X, Y] &:= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] && \textit{(covariance)} \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \end{aligned}$$

If $\mathbb{E}[X] \neq 0$, the **index of dispersion** and the **coefficient of variation** of X are

$$\begin{aligned} \mathbb{D}[X] &:= \frac{\mathbb{V}\mathrm{ar}[X]}{\mathbb{E}[X]} && \textit{(index of dispersion)} \\ \mathbb{CV}[X] &:= \frac{\sqrt{\mathbb{V}\mathrm{ar}[X]}}{\mathbb{E}[X]} && \textit{(coefficient of variation)} \end{aligned}$$

If the functions F and G are defined by

$$F\colon \mathbb{R} \to [0,1]\colon\ t \longmapsto \mathbb{P}(X \leq t)\,,$$
$$G\colon \mathcal{G} \to \mathbb{R}\colon\ t \longmapsto \mathbb{E}\Big[t^X\Big]$$

with $[-1,1] \subseteq \mathcal{G} \subseteq \mathbb{R}$, then F is the **cumulative distribution function of X** and G is the **probability-generating function of X**.

2.4.2. Statistical Experiment and Likelihood Function

In the stochastic literature, e. g. [Als06, p. 3] and [Geo07, p. 196], a **statistical experiment** or **statistical model** is defined as a triple $(\mathfrak{X}, \mathcal{A}, (P_\vartheta)_{\vartheta\in\Theta})$ with a non-empty set $\mathfrak{X}$ of possible observations called sample space, a σ-algebra $\mathcal{A}$ on $\mathfrak{X}$, and a family $(P_\vartheta)_{\vartheta\in\Theta}$ of probability measures on $(\mathfrak{X}, \mathcal{A})$ parameterized with the elements of parameter space Θ. If X is a random variable from the measurable space $(\Omega, \mathfrak{A})$ to $(\mathfrak{X}, \mathcal{A})$, and $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ is a family of probability measures on $(\Omega, \mathfrak{A})$ such that for all $\vartheta \in \Theta$ it holds

$$\mathbb{P}_\vartheta(X \in A) = P_\vartheta(A) \qquad \forall A \in \mathcal{A},$$

or in a more common notation

$$\mathbb{P}_\vartheta(X \in \cdot\,) = P_\vartheta,$$

then $(\mathfrak{X}, \mathcal{A}, (P_\vartheta)_{\vartheta\in\Theta})$ is called a **statistical experiment based on observation X**, and it can be written $\big(\mathfrak{X}, \mathcal{A}, (\mathbb{P}_\vartheta(X \in \cdot\,))_{\vartheta\in\Theta}\big)$ instead [Als06, p. 3].

Under these assumptions the **likelihood function of X given observation value x** ($x \in \mathfrak{X}$) is defined by

$$\mathbb{L}(\,\cdot\,; x)\colon \Theta \to [0,1]\colon\ \vartheta \longmapsto \begin{cases} \mathbb{P}_\vartheta(X = x)\,, & \text{if } X \text{ is discrete,} \\ f_\vartheta(x), & \text{if } X \text{ is continuous,} \end{cases}$$

where f_ϑ denotes the probability density function of X under $\mathbb{P}_\vartheta$ [Als06, p. 3]. Furthermore, the **log-likelihood function of X given observation value x** ($x \in \mathfrak{X}$) means the natural logarithm of the likelihood function,

$$\ell(\,\cdot\,; x)\colon \Theta \to \mathbb{R}\colon\ \vartheta \longmapsto \log(\mathbb{L}(\vartheta; x))\,.$$

2.4.3. Fisher Information and Unbiased, Consistent and Efficient Estimators

Consider the statistical experiment $\big(\mathfrak{X}, \mathcal{A}, (\mathbb{P}_\vartheta(X \in \cdot\,))_{\vartheta\in\Theta}\big)$ with differentiable log-likelihood function ℓ. As long as the parameter space Θ is an open subset of

$\mathbb{R}^n$ ($n \in \mathbb{N}$), the term

$$I(\vartheta) := \left(\mathbb{E}_\vartheta\left[\tfrac{\partial}{\partial\vartheta_i}\ell(\vartheta;X)\,\tfrac{\partial}{\partial\vartheta_j}\ell(\vartheta;X)\right]\right)_{1\le i,j\le n}$$

is called **Fisher information of X in $\vartheta = (\vartheta_1, \ldots, \vartheta_n) \in \Theta$** [Als06, p. 60][LC98, p. 115]. The Fisher information is a measure of how well the true parameter value can be estimated. To reveal this, let X be a continuous random variable with probability density function f_ϑ and let be $n = 1$. The term

$$\tfrac{\partial\ell}{\partial\vartheta}(\vartheta;x) = \frac{\frac{\partial}{\partial\vartheta}f_\vartheta(x)}{f_\vartheta(x)}$$

is the relative rate of how strong the density in x changes as function with respect to ϑ. Especially, if ϑ_0 is the true parameter, the lower the value of $\frac{\partial\ell}{\partial\vartheta}(\vartheta_0, x)$, the less $f_{\vartheta_0}(x)$ changes relatively as function with respect to ϑ, and the more plausible an estimated value far away from ϑ_0 becomes. Conversely, if the value of $\frac{\partial\ell}{\partial\vartheta}(\vartheta_0, x)$ is high, only estimated values of ϑ near the true parameter ϑ_0 appear acceptable.

Besides this heuristic derivation there are some practical applications of the Fisher information. For example, for $n = 1$ the **information inequality** [LC98, pp. 120/127] states that under some regularity conditions an estimator $\hat{\vartheta}$ of ϑ satisfies

$$\mathbb{V}\mathrm{ar}_\vartheta\left[\hat{\vartheta}(X)\right] \ge \frac{\frac{\partial}{\partial\vartheta}\,\mathbb{E}_\vartheta\left[\hat{\vartheta}(X)\right]}{I(\vartheta)} \qquad \forall\vartheta \in \Theta.$$

In older works, this inequality is called **Cramér-Rao inequality**[LC98, p. 143]. According to this, the right-hand side of the inequality is called **Cramér-Rao lower bound**.

An **unbiased estimator** $\hat{\vartheta}$ of ϑ, i. e. $\mathbb{E}_\vartheta\left[\hat{\vartheta}(X)\right] = \vartheta$ for all $\vartheta \in \Theta$ [LC98, p. 5], which achieves equality on the information inequality,

$$\mathbb{V}\mathrm{ar}_\vartheta\left[\hat{\vartheta}(X)\right] = \frac{1}{I(\vartheta)} \qquad \forall\vartheta \in \Theta,$$

is denominated an **efficient estimator** [Lin05, p. 77] (a more general definition of efficiency see [Bor99, p. 144]).

Furthermore, let $X = (X_i)_{i\in\mathbb{N}}$ be a random series with statistically independent and identically distributed X_i, let I_1 be the Fisher information of X_1 and let $x = (x_i)_{i\in\mathbb{N}}$ be a realization of X. Then, again under some regularity conditions, any sequence $\left(\hat{\vartheta}_m(x)\right)_{m\in\mathbb{N}}$ of roots of the likelihood equation, i. e. $\frac{\partial\ell}{\partial\vartheta}\left(\hat{\vartheta}_m(x); (x)\right) = 0$ for all $m \in \mathbb{N}$, which is **consistent**, i. e. $\hat{\vartheta}_m(X) \xrightarrow{P} \vartheta$ for $m \to \infty$ [LC98, p. 54], satisfies

$$\sqrt{mI_1(\vartheta)}\,\left(\hat{\vartheta}_m(X) - \vartheta\right) \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{for } m \to \infty$$

(convergences each with respect to $\mathbb{P}_\vartheta$) [LC98, p. 449]. Such an estimator is called an **asymptotically efficient estimator** [LC98, p. 439].

2.4.4. Maximum Likelihood Method

The estimation method today known as **maximum likelihood** in its current form was introduced and worked out by Ronald A. Fisher (see Aldrich [Ald97] and Hald [Hal99] for an overview of the history of maximum likelihood). Maximum likelihood follows a simple strategy: assume that the observation is a typical realization of the underlying experiment, then it is plausible to select the one distribution from a specified distribution family that yields the greatest probability for the observed data.

More precisely, let $\big(\mathfrak{X}, \mathcal{A}, (\mathbb{P}_\vartheta(X \in \cdot\,))_{\vartheta\in\Theta}\big)$ be a statistical experiment based on X. The distribution of the elements of $\mathfrak{X}$ is known up to a parameter $\vartheta \in \Theta$. The ambition is to find the true parameter that characterizes the distribution of X. Given a realization x of X ($x \in \mathfrak{X}$) the maximum likelihood method chooses the parameter(s) $\hat{\vartheta} \in \Theta$ which maximizes the likelihood function given x,

$$\hat{\vartheta}(x) := \arg\max_{\vartheta\in\Theta} \mathbb{L}(\vartheta; x) = \arg\max_{\vartheta\in\Theta} \ell(\vartheta; x).$$

If and only if the maximizer of the likelihood function exists and is unique, then $\hat{\vartheta}(x)$ is called **maximum likelihood estimator of ϑ based on x** [Als06, p. 23].

The maximum likelihood method is so common, because it is very versatile in its application. In many situations the likelihood function has got a unique maximum which is the only requirement for calculating the maximum likelihood estimator. The method also manages censoring and truncation, and the observation does not need to be a realization of identically distributed random variables. In many situations the maximum likelihood estimator is consistent [LC98, p. 445] and (asymptotically) efficient (see Section 2.4.3).

2.4.5. Distributions

Several specific distributions appear in this thesis. To avoid confusion, the following definition specifies the required ones in detail. In addition, it lists some important characteristics of them. The definitions and facts below are taken from [JKK05, Con89, JKB94, KN00].

2.4.2 Definition & Fact. Let X be a random variable on the probability space $(\Omega, \mathfrak{A}, \mathbb{P})$.

1. X is **binomially distributed** with r trials and success probability q ($r \in \mathbb{N}$, $q \in (0,1)$), in short $X \sim \mathrm{Bin}(r, q)$, if and only if its support is $\{0, \ldots, r\}$ and its probability mass function is

$$\mathbb{P}(X = n) = \binom{r}{n} q^n \,(1-q)^{r-n} \qquad \forall n \in \{0, \ldots, r\}.$$

Then, expectation $\mathbb{E}$, variance $\mathbb{V}\mathrm{ar}$ and probability-generating function G of X are

$$\mathbb{E}[X] = rq, \quad \mathbb{V}\mathrm{ar}[X] = rq(1-q) \qquad \text{and} \qquad G(t) = (1 + q(t-1))^r \quad \forall t \in \mathbb{R}.$$

2. X is **Poisson distributed** with mean λ ($\lambda \in \mathbb{R}_{>0}$), in short $X \sim \mathrm{Poi}(\lambda)$, if and only if its support is $\mathbb{N}_0$ and its probability mass function is

$$\mathbb{P}(X = n) = \mathrm{e}^{-\lambda} \, \frac{\lambda^n}{n!} \qquad \forall n \in \mathbb{N}_0.$$

Then, expectation $\mathbb{E}$, variance $\mathbb{V}\mathrm{ar}$ and probability-generating function G of X are

$$\mathbb{E}[X] = \lambda, \quad \mathbb{V}\mathrm{ar}[X] = \lambda \qquad \text{and} \qquad G(t) = \mathrm{e}^{\lambda(t-1)} \quad \forall t \in \mathbb{R}.$$

3. X is **negative binomially distributed** with exponent ϱ and mean μ ($\varrho, \mu \in \mathbb{R}_{>0}$), in short $X \sim \mathrm{NBin}(\varrho, \mu)$, if and only if its support is $\mathbb{N}_0$ and its probability mass function is

$$\mathbb{P}(X = n) = \frac{\Gamma(\varrho + n)}{n! \; \Gamma(\varrho)} \left(\frac{\varrho}{\varrho + \mu}\right)^{\varrho} \left(\frac{\mu}{\varrho + \mu}\right)^{n} \qquad \forall n \in \mathbb{N}_0$$

with gamma function Γ [AS65, p 255]. Then, expectation $\mathbb{E}$, variance $\mathbb{V}\mathrm{ar}$ and the probability-generating function G of X are

$$\mathbb{E}[X] = \mu, \quad \mathbb{V}\mathrm{ar}[X] = \mu \left(1 + \frac{\mu}{\varrho}\right) \qquad \text{and} \qquad G(t) = \left(\frac{\varrho}{\varrho - \mu(t-1)}\right)^{\varrho}$$

for all $t \in \left(-\frac{\varrho+\mu}{\mu}, \frac{\varrho+\mu}{\mu}\right)$.

4. X is **generalized Poisson distributed** with parameters θ and λ ($\theta \in \mathbb{R}_{>0}$, $\lambda \in [0, 1)$), in short $X \sim \mathrm{GPoi}(\theta, \lambda)$, if and only if its support is $\mathbb{N}_0$ and its probability mass function is

$$\mathbb{P}(X = n) = \mathrm{e}^{-\theta - n\lambda} \, \frac{\theta \, (\theta + n\lambda)^{n-1}}{n!} \qquad \forall n \in \mathbb{N}_0.$$

Then, expectation $\mathbb{E}$, variance $\mathbb{V}\mathrm{ar}$ and probability-generating function G of X are

$$\mathbb{E}[X] = \frac{\theta}{1 - \lambda}, \quad \mathbb{V}\mathrm{ar}[X] = \frac{\theta}{(1-\lambda)^3} \qquad \text{and} \qquad G(t) = \mathrm{e}^{-\theta\left(1 + \frac{1}{\lambda} W\left(-\lambda \mathrm{e}^{-\lambda} t\right)\right)}$$

for all $t \in [-1, 1]$, where W denotes the (principle branch, i. e. $W \geq -1$, of the) Lambert W function [CGH$^+$96] defined by the equation $x = W(x) \, \mathrm{e}^{W(x)}$ for all $x \in \mathbb{R}$.

5. X is **logarithmically distributed** with parameter q ($q \in (0,1)$), in short $X \sim \mathrm{Log}(q)$, if and only if its support is $\mathbb{N}$ and its probability mass function is

$$\mathbb{P}(X = n) = \frac{-1}{\log(1-q)}\,\frac{q^n}{n} \qquad \forall n \in \mathbb{N}.$$

Then, expectation $\mathbb{E}$ and variance $\mathbb{V}\mathrm{ar}$ of X are

$$\mathbb{E}[X] = \frac{-1}{\log(1-q)}\,\frac{q}{1-q} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}[X] = -q\,\frac{q + \log(1-q)}{(1-q)^2 \log(1-q)^2},$$

and the probability-generating function of X is

$$G(t) = \frac{\log(1-qt)}{\log(1-q)} \qquad \forall t \in \left(-\tfrac{1}{q}, \tfrac{1}{q}\right).$$

6. X is (one-parameter) **generalized extreme value distributed** with shape ξ ($\xi \in \mathbb{R}$), in short $X \sim \mathrm{GEV}(\xi)$, if and only if its support is the set $\{x \in \mathbb{R} \,|\, 1 + \xi x > 0\}$ and its cumulative distribution function is

$$F(x) = \begin{cases} \mathrm{e}^{-(1+\xi x)^{-\frac{1}{\xi}}}, & \text{if } \xi \neq 0, \\ \mathrm{e}^{-\mathrm{e}^{-x}}, & \text{if } \xi = 0, \end{cases} \qquad \forall x \in \begin{cases} \mathbb{R}_{>-\frac{1}{\xi}}, & \text{if } \xi > 0, \\ \mathbb{R}, & \text{if } \xi = 0, \\ \mathbb{R}_{<-\frac{1}{\xi}}, & \text{if } \xi < 0. \end{cases}$$

Then, expectation $\mathbb{E}$ and variance $\mathbb{V}\mathrm{ar}$ of X are

$$\mathbb{E}[X] = \begin{cases} \infty, & \text{if } \xi \geq 1, \\ \gamma, & \text{if } \xi = 0, \\ \frac{\Gamma(1-\xi)-1}{\xi}, & \text{else}, \end{cases} \quad \text{and} \quad \mathbb{V}\mathrm{ar}[X] = \begin{cases} \infty, & \text{if } \xi \geq \frac{1}{2} \\ \frac{\pi^2}{6}, & \text{if } \xi = 0, \\ \frac{\Gamma(1-2\xi)-\Gamma(1-\xi)^2}{\xi^2}, & \text{else}, \end{cases}$$

with Euler-Mascheroni constant $\gamma = 0.57721\ldots$ [UC11] and gamma function Γ [AS65, p 255].

7. X is (two-parameter) **generalized Pareto distributed** with shape ξ and scale β ($\xi \in \mathbb{R}$, $\beta \in \mathbb{R}_{>0}$), in short $X \sim \mathrm{GPar}(\xi, \beta)$, if and only if its support is $\mathbb{R}_{\geq 0}$ (if $\xi \geq 0$) or $\left[0, \frac{\beta}{|\xi|}\right)$ (if $\xi < 0$) and its cumulative distribution function is

$$F(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta} x\right)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0, \\ 1 - \mathrm{e}^{-\frac{1}{\beta} x}, & \text{if } \xi = 0, \end{cases} \qquad \forall x \in \begin{cases} \mathbb{R}_{\geq 0}, & \text{if } \xi \geq 0, \\ \left[0, \frac{\beta}{|\xi|}\right), & \text{if } \xi < 0. \end{cases}$$

Then, expectation $\mathbb{E}$ and variance $\mathbb{V}\mathrm{ar}$ of X are

$$\mathbb{E}[X] = \begin{cases} \frac{\beta}{1-\xi}, & \text{if } \xi < 1, \\ \infty, & \text{if } \xi \geq 1, \end{cases} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}[X] = \begin{cases} \frac{\beta^2}{(1-\xi)^2(1-2\xi)}, & \text{if } \xi < \frac{1}{2}, \\ \infty, & \text{if } \xi \geq \frac{1}{2}. \end{cases}$$

8. X is **gamma distributed** with shape a and scale b ($a, b \in \mathbb{R}_{>0}$), in short $X \sim \Gamma(a, b)$, if and only if its support is $\mathbb{R}_{\geq 0}$ and its probability density function is

$$f(x) = \frac{x^{a-1}\,\mathrm{e}^{-\frac{1}{b}x}}{b^a\,\Gamma(a)} \qquad \forall x \in \mathbb{R}_{\geq 0}$$

with gamma function Γ [AS65, p 255]. Then, expectation $\mathbb{E}$ and variance $\mathbb{V}\mathrm{ar}$ of X are

$$\mathbb{E}[X] = ab \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}[X] = ab^2.$$

9. X is **normally distributed** with mean μ and variance σ^2 ($\sigma^2 \in \mathbb{R}_{>0}$, $\mu \in \mathbb{R}$), in short $X \sim \mathcal{N}(\mu, \sigma^2)$, if and only if its support is $\mathbb{R}$ and its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in \mathbb{R}.$$

Then, expectation $\mathbb{E}$ and variance $\mathbb{V}\mathrm{ar}$ of X are

$$\mathbb{E}[X] = \mu \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}[X] = \sigma^2.$$

The distribution $\mathcal{N}(0, 1)$ is also called **standard normal distribution**.

2.4.6. Cumulants

The **second characteristic function** of a random variable X,

$$\kappa(t) := \log\left(\mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right]\right) = \sum_{n=1}^{\infty} \kappa_n[X]\,\frac{(\mathrm{i}t)^n}{n!} \qquad \forall t \in \mathbb{R}$$

[Luk70, pp. 26–27], generates the cumulants of X (if they exist),

$$\kappa_n[X] = \frac{\frac{\mathrm{d}^n\kappa}{\mathrm{d}t^n}(0)}{\mathrm{i}^n} \qquad \forall n \in \mathbb{N}.$$

Cramér [Cra62, pp. 186–187] repeats this definition and indicates the first four cumulants:

$$\kappa_1[X] = \mathbb{E}[X]\,, \qquad \kappa_2[X] = \mathbb{V}\mathrm{ar}[X]\,, \qquad \kappa_3[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^3\right],$$
$$\kappa_4[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^4\right] - 3\,\mathbb{V}\mathrm{ar}[X]^2\,.$$

Conversely, moments of X are polynomials in cumulants [Cra62, pp. 186–187],

$$\mathbb{E}[X] = \kappa_1[X]\,, \qquad \mathbb{E}\left[X^2\right] = \kappa_2[X] + \kappa_1[X]^2\,,$$
$$\mathbb{E}\left[X^3\right] = \kappa_3[X] + 3\,\kappa_2[X]\,\kappa_1[X] + \kappa_1[X]^3$$
$$\mathbb{E}\left[X^4\right] = \kappa_4[X] + 4\,\kappa_3[X]\,\kappa_1[X] + 3\,\kappa_2[X]^2 + 6\,\kappa_2[X]\,\kappa_1[X]^2 + \kappa_1[X]^4\,.$$

Hald [Hal00] shows that the nth cumulant is additive and homogeneous of degree n, i.e. for statistically independent random variables $X_1, \dots X_m$ and constants $c_1, \dots, c_m \in \mathbb{R}$ ($m \in \mathbb{N}$) it holds

$$\kappa_n\Big[\textstyle\sum_{j=1}^m c_j X_j\Big] = \sum_{j=1}^m {c_j}^n\, \kappa_n[X_j] \qquad \forall n \in \mathbb{N}.$$

Billinger [Bil69] found a law of total cumulants, which helps to calculate cumulants from conditional cumulants. The general formula of this law yields for the first four cumulants

$$\begin{aligned}
\kappa_1[X] =& \kappa_1\big[\kappa_1[X|Y]\big], \\
\kappa_2[X] =& \kappa_1\big[\kappa_2[X|Y]\big] + \kappa_2\big[\kappa_1[X|Y]\big], \\
\kappa_3[X] =& \kappa_1\big[\kappa_3[X|Y]\big] + \kappa_3\big[\kappa_1[X|Y]\big] + 3\ \mathbb{C}\mathrm{ov}\Big[\kappa_1[X|Y], \kappa_2[X|Y]\Big], \\
\kappa_4[X] =& \kappa_1\big[\kappa_4[X|Y]\big] + \kappa_4\big[\kappa_1[X|Y]\big] + 3\,\kappa_2\big[\kappa_2[X|Y]\big] \\
& + 4\ \mathbb{C}\mathrm{ov}\Big[\kappa_1[X|Y], \kappa_3[X|Y]\Big] + 6\ \mathbb{C}\mathrm{ov}\Big[\kappa_1[X|Y]^2, \kappa_2[X|Y]\Big] \\
& - 12\ \kappa_1\big[\kappa_1[X|Y]\big]\ \mathbb{C}\mathrm{ov}\Big[\kappa_1[X|Y], \kappa_2[X|Y]\Big],
\end{aligned}$$

where the conditional cumulants are defined via conditional moments [Dur10, p. 221 *et seqq.*][Als05, p. 284 *et seqq.*]. If the term $\kappa_1[X|Y]$ is almost surely constant, it follows

$$\kappa_n[X] = \kappa_1\big[\kappa_n[X|Y]\big] + 3\,\kappa_2\big[\kappa_2[X|Y]\big]\,\mathbb{1}_{\{4\}}(n) \qquad \forall n \in \{1, \dots, 4\},$$

because for any constant $c \in \mathbb{R}$ it is $\kappa_n[c] = 0$ for all $n \in \mathbb{N}_{\geq 2}$.

2.4.7. Extreme Value Theory

When dealing with extreme events like floods, accidents, records, etc., the so-called extreme value theory gives suitable instruments for analysis. Stuart Coles states a characterization of this mathematical discipline in the preface of *An introduction to Statistical Modeling of Extreme Values* [Col07]:

> *"Extreme value theory is unique as a statistical discipline in that it develops techniques and models for describing the unusual rather than the usual. As an abstract study of random phenomena, the subject can be traced back to the 20th century. It was not until the 1950's that the methodology was proposed in any serious way for the modeling of genuine physical phenomena. It is no coincidence that early applications of extreme value models were primarily in the field*

> *of civil engineering: engineers had always been required to design their structures so that they would withstand the forces that might reasonably be expected to impact upon them. Extreme value theory provided a framework in which an estimate of anticipated forces could be made using historical data."*

Among other things, extreme value theory examines the (approximate) distribution of the maximum of random variables. To demonstrate this, let $X_1, X_2, \dots$ be statistically independent random variables with common cumulative distribution function F and with finite variance. The sum and the maximum of the first n random variables ($n \in \mathbb{N}$) shall be denoted by

$$S_n := \sum_{i=1}^{n} X_i \qquad \text{and} \qquad M_n := \max_{1 \le i \le n} X_i$$

respectively. Coles [Col07, p. 45] notes that the cumulative distribution function of M_n is given by

$$\mathbb{P}(M_n \le x) = F(x)^n \qquad \forall x \in \mathbb{R}.$$

This term depends on F though, which is unknown in many situations. When dealing with the sum S_n, the well-known Central Limit Theorem [LC98, p. 58] allows to approximate the distribution of S_n through a normal distribution, i. e. there are sequences of constants $(a_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}_{>0}$ and $(b_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}$ (e. g. $a_n = \mathbb{V}\mathrm{ar}[S_n]$ and $b_n = \mathbb{E}[S_n]$) so that

$$\mathbb{P}\left(\tfrac{S_n - b_n}{a_n} \le x\right) \xrightarrow{n \to \infty} F_{\mathcal{N}(0,1)}(x) \qquad \forall x \in \mathbb{R},$$

where $F_{\mathcal{N}(0,1)}$ is the cumulative distribution function of a standard normal distribution.

On the other hand, the Fisher–Tippett Theorem [Col07, p. 46], also known as Fisher–Tippett–Gnedenko Theorem [HF06, p. 6], indicates the following: if there are series of constants $(a_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}_{>0}$ and $(b_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}$ such that

$$\mathbb{P}\left(\tfrac{M_n - b_n}{a_n} \le x\right) \xrightarrow{n \to \infty} H(x)$$

for all continuity points of H, where H is a nondegenerated cumulative distribution function, then H is either part of the Fréchet, Gumbel or Weibull distribution family. These so-called **extreme value distribution families** [Col07, p. 47] are represented by the cumulative distribution functions

$$F_{\mathrm{Fre}(\alpha)} \colon \mathbb{R} \to [0,1] \colon\; x \longmapsto \mathrm{e}^{-x^{-\alpha}}\, \mathbb{1}_{\mathbb{R}_{\ge 0}}(x) \qquad \textit{(Fréchet)}$$

$$F_{\mathrm{Gum}} \colon \mathbb{R} \to [0,1] \colon\; x \longmapsto \mathrm{e}^{-\mathrm{e}^{-x}}\, \mathbb{1}_{\mathbb{R}_{\ge 0}}(x) \qquad \textit{(Gumbel)}$$

$$F_{\mathrm{Wei}(\alpha)} \colon \mathbb{R} \to [0,1] \colon\; x \longmapsto \mathrm{e}^{-(-x)^{\alpha}}\, \mathbb{1}_{\mathbb{R}_{<0}}(x) + \mathbb{1}_{\mathbb{R}_{\ge 0}}(x) \qquad \textit{(Weibull)}$$

where $\alpha \in \mathbb{R}_{>0}$. In this situation one says that F is in the domain of attraction of one of the three distribution families.

The Fréchet, Gumbel and Weibull families can be combined into a single family called **generalized extreme value distribution family**. If $F_{\mathrm{GEV}(\xi)}$ is the cumulative distribution function of the one-parameter generalized extreme value distribution as given in Definition 2.4.2, it holds

$$F_{\mathrm{GEV}(\xi)}(x) = \begin{cases} F_{\mathrm{Fre}(1/\xi)}(1+\xi x), & \text{if } \xi > 0, \\ F_{\mathrm{Gum}}(x), & \text{if } \xi = 0, \\ F_{\mathrm{Wei}(-1/\xi)}(-1-\xi x), & \text{if } \xi < 0, \end{cases} \qquad \forall x \in \mathbb{R}.$$

Summarized, the Fisher–Tippett Theorem read as follows:

2.4.3 Theorem (Fisher–Tippett, [Col07, p. 46]). *Let $X_1, X_2, \ldots$ be a sequence of statistically independent random variables with common cumulative distribution function F. For any $n \in \mathbb{N}$, let M_n be the maximum of the first n random variables, $M_n := \max\{X_i \mid 1 \leq i \leq n\}$. Suppose that there are sequences of constants $(a_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}_{>0}$ and $(b_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ and a non-degenerated cumulative distribution function H such that*

$$\mathbb{P}\left(\tfrac{M_n - b_n}{a_n} \leq x\right) = F(a_n x + b_n)^n \xrightarrow{n\to\infty} H(x)$$

for each continuity point of H. Then there are constants a and b ($a \in \mathbb{R}_{>0}$, $b \in \mathbb{R}$) such that

$$H(ax+b) = F_{\mathrm{GEV}(\xi)}(x) \qquad \forall x \in \mathbb{R},$$

where $F_{\mathrm{GEV}(\xi)}$ is the cumulative distribution function of the extreme value distribution $\mathrm{GEV}(\xi)$.

Coles [Col07, pp. 51–52] as well as de Haan and Ferreira [HF06, pp. 11/34] present some examples of distributions which are in the domain of attraction of the generalized extreme value distribution. Thus, the Cauchy distribution is in the Fréchet domain of attraction, exponential, gamma and normal distribution are in the Gumbel domain of attraction, and beta and uniform distribution are in the Weibull domain of attraction. Furthermore, de Haan and Ferreira [HF06] list a lot of criteria to decide in which domain of attraction a distribution lies.

Here, the Fisher–Tippett Theorem is only a preliminary for a related theorem which is very useful with regard to modeling SOLEs. Pickands [Pic75] was the first to realize the connection between the characteristics of random maxima and the generalized Pareto distribution as mentioned in Definition 2.4.2. According to that, large values above a high threshold are approximately generalized Pareto distributed provided that the appropriate exact distribution is in the generalized extreme value domain of attraction. The following formulation of this Theorem can be found in [Sor04, p. 30], for example. More heuristic versions are written down in [Col07, HF06].

2.4.4 Theorem (Pickands–Balkema–de Haan, [Sor04, p. 30]). *Let X_1, $X_2, \ldots$ be a sequence of statistically independent random variables with common cumulative distribution function F which is continuous at*

$$x_F := \sup\{x \in \mathbb{R} \mid F(x) < 1\}.$$

For any $u \in \mathbb{R}$, define the conditional excess distribution function

$$F_u \colon \mathbb{R} \to [0,1]\colon\ x \longmapsto \mathbb{P}(X_1 \le u + x \mid X_1 > u)\,.$$

Now, the following statements are equivalent:

(i) There is a function $\beta(u) \in \mathbb{R}_{>0}$ such that

$$\lim_{u \nearrow x_F} \sup_{0<x<x_F-u} \left|F_u(x) - F_{\mathrm{GPar}(\xi,\beta(u))}(x)\right| = 0,$$

where $F_{\mathrm{GPar}(\xi,\beta(u))}$ denotes the cumulative distribution function of the (two-parameter) generalized Pareto distribution with shape ξ and scale $\beta(u)$.

(ii) F satisfies the Fisher-Tippett Theorem 2.4.3 with extreme value parameter ξ.

3. A Model for Supra Operating Load Events

This chapter evolves a full model for analyzing supra operating load events (SOLE) based on the available data as introduced in Chapter 2. Section 3.1 defines the necessary mathematical framework. Section 3.2 lists the requirements which are sufficient to answer the main question from Section 2.3 concerning the distribution of SOLEs. The main factors for this are the distributions of both number and severity of SOLEs. Section 3.5 presents several approaches for the distribution of numbers of SOLEs. Moreover, a hypothesis test is created which helps to decide whether the number of SOLEs might be Poisson distributed. Analogously, Section 3.6 introduces a suggestion for the distribution of the severity of SOLEs. Beforehand, Section 3.3 examines the question of whether the number of SOLEs in two or more disjoint ranges are independent, and Section 3.4 answers the main question concerning the distribution of SOLEs including the observed maximum event.

3.1. SOLEs in Mathematical Terminology

In automotive environment, a supra operating load event (SOLE) designates an incident in traffic where extreme loads act on a specific vehicle component (see Section 2.1). Such an event is classified by its severity, i. e. the exact absolute load magnitude. Only occurrences with a value of load larger than a specified threshold earn the prefix *supra*.

Of course, the severity of an arbitrary, randomly observed SOLE is not predictable exactly. Indeed, it is a stochastic phenomenon. This yields the mathematical interpretation of a SOLE as a random variable with real function values above a given threshold.

3.1.1 Definition. A **supra operating load event (SOLE)** S_{sev} is a random variable from a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ to the measurable space $(\mathcal{S}, \mathfrak{S})$,

$$S_{\text{sev}} : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\mathcal{S}, \mathfrak{S}),$$

where the **severity space** $\mathcal{S}$ is the set of all possible severities of a SOLE, and the **severity σ-algebra** $\mathfrak{S}$ shall be the Borel σ-algebra on $\mathcal{S}$,

$$(\mathcal{S}, \mathfrak{S}) := (\mathbb{R}_{>u_{\text{sev}}}, \mathfrak{B}_{>u_{\text{sev}}})$$

with **severity threshold** u_{sev} ($u_{\text{sev}} \in \mathbb{R}_{>0}$). F_{sev} is the cumulative distribution function of S_{sev},

$$F_{\text{sev}} \colon \mathbb{R} \to [0,1] \colon \ t \longmapsto \mathbb{P}(S_{\text{sev}} \leq t)\,.$$

For all $A \in \mathfrak{S}$, p_A denotes the probability that a SOLE lies within A,

$$p_A := \mathbb{P}(S_{\text{sev}} \in A)\,.$$

The nature of SOLEs is not completely determined by their severity. The focus must also be set on the frequency of their occurrence. The absolute maximum load magnitude during an observation period depends not only on the possible severity of any single event, but also on the number of events occurring during the observation period. Of course, this number is a stochastic quantity, too.

Because SOLEs are incidences in traffic that only take place if the vehicle is on the move, the observation period is specified in mileage (see Section 2.2). It thus makes sense to define the number of events during one distance unit, which shall be one kilometer.

3.1.2 Definition. The **number of supra operating load events during one kilometer** N_{num} is a random variable from a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ to the measurable space $(\mathbb{N}_0, \mathfrak{P}_0)$,

$$N_{\text{num}} \colon (\Omega, \mathfrak{A}, \mathbb{P}) \to (\mathbb{N}_0, \mathfrak{P}_0).$$

The cumulative distribution function of N_{num} is denoted by F_{num},

$$F_{\text{num}} \colon \mathbb{R} \to [0,1] \colon \ t \longmapsto \mathbb{P}(N_{\text{num}} \leq t)\,.$$

G_{num} is the probability-generating function of N_{num},

$$G_{\text{num}} \colon \mathcal{G}_{\text{num}} \to \mathbb{R} \colon \ t \longmapsto \mathbb{E}\Big[t^{N_{\text{num}}}\Big] = \sum_{n=0}^{\infty} t^n \, \mathbb{P}(N_{\text{num}} = n)$$

with domain $\mathcal{G}_{\text{num}}$ ($[-1,1] \subseteq \mathcal{G}_{\text{num}} \subseteq \mathbb{R}$).

The last definition, or rather the nomenclature in the last definition, only is reasonable under the assumption that the occurrence rate of SOLEs does not change over a period of time. This supposition ensures that a distance is indistinguishable from any other distance of the same length with regard to the number of SOLEs. One can argue whether this treatment is realistic or not, because a vehicle with a high mileage may be treated with less care than a brand-new one, which could lead to a higher occurrence rate for SOLEs. But there is no quantified information about such a process, so, a change of rate is excluded.

Similar to the rate, the severity of a SOLE shall be independent of the mileage. Furthermore, by definition, a SOLE is very rare. So, it is plausible that the severity of one SOLE is not influenced by the characteristics of any other. Let us summarize the mentioned assumptions.

3.1.3 Assumption. (A1) The occurrence rate of SOLEs does not depend on the mileage.

(A2) The occurrence rate of SOLEs is not influenced by the number, severity and mileage of previous events.

(A3) The severity of a SOLE does not depend on the mileage.

(A4) The severity of a SOLE is independent of the number, severity and mileage of previous events.

With these assumptions, the number of SOLEs during l kilometers is just the sum of l statistically independent and identically distributed random variables distributed according to N_{num}, because the mileage is measured in integer numbers of kilometers (see Section 2.2).

3.1.4 Definition. Suppose, $N_1, N_2, \ldots$ and $S_1, S_2, \ldots$ are statistically independent random variables with $N_i \sim F_{\mathrm{num}}$ and $S_i \sim F_{\mathrm{sev}}$ for all $i \in \mathbb{N}$. For all $l \in \mathbb{N}$ and $A \in \mathfrak{S}$, the random variables N^{*l}_{num} and $Z_{l,A}$ are defined as

$$N^{*l}_{\mathrm{num}} := \sum_{i=1}^{l} N_i \qquad \text{and} \qquad Z_{l,A} := \sum_{i=1}^{N^{*l}_{\mathrm{num}}} \mathbb{1}_A(S_i)\,.$$

The probability-generating functions of $Z_{l,A}$ and $(Z_{l,A_1}, \ldots, Z_{l,A_d})$ shall be denoted by $G_{l,A}$ and $G_{l,A_1\ldots A_d}$ respectively ($A_1, \ldots, A_d \in \mathfrak{S}$, $d \in \mathbb{N}_{\geq 2}$),

$$G_{l,A} \colon \mathcal{G}_{l,A} \to \mathbb{R}\colon\ t \longmapsto \mathbb{E}\left[t^{Z_{l,A}}\right]$$

$$G_{l,A_1\ldots A_d} \colon \mathcal{G}_{l,A_1\ldots A_d} \to \mathbb{R}^d\colon\ (t_1, \ldots t_d) \longmapsto \mathbb{E}\left[\prod_{k=1}^{d} t_k^{Z_{l,A_k}}\right]$$

with domains $\mathcal{G}_{l,A}$ of $G_{l,A}$ and $\mathcal{G}_{l,A_1\ldots A_d}$ of $G_{l,A_1\ldots A_d}$ ($[-1,1] \subseteq \mathcal{G}_{l,A} \subseteq \mathbb{R}$, $[-1,1]^d \subseteq \mathcal{G}_{l,A_1\ldots A_d} \subseteq \mathbb{R}^d$).

Under assumptions (A1) and (A2) in Assumption 3.1.3, N^{*l}_{num} is the number of SOLEs during l kilometers and, if additionally the assumptions (A3) and (A4) hold, $Z_{l,A}$ is the number of events with severity in A during l kilometers.

3.2. Distribution of Counts per Range and Mileage

By definition, the question of how many events can be observed within a range A during l kilometers can be answered if the distribution of $Z_{l,A}$ is known (see Definition 3.1.4). The following three points are together a sufficient condition for specifying this distribution:

- The approach is based on the assumptions (A1)-(A4) in Assumption 3.1.3.
- The distribution of the total number of SOLEs during one kilometer, F_{num}, is known.
- The distribution of a SOLE, F_{sev}, is known.

If these three statements are true, the probabilities p_A and the distribution of N_{num}^{*l} are known. The next proposition verifies that this knowledge is sufficient for determining the distribution of $Z_{l,A}$. In addition, the proposition indicates the probability-generating function of $Z_{l,A}$, since it plays an important part in several derivations and argumentations below (e. g. Example 3.2.2, Theorem 3.3.2).

3.2.1 Proposition. *Let be $l \in \mathbb{N}$ and $A \in \mathfrak{S}$.*

1. *The distribution of $Z_{l,A}$ is given by*

$$\mathbb{P}(Z_{l,A} = z) = \frac{{p_A}^z}{z!} \sum_{n=0}^{\infty} \left(\frac{(1-p_A)^n}{n!} \; (n+z)! \; \mathbb{P}\Big(N_{\text{num}}^{*l} = n+z\Big) \right)$$

   for all $z \in \mathbb{N}_0$.

2. *If N_{num} is integrable, then expectation and variance of $Z_{l,A}$ exist and*

$$\mathbb{E}[Z_{l,A}] = l\, p_A\, \mathbb{E}[N_{\text{num}}]\,,$$

$$\mathbb{V}\text{ar}[Z_{l,A}] = l\, p_A\, \mathbb{E}[N_{\text{num}}] + l\, {p_A}^2 \left(\mathbb{V}\text{ar}[N_{\text{num}}] - \mathbb{E}[N_{\text{num}}] \right).$$

3. *The probability-generating function of $Z_{l,A}$ is given by*

$$G_{l,A}(t) = G_{\text{num}}(1 + p_A(t-1))^l \qquad \forall t \in \mathcal{G}_{l,A}$$

   with domain $\mathcal{G}_{l,A} = \{t \in \mathbb{R} \mid (1 + p_A(t-1)) \in \mathcal{G}_{\text{num}}\}$.

Proof. **1.:** Suppose, the total number of SOLEs during l kilometers is $N_{\text{num}}^{*l} = n$ ($n \in \mathbb{N}$). Each of the n events lies either in the range A (with probability p_A) or in its complement (with probability $1 - p_A$). Hence, the conditional random variable $Z_{l,A}$ given $N_{\text{num}}^{*l} = n$ is binomially distributed,

$$\mathbb{P}\Big(Z_{l,A} = z \;\Big|\; N_{\text{num}}^{*l} = n\Big) = \frac{n!}{z!\,(n-z)!}\; {p_A}^{z}\,(1-p_A)^{n-z} \qquad \forall z \in \mathbb{N}_{\leq n}.$$

The relation

$$\mathbb{P}(Z_{l,A} = z) = \sum_{n=0}^{\infty} \mathbb{P}\Big(Z_{l,A} = z \,\Big|\, N_{\text{num}}^{*l} = n + z\Big)\; \mathbb{P}\Big(N_{\text{num}}^{*l} = n + z\Big)$$

proves the first statement.

2.: Since N_{num} and, therefore, N_{num}^{*l} are integrable, the evident fact

$$0 \leq Z_{l,A} \leq N_{\text{num}}^{*l} \qquad \mathbb{P}\text{-a.s.}$$

ensures the existence of expectation and variance of $Z_{l,A}$. The values can be calculated easily by using their conditional versions. The statistical independence of the N_i and the S_i in the definition of $Z_{l,A}$ (see Definition 3.1.4) leads to

$$\mathbb{E}\Big[Z_{l,A} \,\Big|\, N_{\text{num}}^{*l}\Big] = \sum_{i=1}^{N_{\text{num}}^{*l}} \mathbb{E}[\mathbb{1}_A(S_i)] = p_A\, N_{\text{num}}^{*l} \qquad \mathbb{P}\text{-a.s.},$$

$$\mathbb{V}\text{ar}\Big[Z_{l,A} \,\Big|\, N_{\text{num}}^{*l}\Big] = \sum_{i=1}^{N_{\text{num}}^{*l}} \mathbb{V}\text{ar}[\mathbb{1}_A(S_i)] = p_A\,(1-p_A)\, N_{\text{num}}^{*l} \qquad \mathbb{P}\text{-a.s.}$$

The law of total expectation [Wei05, pp. 380–383],

$$\mathbb{E}[Z_{l,A}] = \mathbb{E}\Big[\mathbb{E}\Big[Z_{l,A} \,\Big|\, N_{\text{num}}^{*l}\Big]\Big] = p_A\, \mathbb{E}\Big[N_{\text{num}}^{*l}\Big] = l\, p_A\, \mathbb{E}[N_{\text{num}}]\,,$$

and the law of total variance [Wei05, pp. 385–386],

$$\begin{aligned} \mathbb{V}\text{ar}[Z_{l,A}] &= \mathbb{E}\Big[\mathbb{V}\text{ar}\Big[Z_{l,A} \,\Big|\, N_{\text{num}}^{*l}\Big]\Big] + \mathbb{V}\text{ar}\Big[\mathbb{E}\Big[Z_{l,A} \,\Big|\, N_{\text{num}}^{*l}\Big]\Big] \\ &= l\, p_A\,(1-p_A)\, \mathbb{E}[N_{\text{num}}] + l\, {p_A}^2\, \mathbb{V}\text{ar}[N_{\text{num}}] \end{aligned}$$

prove the second part of the proposition.

3.: For all $t \in \mathbb{R}$, the definitions of N_{num}^{*l} and $Z_{l,A}$ ensure

$$\mathbb{E}\Big[t^{Z_{l,A}} \Big|\, N_{\text{num}}^{*l}\Big] = \prod_{i=1}^{N_{\text{num}}^{*l}} \mathbb{E}\Big[t^{\mathbb{1}_A(S_i)}\Big] = (1 - p_A + p_A t)^{N_{\text{num}}^{*l}} = \prod_{i=1}^{l} (1 + p_A(t-1))^{N_i}\,.$$

Whenever $s := (1 + p_A(t-1)) \in \mathcal{G}_{\text{num}}$, the expectation of the term above exists,

$$\mathbb{E}\Big[t^{Z_{l,A}}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[t^{Z_{l,A}}\,\Big|\,N_{\text{num}}^{*l}\Big]\Big] = \prod_{i=1}^{l}\mathbb{E}\Big[s^{N_i}\Big] = G_{\text{num}}(s)^l.$$

□

In some standard cases the distribution of $Z_{l,A}$ is in the same distribution family as the distribution of N_{num}, i.e. only the values of the distribution parameters differ. That happens, for example, if N_{num} is Poisson, binomially or negative binomially distributed (see (a)-(c) in Example 3.2.2 below). However, this behavior is not transferable to the general case (see (d)-(e) in Example 3.2.2 below). The conjecture is that Poisson, binomial and negative binomial distribution are the only ones with these characteristics.

The third statement in Proposition 3.2.1 provides a criterion to verify whether the distributions of N_{num} and $Z_{l,A}$ only differ in their parameter values. All to do is prove if a transformation of the distribution parameters changes the probability-generating function $G_{\text{num}}(t)$ to be the term $G_{\text{num}}(1 + p_A(t-1))^l$.

The following example applies this criterion to the Poisson, binomial and negative binomial distribution. Using the example of the generalized Poisson distribution, a technique is shown how to prove that N_{num} and $Z_{l,A}$ are not in the same distribution family (this fact can be gathered with help of Ambagaspitiya and Balakrishnan [AB94], too). The logarithmic distribution does not satisfy the criterion above, too.

3.2.2 Example. (a) Suppose, N_{num} is binomially distributed with r trials and success probability q ($r \in \mathbb{N}$, $q \in (0,1)$), then the corresponding probability-generating function is (see Definition 2.4.2)

$$G_{\text{num}}(t) = (1 + q(t-1))^r \qquad \forall t \in \mathbb{R}.$$

Hence, Proposition 3.2.1 yields

$$G_{l,A}(t) = (1 + qp_A(t-1))^{rl} \qquad \forall t \in \mathbb{R}.$$

Thus, $G_{l,A}$ is the probability-generating function of a binomial distribution with rl trials and success probability qp_A. Due to the fact that a distribution is clearly defined by its probability-generating function [Als05, p. 222], $Z_{l,A}$ has to be binomially distributed, too, more precisely $Z_{l,A} \sim \text{Bin}(rl, qp_A)$.

(b) Suppose, N_{num} is Poisson distributed with mean λ ($\lambda \in \mathbb{R}_{>0}$), then the corresponding probability-generating function is (see Definition 2.4.2)

$$G_{\text{num}}(t) = e^{\lambda(t-1)} \qquad \forall t \in \mathbb{R}.$$

Hence, Proposition 3.2.1 yields

$$G_{l,A}(t) = \mathrm{e}^{\lambda l p_A (t-1)} \qquad \forall t \in \mathbb{R}.$$

Thus, $Z_{l,A}$ is Poisson distributed, too, more precisely $Z_{l,A} \sim \mathrm{Poi}(\lambda l p_A)$.

(c) Suppose, N_{num} is negative binomially distributed with exponent ϱ and mean μ ($\varrho, \mu \in \mathbb{R}_{>0}$), then the corresponding probability-generating function is (see Definition 2.4.2)

$$G_{\mathrm{num}}(t) = \left(\frac{\varrho}{\varrho - \mu(t-1)}\right)^{\varrho} \qquad \forall t \in \left(-\tfrac{\varrho+\mu}{\mu}, \tfrac{\varrho+\mu}{\mu}\right).$$

Hence, Proposition 3.2.1 yields

$$G_{l,A}(t) = \left(\frac{\varrho}{\varrho - \mu p_A (t-1)}\right)^{\varrho l} = \left(\frac{\varrho l}{\varrho l - \mu l p_A (t-1)}\right)^{\varrho l}$$

for all $t \in \left(-\frac{\varrho+\mu(2-p_A)}{\mu p_A}, \frac{\varrho+\mu p_A}{\mu p_A}\right)$. Thus, $Z_{l,A}$ is negative binomially distributed, too, more precisely $Z_{l,A} \sim \mathrm{NBin}(\varrho l, \mu l p_A)$.

(d) Suppose that N_{num} is generalized Poisson distributed with parameters θ and λ ($\theta \in \mathbb{R}_{>0}$, $\lambda \in (0,1)$), then the corresponding probability-generating function is (see Definition 2.4.2)

$$G_{\mathrm{num}}(t) = \mathrm{e}^{-\theta\left(1+\frac{1}{\lambda}W\left(-\lambda \mathrm{e}^{-\lambda} t\right)\right)} \qquad \forall t \in [-1,1].$$

Hence, Proposition 3.2.1 yields

$$G_{l,A}(t) = \mathrm{e}^{-\theta l\left(1+\frac{1}{\lambda}W\left(-\lambda \mathrm{e}^{-\lambda}(1+p_A(t-1))\right)\right)} \qquad \forall t \in \left[-\tfrac{2-p_A}{p_A}, 1\right].$$

If $Z_{l,A}$ was generalized Poisson distributed, too, then parameters $\theta^* \in \mathbb{R}_{>0}$ and $\lambda^* \in [0,1)$ would exist such that

$$G_{l,A}(t) = \mathrm{e}^{-\theta^*\left(1+\frac{1}{\lambda^*}W\left(-\lambda^* \mathrm{e}^{-\lambda^*} t\right)\right)} \qquad \forall t \in [-1,1].$$

In particular, due to $W(0) = 0$, it would hold

$$\theta^* = -\log(G_{l,A}(0)) = \theta l \left(1 + \frac{1}{\lambda} W\left(-\lambda \mathrm{e}^{-\lambda}(1-p_A)\right)\right) \tag{3.1}$$

On the other hand, if $Z_{l,A}$ really was generalized Poisson distributed with parameters θ^* and λ^* from above, it would be

$$\mathbb{E}[N_{\mathrm{num}}] = \frac{\theta}{1-\lambda}, \qquad \mathbb{E}[Z_{l,A}] = \frac{\theta^*}{1-\lambda^*}$$

and

$$\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}] = \frac{\theta}{(1-\lambda)^3}, \qquad \mathbb{V}\mathrm{ar}[Z_{l,A}] = \frac{\theta^*}{(1-\lambda^*)^3}.$$

These values inserted into the formulas in the second statement of Proposition 3.2.1 would yield

$$\theta^* = \theta l \underbrace{\frac{p_A}{\sqrt{(1-p_A)\,(1-\lambda)^2 + p_A}}}_{=:f(\lambda,p_A)}. \tag{3.2}$$

The two expressions of θ^* in Equation (3.1) and Equation (3.2) above would result in the relation

$$1 + \frac{1}{\lambda} W\Big(-\lambda \mathrm{e}^{-\lambda}\,(1-p_A)\Big) = f(\lambda, p_A),$$

which is equivalent to

$$1 = \frac{1 - f(\lambda, p_A)}{1 - p_A}\ \mathrm{e}^{\lambda f(\lambda, p_A)}.$$

However, it can be shown that the right-hand side of the last equation exceeds 1 for every $\lambda \in (0,1)$ and every $p_A \in (0,1)$. Consequently, $Z_{l,A}$ cannot be generalized Poisson distributed.

(e) Suppose, the random variable $X := N_{\mathrm{num}} + 1$ is logarithmically distributed with parameter q ($q \in (0,1)$), then the corresponding probability-generating function is (see Definition 2.4.2)

$$G_X(t) = \frac{\log(1-qt)}{\log(1-q)} \qquad \forall t \in \left(-\tfrac{1}{q}, \tfrac{1}{q}\right).$$

By definition of a probability-generating function (see Definition 2.4.1), it then must be

$$G_{\mathrm{num}}(t) = \begin{cases} \frac{G_X(t)}{t} = \frac{\log(1-qt)}{t\ \log(1-q)}, & \text{if } t \in \left(-\tfrac{1}{q}, 0\right) \cup \left(0, \tfrac{1}{q}\right), \\ \lim_{t\to 0} \frac{G_X(t)}{t} = \frac{-q}{\log(1-q)}, & \text{if } t = 0. \end{cases}$$

If $Z_{1,A} + 1$ was logarithmically distributed, too, then a parameter $q^* \in (0,1)$ would exist such that the probability-generating function of $Z_{1,A}$ is

$$G_{1,A}(t) = \begin{cases} \frac{\log\left(1-q^* t\right)}{t\ \log(1-q^*)}, & \text{if } t \in \left(-\frac{1}{q^*}, 0\right) \cup \left(0, \frac{1}{q^*}\right), \\ \frac{-q^*}{\log(1-q^*)}, & \text{if } t = 0. \end{cases}$$

In particular, it would hold

$$q^* = 1 - \exp\Big(-\tfrac{q^*}{G_{1,A}(0)}\Big), \tag{3.3}$$

and therefore

$$\mathbb{E}[Z_{1,A}] = \frac{-1}{\log(1-q^*)}\,\frac{q^*}{1-q^*} - 1 = G_{1,A}(0)\left(\exp\left(\tfrac{q^*}{G_{1,A}(0)}\right) - 1\right) - 1$$

(see Definition 2.4.2). This last relation is equivalent to

$$q^* = G_{1,A}(0)\,\log\left(\tfrac{\mathbb{E}[Z_{1,A}]+G_{1,A}(0)+1}{G_{1,A}(0)}\right).$$

Together with Equation (3.3) above it would follow

$$\begin{aligned} 1 &= \frac{q^*}{1-\exp\left(-\tfrac{q^*}{G_{1,A}(0)}\right)} \\ &= G_{1,A}(0)\,\frac{\mathbb{E}[Z_{1,A}]+G_{1,A}(0)+1}{\mathbb{E}[Z_{1,A}]+1}\,\log\left(\tfrac{\mathbb{E}[Z_{1,A}]+G_{1,A}(0)+1}{G_{1,A}(0)}\right). \end{aligned} \tag{3.4}$$

However, Proposition 3.2.1 yields

$$\mathbb{E}[Z_{1,A}] = \frac{-1}{\log(1-q)}\,\frac{qp_A}{1-q} - p_A \qquad \text{and} \qquad G_{1,A}(0) = \frac{\log(1-q+qp_A)}{(1-p_A)\,\log(1-q)},$$

and with that it can be shown that the right-hand side of Equation (3.4) exceeds 1 for every $q \in (0,1)$ and every $p_A \in (0,1)$. Consequently, $Z_{l,A}+1$ cannot be logarithmically distributed.

With regard to the statistical analysis in Chapter 4, let us generalize Proposition 3.2.1. The next theorem provides the common distribution of the number of SOLEs in several regions. It is easy to see that Proposition 3.2.1 is a special case of it.

3.2.3 Theorem. *Let be $l \in \mathbb{N}$ and let $A_1, \dots A_d \in \mathfrak{S}$ be disjoint measurable sets ($d \in \mathbb{N}_{\geq 2}$).*

1. *The common distribution of $(Z_{l,A_1}, \dots, Z_{l,A_d})$ is given by*

$$\begin{aligned} \mathbb{P}(Z_{l,A_1} = z_1, \dots, Z_{l,A_d} = z_d) &= \left(\prod_{k=1}^{d} \frac{{p_{A_k}}^{z_k}}{z_k!}\right) \\ &\cdot \sum_{n=0}^{\infty}\left(\frac{\left(1-\sum_{k=1}^{d} p_{A_k}\right)^n}{n!}\,\left(n+\sum_{k=1}^{d} z_k\right)!\;\mathbb{P}\left(N^{*l}_{\mathrm{num}} = n+\sum_{k=1}^{d} z_k\right)\right) \end{aligned}$$

for all $(z_1, \dots, z_d) \in {\mathbb{N}_0}^d$.

2. If it is $\mathbb{P}\Big(S_{\text{sev}} \in \bigcup_{k=1}^{d} A_k\Big) = 1$*, then*

$$\begin{aligned}&\mathbb{P}(Z_{l,A_1} = z_1, \dots, Z_{l,A_d} = z_d)\\ &\qquad = \left(\prod_{k=1}^{d} {p_{A_k}}^{z_k}\right) \mathbb{P}\Big(N^{*l}_{\text{num}} = \textstyle\sum_{k=1}^{d} z_k\Big) \dfrac{\big(\sum_{k=1}^{d} z_k\big)!}{\prod_{k=1}^{d} z_k!}\end{aligned}$$

for all $(z_1, \dots, z_d) \in {\mathbb{N}_0}^d$.

3. The probability-generating function of $(Z_{l,A_1}, \dots, Z_{l,A_d})$ *is given by*

$$G_{l,A_1\dots A_d}(t_1, \dots, t_d) = G_{\text{num}}\Big(1 + \textstyle\sum_{k=1}^{d} p_{A_k}(t_k - 1)\Big)^l$$

for all $(t_1, \dots, t_d) \in \mathcal{G}_{l,A_1\dots A_d}$ *with domain*

$$\mathcal{G}_{l,A_1\dots A_d} = \left\{ (t_1, \dots, t_d) \in \mathbb{R}^d \;\middle|\; 1 + \sum_{k=1}^{d} p_{A_k}(t_k - 1) \in \mathcal{G}_{\text{num}} \right\}.$$

***Proof*. 1.:** The calculation of the distribution of $(Z_{l,A_1}, \dots, Z_{l,A_d})$ runs similar to the one in Proposition 3.2.1. Again, suppose the number of SOLEs during l kilometers is $N^{*l}_{\text{num}} = n$ ($n \in \mathbb{N}$). Then each event lies either in one of the sets A_k with probability p_{A_k} or in the set

$$A_{d+1} := \mathcal{S} \setminus (A_1 \cup \ldots \cup A_d)$$

with probability $p_{A_{d+1}}$. Hence, the distribution of $(Z_{l,A_1}, \dots, Z_{l,A_{d+1}})$ given $N^{*l}_{\text{num}} = n$ is multinomially distributed,

$$\begin{aligned}&\mathbb{P}\Big(Z_{l,A_1} = z_1, \dots, Z_{l,A_{d+1}} = z_{d+1} \;\Big|\; N^{*l}_{\text{num}} = n\Big)\\ &\qquad = n! \left(\prod_{k=1}^{d+1} \frac{{p_{A_k}}^{z_k}}{z_k!}\right) \mathbb{1}_{\{n\}}\Big(\textstyle\sum_{k=1}^{d+1} z_k\Big)\end{aligned}$$

for all $(z_1, \dots z_{d+1}) \in {\mathbb{N}_0}^{d+1}$. The remark

$$\begin{aligned}&\mathbb{P}(Z_{l,A_1} = z_1, \dots, Z_{l,A_d} = z_d) =\\ &\sum_{z_{d+1}=0}^{\infty} \mathbb{P}\Big(Z_{l,A_1} = z_1, \dots, Z_{l,A_{d+1}} = z_{d+1} \Big| N^{*l}_{\text{num}} = \textstyle\sum_{k=1}^{d+1} z_k\Big) \mathbb{P}\Big(N^{*l}_{\text{num}} = \textstyle\sum_{k=1}^{d+1} z_k\Big)\end{aligned}$$

finishes the proof of the first statement if it is kept in mind that it holds $p_{A_{d+1}} = 1 - \sum_{k=1}^{d} p_{A_k}$.

2.: Here it holds $\sum_{k=1}^{d} p_{A_k} = 1$. Thus, all addends in the first statement of this theorem are equal to 0 except for $n = 0$.

3.: For $(t_1, \ldots, t_d) \in \mathbb{R}^d$, the definitions of N_{num}^{*l} and $Z_{l,A}$ (see Definition 3.1.4) ensure

$$\mathbb{E}\Big[\prod_{k=1}^{d} t_k^{Z_{l,A_k}} \Big| N_{\text{num}}^{*l}\Big] = \prod_{i=1}^{N_{\text{num}}^{*l}} \mathbb{E}\Big[\prod_{k=1}^{d} t_k^{\mathbb{1}_{A_k}(S_i)}\Big]$$
$$= \prod_{i=1}^{l} \prod_{j=1}^{N_i} \left(1 - \sum_{k=1}^{d} p_{A_k} + \sum_{k=1}^{d} p_{A_k} t_k\right).$$

Whenever $s := \left(1 + \sum_{k=1}^{d} p_{A_k}(t_k - 1)\right) \in \mathcal{G}_{\text{num}}$, the expectation of the term above exists:

$$\mathbb{E}\left[\prod_{k=1}^{d} t_k^{Z_{l,A_k}}\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{k=1}^{d} t_k^{Z_{l,A_k}} \middle| N_{\text{num}}^{*l}\right]\right] = \prod_{i=1}^{l} \mathbb{E}\left[s^{N_i}\right] = G_{\text{num}}(s)^l.$$

□

3.3. (In-)dependence of Number of SOLEs in Disjoint Ranges

Whether or not the numbers of SOLEs in two or more disjoint subsets of $\mathcal{S}$ are statistically independent depends on the chosen distribution F_{num} for sure. Theorem 3.3.2 below shows that Z_{l,A_1} and Z_{l,A_2} ($A_1, A_2 \in \mathfrak{S}$, $A_1 \cap A_2 = \emptyset$, $p_{A_1}, p_{A_2} > 0$) are statistically independent if and only if N_{num} is Poisson distributed.

Before verifying this, the next result deals with a weaker condition than statistical independence: uncorrelatedness. Lemma 3.3.1 shows that Z_{l,A_1} and Z_{l,A_2} are uncorrelated if and only if the index of dispersion of N_{num} is equal to 1.

3.3.1 Lemma. *Let be $l \in \mathbb{N}$ and let $A_1, A_2 \in \mathfrak{S}$ be disjoint measurable sets. If N_{num} is integrable, then the covariance of Z_{l,A_1} and Z_{l,A_2} is given by*

$$\mathbb{C}\text{ov}[Z_{l,A_1}, Z_{l,A_2}] = l\, p_{A_1} p_{A_2} \big(\mathbb{V}\text{ar}[N_{\text{num}}] - \mathbb{E}[N_{\text{num}}]\big).$$

Proof. The definition of a probability-generating function [UC11] ensures

$$\lim_{t_1, t_2 \nearrow 1} \frac{\partial^2 G_{l,A_1 A_2}}{\partial t_1 \partial t_2}(t_1, t_2) = \mathbb{E}[Z_{l,A_1}\, Z_{l,A_2}]\,.$$

However, the probability-generating function G_{l,A_1A_2} is given in Theorem 3.2.3,

$$G_{l,A_1A_2}(t_1,t_2) = G_{\text{num}}\big(1 + p_{A_1}(t_1-1) + p_{A_2}(t_2-1)\big)^l.$$

With $c_{t_1,t_2} := 1 + \sum_{k=1}^{2} p_{A_k}(t_k - 1)$, the derivative of G_{l,A_1A_2} with respect to both variables is

$$\begin{aligned}\frac{\partial^2 G_{l,A_1A_2}}{\partial t_1 \partial t_2}(t_1,t_2) &= l(l-1)\; p_{A_1}p_{A_2}\; G_{\text{num}}(c_{t_1,t_2})^{l-2}\; \tfrac{\mathrm{d}G_{\text{num}}}{\mathrm{d}t}(c_{t_1,t_2})^2 \\ &\quad + l\; p_{A_1}p_{A_2}\; G_{\text{num}}(c_{t_1,t_2})^{l-1}\; \tfrac{\mathrm{d}^2G_{\text{num}}}{\mathrm{d}t^2}(c_{t_1,t_2}).\end{aligned}$$

Since well-known properties of probability-generating functions ensure

$$\lim_{t\nearrow 1} G_{\text{num}}(t) = 1, \qquad \lim_{t\nearrow 1} \tfrac{\mathrm{d}G_{\text{num}}}{\mathrm{d}t}(t) = \mathbb{E}[N_{\text{num}}]\,,$$
$$\lim_{t\nearrow 1} \tfrac{\mathrm{d}^2G_{\text{num}}}{\mathrm{d}t^2}(t) = \mathbb{E}\big[{N_{\text{num}}}^2\big] - \mathbb{E}[N_{\text{num}}]$$

[UC11], the sought-after derivative is

$$\begin{aligned}\mathbb{E}[Z_{l,A_1}\, Z_{l,A_2}] &= \lim_{t_1,t_2\nearrow 1} \frac{\partial^2 G_{l,A_1A_2}}{\partial t_1 \partial t_2}(t_1,t_2) \\ &= l\, p_{A_1}p_{A_2}\Big(l\,\mathbb{E}[N_{\text{num}}]^2 + \mathbb{V}\text{ar}[N_{\text{num}}] - \mathbb{E}[N_{\text{num}}]\Big).\end{aligned}$$

The expectations of Z_{l,A_1} and Z_{l,A_2} from Proposition 3.2.1 yield the desired result. □

In other words, the index of dispersion of N_{num} controls the covariance of Z_{l,A_1} and Z_{l,A_2}:

$$\mathbb{C}\text{ov}[Z_{l,A_1}, Z_{l,A_2}] \lesseqqgtr 0 \quad\Leftrightarrow\quad \mathbb{V}\text{ar}[N_{\text{num}}] \lesseqqgtr \mathbb{E}[N_{\text{num}}] \quad\Leftrightarrow\quad \mathbb{D}[N_{\text{num}}] \lesseqqgtr 1.$$

This confirms the intuition with regard to the influence of Z_{l,A_1} on Z_{l,A_2}. Suppose, the variance of N_{num} is larger than its expectation. A high number of SOLEs in A_1 during l kilometers then indicates that the total number of events N_{num}^{*l} is great. Hence, the number of SOLEs in A_2 will be high, too. On the other hand, let the variance of N_{num} be less than its expectation. A small variance means that realizations of N_{num}^{*l} are frequently close together. Therefore, an extreme number of SOLEs in A_1 hints at a small number of SOLEs in A_2.

Since uncorrelatedness is a necessary condition for statistical independence, the index of dispersion of N_{num} must be equal to 1 if Z_{l,A_1} and Z_{l,A_2} are independent. However, similar to the general case, independence is not a consequence of uncorrelatedness. As mentioned, Z_{l,A_1} and Z_{l,A_2} are statistically independent only in the Poisson case.

3.3.2 Theorem. *Let be $l \in \mathbb{N}$ and let $A_1, A_2 \in \mathfrak{S}$ be disjoint measurable sets with $p_{A_1}, p_{A_2} > 0$. Suppose $F_{\mathrm{num}}(0) < 1$. Then the following statements are equivalent:*

(i) The random variables Z_{l,A_1} and Z_{l,A_2} are statistically independent.

(ii) The number of SOLEs per kilometer, N_{num}, is Poisson distributed.

If the equivalent statements above hold, then also the random variables $Z_{l,A_1}, \ldots,$ Z_{l,A_d} are statistically independent if $A_3, \ldots, A_d \in \mathfrak{S}$ with $A_1 \cap \ldots \cap A_d = \emptyset$ ($d \in \mathbb{N}_{\geq 2}$).

Proof. (**i**) $\Rightarrow$ (**ii**): Since the random variables Z_{l,A_1} and Z_{l,A_2} are statistically independent, it holds

$$G_{l,A_1A_2}(t_1,t_2) = \mathbb{E}\Big[t_1^{Z_{l,A_1}}\, t_2^{Z_{l,A_2}}\Big] = \mathbb{E}\Big[t_1^{Z_{l,A_1}}\Big] \cdot \mathbb{E}\Big[t_2^{Z_{l,A_2}}\Big] = G_{l,A_1}(t_1) \cdot G_{l,A_2}(t_2)$$

for all $(t_1, t_2) \in (\mathcal{G}_{l,A_1} \times \mathcal{G}_{l,A_2}) \cap \mathcal{G}_{l,A_1A_2}$. The probability-generating functions are given in Proposition 3.2.1 and Theorem 3.2.3. It follows for the stated (t_1, t_2)

$$G_{\mathrm{num}}\big(1 + \textstyle\sum_{k=1}^{2} p_{A_k}(t_k - 1)\big) = \displaystyle\prod_{k=1}^{2} G_{\mathrm{num}}(1 + p_{A_k}(t_k - 1))\,.$$

In particular, since it is $[-1, 1] \subseteq \mathcal{G}_{\mathrm{num}}$, it follows

$$G_{\mathrm{num}}\,(1 - x_1 - x_2) = G_{\mathrm{num}}(1 - x_1) \cdot G_{\mathrm{num}}(1 - x_2) \qquad \forall (x_1, x_2) \in [0, 1]^2.$$

Consequently, for any rational number $\frac{m}{n} \in \mathbb{Q} \cap (0, 1)$ with $m, n \in \mathbb{N}$ it holds

$$\begin{aligned} G_{\mathrm{num}}\big(\tfrac{m}{n}\big) = G_{\mathrm{num}}\big(1 - \tfrac{n-m}{n}\big) &= G_{\mathrm{num}}\big(1 - \tfrac{n-m-1}{n}\big) \cdot G_{\mathrm{num}}\big(1 - \tfrac{1}{n}\big) \\ &= \ldots = G_{\mathrm{num}}\big(1 - \tfrac{1}{n}\big)^{n-m} \\ &= \big(G_{\mathrm{num}}\big(1 - \tfrac{1}{n}\big)^{n}\big)^{1-\frac{m}{n}} \\ &= \Big(G_{\mathrm{num}}\big(1 - \tfrac{2}{n}\big)^{n-1}\Big)^{1-\frac{m}{n}} \\ &= \ldots = G_{\mathrm{num}}(0)^{1-\frac{m}{n}}\,. \end{aligned}$$

$G_{\mathrm{num}}(0)$ cannot be equal to 0, because as a probability-generating function G_{num} is continuous and $G_{\mathrm{num}}(1) = 1$. Thus, if $G_{\mathrm{num}}(0)$ was equal to 0, it would be

$$1 = G_{\mathrm{num}}(1) = \lim_{n\to\infty} G_{\mathrm{num}}\big(\tfrac{n-1}{n}\big) = \lim_{n\to\infty} G_{\mathrm{num}}(0)^{\frac{1}{n}} = 0 \qquad \lightning\,.$$

Hence, $G_{\mathrm{num}}(0)$ is positive, and the continuity of G_{num} ensures that even for all real $t \in [0, 1]$ the equation

$$G_{\mathrm{num}}(t) = G_{\mathrm{num}}(0)^{1-t} = \mathrm{e}^{-\log(G_{\mathrm{num}}(0))\,(t-1)}$$

holds. But since it is $1 > F_{\mathrm{num}}(0) = G_{\mathrm{num}}(0)$, this is exactly the probability-generating function of a Poisson distribution (see Definition 2.4.2). Due to the fact that a distribution is well-defined by its probability-generating function on $[0, 1]$ [Als05, 222], N_{num} has to be Poisson.

(ii) $\Rightarrow$ **(i)**: Suppose that N_{num} is Poisson distributed with mean λ ($\lambda \in \mathbb{R}_{>0}$). Example 3.2.2 shows that $Z_{l,A}$ is Poisson distributed with mean $\lambda l p_A$. Together with the results of Theorem 3.2.3 it follows

$$\mathbb{P}(Z_{l,A_1} = z_1)\ \mathbb{P}(Z_{l,A_2} = z_2) = \prod_{k=1}^{2} \frac{(\lambda l p_{A_k})^{z_k}}{z_k!}\ \mathrm{e}^{-\lambda l p_{A_k}} = \mathbb{P}(Z_{l,A_1} = z_1, Z_{l,A_2} = z_2)$$

for all $z_1, z_2 \in \mathbb{N}_0$, which is the definition of statistical independence. This result can easily be generalized to more than two variates $Z_{l,A_1}, \ldots, Z_{l,A_d}$. $\square$

3.4. Including the Maximum SOLE

Besides the total number of SOLEs, also the maximum SOLE will be of interest. Since a SOLE is defined as an event with a severity above a threshold u_{sev}, the maximum SOLE just is the maximum of all events above u_{sev}. Like the total number of SOLEs N_{num}, also the maximum SOLE shall be scaled to one kilometer.

3.4.1 Definition. Let $N_1, N_2, \ldots$ and $S_1, S_2, \ldots$ be the same statistically independent random variables as in Definition 3.1.4 ($N_i \sim F_{\mathrm{num}}$, $S_i \sim F_{\mathrm{sev}}$ for all $i \in \mathbb{N}$). Under assumptions (A1)-(A4) in Assumption 3.1.3, the **maximum supra operating load event during one kilometer** M_{sev} and, for all $l \in \mathbb{N}$, the **maximum supra operating load event during** l **kilometers** M_{sev}^{*l} are the random variables defined by

$$M_{\mathrm{sev}} := \begin{cases} \max\left\{S_1, \ldots, S_{N_{\mathrm{num}}}\right\}, & \text{if } N_{\mathrm{num}} > 0, \\ 0, & \text{if } N_{\mathrm{num}} = 0, \end{cases}$$

$$M_{\mathrm{sev}}^{*l} := \begin{cases} \max\left\{S_1, \ldots, S_{N_{\mathrm{num}}^{*l}}\right\}, & \text{if } N_{\mathrm{num}}^{*l} > 0, \\ 0, & \text{if } N_{\mathrm{num}}^{*l} = 0. \end{cases}$$

Remember that assumptions (A1)-(A4) in Assumption 3.1.3 and the knowledge of F_{num} and F_{sev} are sufficient for specifying the distribution of $Z_{l,A}$ (see page 22). This can be transferred to the maximum SOLE. The probability-generating function G_{num} plays an important part again as can be seen in the following proposition.

3.4.2 Proposition. *Let be $l \in \mathbb{N}$. The distribution of M_{sev}^{*l} is given by*

$$\mathbb{P}\Big(M_{\text{sev}}^{*l} \le t\Big) = \begin{cases} \sum_{n=0}^{\infty} F_{\text{sev}}(t)^n \ \mathbb{P}\big(N_{\text{num}}^{*l} = n\big) = G_{\text{num}}(F_{\text{sev}}(t))^l \,, & \text{if } t \in \mathbb{R}_{\ge 0} \\ 0, & \text{if } t \in \mathbb{R}_{<0}. \end{cases}$$

***Proof*.** Since M_{sev}^{*l} is nonnegative, the cumulative distribution function of M_{sev}^{*l} is equal to 0 on $\mathbb{R}_{<0}$. Furthermore, due to $S_{\text{sev}} > u_{\text{sev}}$,

$$\mathbb{P}\Big(M_{\text{sev}}^{*l} \le t\Big) = \mathbb{P}\Big(N_{\text{num}}^{*l} = 0\Big) = \mathbb{P}(N_{\text{num}} = 0)^l = G_{\text{num}}(0)^l \qquad \forall t \in [0, u_{\text{sev}}].$$

Keeping in mind that $F_{\text{sev}}(t) = 0$ for all $t \in [0, u_{\text{sev}}]$, the proof is done for all $t \in \mathbb{R}_{\le u_{\text{sev}}}$.

For all $t \in \mathcal{S}$ $(= \mathbb{R}_{>u_{\text{sev}}})$, the first equation is established by the obvious equivalence

$$M_{\text{sev}}^{*l} \le t \qquad \Leftrightarrow \qquad Z_{l,(t,\infty)} = 0$$

together with the result of Proposition 3.2.1 and the actuality

$$1 - p_{(t,\infty)} = 1 - (1 - F_{\text{sev}}(t)) = F_{\text{sev}}(t).$$

For the proof of the second equation, first note that $N_{\text{num}}^{*l} = Z_{l,\mathcal{S}}$, which yields with the definition of a probability-generating function

$$\sum_{n=0}^{\infty} F_{\text{sev}}(t)^n \ \mathbb{P}\Big(N_{\text{num}}^{*l} = n\Big) = G_{l,\mathcal{S}}(F_{\text{sev}}(t)) \,.$$

Due to $p_{\mathcal{S}} = 1$, Proposition 3.2.1 now states

$$G_{l,\mathcal{S}}(F_{\text{sev}}(t)) = G_{\text{num}}(1 + p_{\mathcal{S}}(F_{\text{sev}}(t) - 1))^l = G_{\text{num}}(F_{\text{sev}}(t))^l \,.$$

□

The following example shows how easy it is to fix the distribution of the maximum SOLE during a given mileage when only the probability-generating function of N_{num} and the cumulative distribution function of S_{sev} are known.

3.4.3 Example. (a) Suppose, N_{num} is binomially distributed with r trials and success probability q ($r \in \mathbb{N}$, $q \in (0,1)$). Then

$$\mathbb{P}\Big(M_{\text{sev}}^{*l} \le t\Big) = \big(1 + q(F_{\text{sev}}(t) - 1)\big)^{rl} \ \mathbb{1}_{\mathbb{R}_{\ge 0}}(t) \qquad \forall t \in \mathbb{R}.$$

(b) Suppose, N_{num} is Poisson distributed with mean λ ($\lambda \in \mathbb{R}_{>0}$). Then

$$\mathbb{P}\Big(M_{\text{sev}}^{*l} \le t\Big) = \mathrm{e}^{\lambda l (F_{\text{sev}}(t) - 1)} \ \mathbb{1}_{\mathbb{R}_{\ge 0}}(t) \qquad \forall t \in \mathbb{R}.$$

(c) Suppose, N_{num} is negative binomially distributed with exponent ϱ and mean μ ($\varrho, \mu \in \mathbb{R}_{>0}$). Then

$$\mathbb{P}\Big(M_{\text{sev}}^{*l} \le t\Big) = \left(\frac{\varrho}{\varrho - \mu(F_{\text{sev}}(t) - 1)}\right)^{\varrho l} \ \mathbb{1}_{\mathbb{R}_{\ge 0}}(t) \qquad \forall t \in \mathbb{R}.$$

More interesting than the single distribution of the maximum SOLE is the simultaneous distribution of maximum SOLE and number of SOLEs per range. Assume, the severity space $\mathcal{S}$ is divided into three disjoint intervals,

$$\mathcal{S} = (u_{\text{sev}}, t_1] \cup (t_1, t_2] \cup (t_2, \infty),$$

where $u_{\text{sev}} < t_1 < t_2 < \infty$. The number of SOLEs per interval during l kilometers ($l \in \mathbb{N}$) shall be

$$Z_{l,(u_{\text{sev}},t_1]} = z_1, \qquad Z_{l,(t_1,t_2]} = z_2, \qquad Z_{l,(t_2,\infty)} = z_3.$$

The maximum SOLE $t_{\max}$ shall be located between t_1 and t_2, $t_{\max} \in (t_1, t_2)$. Consequently, the third interval must be empty, $z_3 = 0$, while at least one SOLE, the maximum one, is observed within the second interval, $z_2 \geq 1$. Actually, the $z_2 - 1$ SOLEs which are located in (t_1, t_2) next to the maximum SOLE lie between t_1 and $t_{\max}$,

$$Z_{l,(u_{\text{sev}},t_1]} = z_1, \qquad Z_{l,(t_1,t_{\max}]} = z_2, \qquad Z_{l,(t_{\max},\infty)} = 0, \qquad M^{*l}_{\text{sev}} = t_{\max}.$$

Generally, one gets the following result.

3.4.4 Theorem. *Let be $l \in \mathbb{N}$ and let $A_1, \ldots, A_d \in \mathfrak{S}$ be disjoint measurable sets ($d \in \mathbb{N}_{\geq 2}$) such that there exist $a \in \mathbb{R}_{>u_{\text{sev}}}$, $b \in \mathbb{R}_{>a}$ and $c \in \{0, \ldots, d\}$ (if $c = d$, then $b \to \infty$) with*

$$\bigcup_{k=1}^{c} A_k \subseteq (u_{\text{sev}}, a] \qquad \text{and} \qquad \bigcup_{k=c+1}^{d} A_k = (b, \infty).$$

Suppose E denotes the event $E := \{Z_{l,A_1} = z_1, \ldots, Z_{l,A_c} = z_c\}$ for an arbitrary $(z_1, \ldots, z_c) \in {\mathbb{N}_0}^c$. Then it holds for all $z \in \mathbb{N}$

$$\mathbb{P}\Big(M^{*l}_{\text{sev}} \leq t,\ E,\ Z_{l,(a,b]} = z,\ Z_{l,A_{c+1}} = 0,\ \ldots,\ Z_{l,A_d} = 0\Big)$$
$$= \begin{cases} \mathbb{P}\big(E,\ Z_{l,(a,b]} = z,\ Z_{l,(b,\infty)} = 0\big)\ \mathbb{1}_{(b,\infty)}(t)\,, & \text{if } t \in \mathbb{R} \setminus (a,b], \\ \mathbb{P}\big(E,\ Z_{l,(a,t]} = z,\ Z_{l,(t,\infty)} = 0\big)\,, & \text{if } t \in (a,b]. \end{cases}$$

In addition, if F_{sev} is absolutely continuous and f_{sev} is (almost everywhere) the derivative of F_{sev}, then the common probability density function of M^{*l}_{sev} and $(Z_{l,A_1}, \ldots Z_{l,A_d})$ is

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\Big(M^{*l}_{\text{sev}} \leq t,\ E,\ Z_{l,(a,b]} = z,\ Z_{l,A_{c+1}} = 0,\ \ldots,\ Z_{l,A_d} = 0\Big)$$
$$= \frac{z\, f_{\text{sev}}(t)}{p_{(a,t]}}\ \mathbb{P}\big(E,\ Z_{l,(a,t]} = z,\ Z_{l,(t,\infty)} = 0\big)\ \mathbb{1}_{(a,b]}(t) \qquad \forall_{\text{a.s.}} t \in \mathbb{R}.$$

Proof. Due to $\{Z_{l,A_c+1} = 0, \ldots, Z_{l,A_d} = 0\} = \{Z_{l,(b,\infty)} = 0\}$, the random variables which are equal to 0 can be combined with each other to one random variable. The first result comes from the relation

$$\left\{M^{*l}_{\mathrm{sev}} \le t,\, Z_{l,(a,b]} = z,\, Z_{l,(b,\infty)} = 0\right\} = \begin{cases} \emptyset, & \text{if } t \le a \\ \left\{Z_{l,(a,t]} = z,\, Z_{l,(t,\infty)} = 0\right\}, & \text{if } t \in (a,b] \\ \left\{Z_{l,(a,b]} = z,\, Z_{l,(b,\infty)} = 0\right\}, & \text{if } t > b. \end{cases}$$

The rest of the theorem follows with Theorem 3.2.3, which states

$$\mathbb{P}\big(E,\, Z_{l,(a,t]} = z,\, Z_{l,(t,\infty)} = 0\big) = \left(\frac{p_{(a,t]}}{p_{(a,b]}}\right)^{z} \mathbb{P}\big(E,\, Z_{l,(a,b]} = z,\, Z_{l,(b,\infty)} = 0\big)\,,$$

and the fact

$$\tfrac{\mathrm{d}}{\mathrm{d}t} p_{(a,t]} = \tfrac{\mathrm{d}}{\mathrm{d}t}\big(F_{\mathrm{sev}}(t) - F_{\mathrm{sev}}(a)\big) = f_{\mathrm{sev}}(t)\,.$$

Thus, the derivative with respect to t is

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\big(E,\, Z_{l,(a,t]} = z,\, Z_{l,(t,\infty)} = 0\big) &= z\, f_{\mathrm{sev}}(t)\ \frac{{p_{(a,t]}}^{z-1}}{{p_{(a,b]}}^{z}}\ \mathbb{P}\big(E,\, Z_{l,(a,b]} = z,\, Z_{l,(b,\infty)} = 0\big) \\ &= \frac{z\, f_{\mathrm{sev}}(t)}{p_{(a,t]}}\ \mathbb{P}\big(E,\, Z_{l,(a,t]} = z,\, Z_{l,(t,\infty)} = 0\big)\,. \end{aligned}$$

□

Especially if it holds $\mathbb{P}\Big(S_{\mathrm{sev}} \in \bigcup_{k=1}^{d} A_k \cup [a,b)\Big) = 1$ in Theorem 3.4.4, the second statement from Theorem 3.2.3 yields together with the result of Theorem 3.4.4

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\Big(M^{*l}_{\mathrm{sev}} \le t,\ E,\ Z_{l,(a,b]} = z,\ Z_{l,A_c+1} = 0, \ldots, Z_{l,A_d} = 0\Big) & \\ = y_1\ f_{\mathrm{sev}}(t)\ \mathbb{P}\big(E,\, Z_{l,(a,t]} = z,\, Z_{l,(t,\infty)} = 0\big)\ \mathbb{1}_{(a,b]}(t) & \\ = y_2\ f_{\mathrm{sev}}(t)\ \mathbb{P}\big(E,\, Z_{l,(a,t]} = z-1,\, Z_{l,(t,\infty)} = 0\big)\ \mathbb{1}_{(a,b]}(t)\,, & \end{aligned} \tag{3.5}$$

where

$$y_1 := \frac{z}{p_{(a,t]}} \qquad \text{and} \qquad y_2 := \frac{\mathbb{P}\big(N^{*l}_{\mathrm{num}} = z + \sum_{k=1}^{c} z_k\big)}{\mathbb{P}\big(N^{*l}_{\mathrm{num}} = z - 1 + \sum_{k=1}^{c} z_k\big)} \left(z + \sum_{k=1}^{c} z_k\right).$$

Hence, the common probability of $(Z_{l,A_1}, \ldots, Z_{l,A_c}, Z_{l,[a,b)}, Z_{l,A_c+1}, \ldots, Z_{l,A_d})$ and M^{*l}_{sev} can be interpreted in two ways (see Equation (3.5) above): either the maximum SOLE itself is counted in the interval $(a, t]$ ($\rightsquigarrow Z_{l,(a,t]} = z$) and a correction factor collects the fact that in truth the location of the maximum SOLE is known in detail ($\rightsquigarrow y_1$), or the maximum SOLE itself is not counted in the interval $(a, t]$ ($\rightsquigarrow Z_{l,(a,t]} = z - 1$) and a correction factor collects the fact that one event is not counted ($\rightsquigarrow y_2$). It is worth mentioning that y_1 depends on F_{sev} and is independent of F_{num} while y_2 only depends on F_{num} and not on F_{sev}. Depending on whether the severity or the number of SOLEs shall be analyzed, the first or the second interpretation is preferred.

3.5. Selecting the Distribution of Number of SOLEs per Kilometer

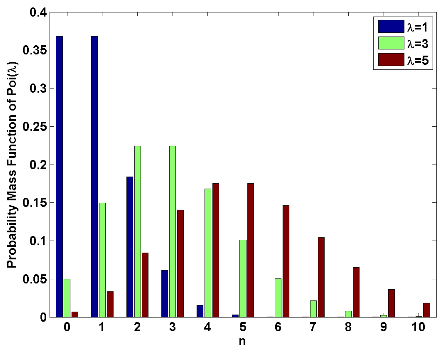
When looking for a convenient distribution for the number of SOLEs per kilometer, F_{num}, the Poisson distribution (see Definition 2.4.2) is an adequate choice. This distribution was first introduced by the French mathematician Siméon D. Poisson published in 1837 in his work *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* (Research on the Probability of Judgments in Criminal and Civil Matters [Poi37]). According to Yang [YHAV07] *"the Poisson model is the usual approach to analysis"* when analyzing counts.

Prem C. Consul gives an outline of the derivation of the Poisson distribution and characterizes it in his introduction to the first chapter of *Generalized Poisson Distributions* [Con89]:

> *"The Poisson probability model has been used in a very wide variety of situations to describe the behavior of living beings as well as the patterns observed in different types of nonliving phenomena. The Poisson distribution is generated by processes in which a large number of cells, squares, leaves, petals, or intervals of time (e. g. seconds, minutes, hours, days) are hit by a relatively small number of events (births, deaths, blood cells, particles of nuclear decay, balls, etc.) such that the occurrence or nonoccurrence of an event in that interval has no effect on the further occurrences or nonoccurrences of events in that interval and that the probability of two or more occurrences in a short interval of time is almost zero; that is, a cell (or interval) with lots of counts is as likely to get another count as a cell with fewer counts or with no counts at all. This principle of randomness implies that the individual organisms or events are scattered by chance alone."*

Due to this characterization, the assumptions (A1)-(A4) in Assumption 3.1.3 make the Poisson distribution a reasonable approach for modeling SOLEs. The only question is whether the concentration of SOLEs really is totally randomly distributed over the space available to them. However, the Poisson approach has some mathematical advantages in SOLE context, e. g. the number of events in an arbitrary subset during any mileage consequently is Poisson, too (see Example 3.2.2), and the Poisson distribution is the unique distribution that makes the number of SOLEs in disjoint ranges statistically independent (see Theorem 3.3.2). Figure 3.1 shows the Poisson distribution for three different values of the mean parameter λ.

Figure 3.1.: *Probability mass functions of Poisson distribution.*



3.5.1. The Index of Dispersion

Even if there are some mathematical advantages, the Poisson approach must be verified. For this attempt, the quotient of variance and expectation, the so-called index of dispersion,

$$\mathbb{D}[N_{\text{num}}] = \frac{\mathbb{V}\text{ar}[N_{\text{num}}]}{\mathbb{E}[N_{\text{num}}]},$$

is a suitable test object, because a Poisson distribution is characterized by an index of dispersion that is equal to 1 (see Definition 2.4.2). Given m independent realizations $n_1, \ldots, n_m \in \mathbb{N}_0$ of N_{num} with $\sum_{j=1}^{m} n_j > 0$ ($m \in \mathbb{N}$), the ratio of sample variance and sample mean,

$$\frac{\frac{1}{m-1} \sum_{j=1}^{m} \left(n_j - \frac{1}{m} \sum_{i=1}^{m} n_i\right)^2}{\frac{1}{m} \sum_{j=1}^{m} n_j},$$

is an evident estimator of $\mathbb{D}[N_{\text{num}}]$, because sample mean and sample variance are unbiased and consistent estimators of the expectation and the variance of a random variable, respectively [LC98, p. 55]. Hence, according to Slutsky's Theorem [Slu25, Cra62, pp. 254–255], the estimator above is consistent, too. A hypothesis test can be constructed from the fact that the term

$$\frac{\sum_{j=1}^{m}\left(N_j - \frac{1}{m}\sum_{i=1}^{m} N_i\right)^2}{\frac{1}{m}\sum_{j=1}^{m} N_j}$$

is well-known to be approximately chi-squared distributed with $(m-1)$ degrees of freedom if $N_1, \ldots, N_m$ are statistically independent and identically Poisson distributed random variables [Ben59, Hoe43, Sel65].

Now suppose, the observation does not consist of realizations of N_{num} but of N_{num}^{*l} with diverse values l (see Section 2.2). Again, let $N_1, N_2, \ldots$ be statistically independent random variables distributed according to N_{num}, and let be $l_1, \ldots, l_{\tilde{m}} \in \mathbb{N}$ $(\tilde{m} \in \mathbb{N})$. Define

$$\tilde{N}_j := \sum_{i=1+\sum_{k=1}^{j-1} l_k}^{\sum_{k=1}^{j} l_k} N_i \qquad \forall j \in \{1, \ldots, \tilde{m}\},$$

then each $\tilde{N}_j$ is as distributed as $N_{\text{num}}^{*l_j}$. Especially, they are not identically distributed as long as the l_j are not all equal to each other. If only realizations of the $\tilde{N}_j$ are available, then the sample mean from above with $m = \sum_{k=1}^{\tilde{m}} l_k$ can be calculated just as well, because

$$\frac{1}{m}\sum_{j=1}^{m} N_j = \frac{1}{\sum_{k=1}^{\tilde{m}} l_k}\sum_{j=1}^{\tilde{m}} \tilde{N}_j.$$

However, the calculation of the sample variance is problematic, because in general it is

$$\sum_{j=1}^{m}\left(N_j - \frac{1}{m}\sum_{i=1}^{m} N_i\right)^2 \neq \sum_{i=1}^{\tilde{m}}\left(\tilde{N}_i - \frac{1}{\sum_{k=1}^{\tilde{m}} l_k}\sum_{j=1}^{\tilde{m}} \tilde{N}_j\right)^2.$$

To fix this problem, the index of dispersion of N_{num} must not be expressed by expectation and variance of the nonobservable random variable N_{num} but by $N_{\text{num}}^{*l_1}, N_{\text{num}}^{*l_2}$, etc. For this purpose, generalize the definition of N_{num}^{*l} in Definition 3.1.4 to the effect that the mileage l is allowed to be a random variable $L\colon (\Omega, \mathfrak{A}, \mathbb{P}) \to (\mathbb{N}, \mathfrak{P})$,

$$l \to L \quad \rightsquigarrow \quad N_{\text{num}}^{*l} \to N_{\text{num}}^{*L} = \sum_{i=1}^{L} N_i.$$

L shall be statistically independent of all the N_i. It can be interpreted as the random variable the mileages $l_1, l_2, \ldots$ are realizations of. It is worth trying to replace the number of SOLEs *per kilometer*, N_{num}, by the number of SOLEs during L kilometers divided by the number of kilometers, N^{*L}_{num}/L. The following lemma is the result of this attempt. In addition, it handles another approach, which becomes interesting in connection to the maximum likelihood estimation in case of a negative binomially distributed N_{num} (see Section 4.3.2, especially Theorem 4.3.7).

3.5.1 Lemma. *Suppose, L is an integrable random variable from a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ to the measurable space $(\mathbb{N}, \mathfrak{P})$, $L\colon (\Omega, \mathfrak{A}, \mathbb{P}) \to (\mathbb{N}, \mathfrak{P})$, which is statistically independent of all other random variables. Then, the index of dispersion of N_{num} can be written as*

$$\mathbb{D}[N_{\text{num}}] = \frac{\mathbb{V}\text{ar}\Big[\frac{N^{*L}_{\text{num}}}{L}\Big]}{\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L^2}\Big]} \qquad \text{and} \qquad \mathbb{D}[N_{\text{num}}] = \frac{\mathbb{E}\Big[\frac{(N^{*L}_{\text{num}})^2}{L}\Big] - \frac{\mathbb{E}[N^{*L}_{\text{num}}]^2}{\mathbb{E}[L]}}{\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L}\Big]}.$$

Proof. Generally, the expectation of N^{*L}_{num}/L^c with $c \in \mathbb{R}$ is

$$\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L^c}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L^c}\,\Big|\, L\Big]\Big] = \mathbb{E}\Big[\tfrac{1}{L^c}\textstyle\sum_{i=1}^{L} \mathbb{E}[\,N_{\text{num}} \,|\, L]\Big] = \mathbb{E}\big[L^{1-c}\big]\,\mathbb{E}[N_{\text{num}}]\,.$$

The variance of N^{*L}_{num}/L is obtained from its conditional variance given L,

$$\mathbb{V}\text{ar}\Big[\frac{N^{*L}_{\text{num}}}{L}\,\Big|\, L\Big] = \frac{1}{L^2}\sum_{i=1}^{L} \mathbb{V}\text{ar}[\,N_{\text{num}} \,|\, L] = \frac{1}{L}\mathbb{V}\text{ar}[N_{\text{num}}] \qquad \mathbb{P}\text{-a.s.},$$

because the law of total variance [Wei05, pp. 385–368] states

$$\mathbb{V}\text{ar}\Big[\frac{N^{*L}_{\text{num}}}{L}\Big] = \mathbb{E}\Big[\mathbb{V}\text{ar}\Big[\frac{N^{*L}_{\text{num}}}{L}\,\Big|\, L\Big]\Big] + \mathbb{V}\text{ar}\Big[\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L}\,\Big|\, L\Big]\Big] = \mathbb{E}\big[\tfrac{1}{L}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}]\,.$$

Finally, the expectation of $\big(N^{*L}_{\text{num}}\big)^2/L$ is

$$\begin{aligned}\mathbb{E}\Big[\frac{(N^{*L}_{\text{num}})^2}{L}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\frac{(N^{*L}_{\text{num}})^2}{L}\,\Big|\, L\Big]\Big] &= \mathbb{E}\Big[L\,\mathbb{V}\text{ar}\Big[\frac{N^{*L}_{\text{num}}}{L}\,\Big|\, L\Big] + L\,\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L}\,\Big|\, L\Big]^2\Big] \\ &= \mathbb{V}\text{ar}[N_{\text{num}}] + \mathbb{E}[L]\,\mathbb{E}[N_{\text{num}}]^2\,.\end{aligned}$$

All these calculations ensure

$$\frac{\mathbb{V}\text{ar}\Big[\frac{N^{*L}_{\text{num}}}{L}\Big]}{\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L^2}\Big]} = \frac{\mathbb{E}\big[\frac{1}{L}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}]}{\mathbb{E}\big[\frac{1}{L}\big]\,\mathbb{E}[N_{\text{num}}]} = \frac{\mathbb{V}\text{ar}[N_{\text{num}}]}{\mathbb{E}[N_{\text{num}}]} = \frac{\mathbb{E}\Big[\frac{(N^{*L}_{\text{num}})^2}{L}\Big] - \frac{\mathbb{E}[N^{*L}_{\text{num}}]^2}{\mathbb{E}[L]}}{\mathbb{E}\Big[\frac{N^{*L}_{\text{num}}}{L}\Big]}.$$

The index of dispersion $\mathbb{D}[N_{\text{num}}]$ is exactly defined as the ratio of variance $\mathbb{V}\text{ar}[N_{\text{num}}]$ and expectation $\mathbb{E}[N_{\text{num}}]$. □

The indices of dispersion of N_{num} and ${}^{N^{*L}_{\mathrm{num}}}/_{L}$ are nearly equal to each other. Only the additional factor ${}^{1}/_{L}$ in the denominator is a necessary correction term. Since the mileage and the total number of SOLEs during this mileage is part of the collected data (see Section 2.2), the index of dispersion can be estimated in the following way: let $(n_1, l_1), \ldots, (n_m, l_m) \in \mathbb{N}_0 \times \mathbb{N}$ $(m \in \mathbb{N})$ be the observable realizations of $\big(N^{*L}_{\mathrm{num}}, L\big)$, then estimate the variance and expectations in Lemma 3.5.1 by sample variance and sample mean as described above (see page 37), so that evident consistent estimators of $\mathbb{D}[N_{\mathrm{num}}]$ are

$$\begin{aligned}\hat{D}_1 = \hat{D}_1\big((n_j, l_j)_{1\le j\le m}\big) &:= \frac{\frac{1}{m-1}\sum_{j=1}^{m}\left(\frac{n_j}{l_j} - \frac{1}{m}\sum_{i=1}^{m}\frac{n_i}{l_i}\right)^2}{\frac{1}{m}\sum_{j=1}^{m}\frac{n_j}{{l_j}^2}} \\ \hat{D}_2 = \hat{D}_2\big((n_j, l_j)_{1\le j\le m}\big) &:= \frac{\frac{1}{m}\sum_{j=1}^{m}\frac{{n_j}^2}{l_j} - \frac{1}{m}\frac{\left(\sum_{j=1}^{m} n_j\right)^2}{\sum_{j=1}^{m} l_j}}{\frac{1}{m}\sum_{j=1}^{m}\frac{n_j}{l_j}}.\end{aligned} \tag{3.6}$$

3.5.2. Confidence Intervals of Sample Index of Dispersion

An estimator of $\mathbb{D}[N_{\mathrm{num}}]$ alone is not a complete instrument to decide whether N_{num} could be Poisson distributed or not. The distribution or, at least, an approximate distribution of this estimator is necessary to get confidence intervals or to perform a hypothesis test. Since the data come from a large measurement campaign, the sample size is very high. Therefore, it is sufficient to state the approximate normal distribution the estimators $\hat{D}_1$ and $\hat{D}_2$ above (see Equation (3.6)) converge to in law.

3.5.2 Theorem. *Let $(L_j)_{j\in\mathbb{N}}$ and $(N_{ij})_{i,j\in\mathbb{N}}$ be statistically independent random variables distributed according to $L_j \sim L$ and $N_{ij} \sim N_{\mathrm{num}}$ for all $i, j \in \mathbb{N}$, where L is the same random variable as in Lemma 3.5.1. For all $j \in \mathbb{N}$ define*

$$\mathcal{N}_j := \sum_{i=1}^{L_j} N_{ij} \qquad \left(\sim N^{*L_j}_{\mathrm{num}}\right).$$

Then, for large values of m, the estimators $\hat{D}_1$ and $\hat{D}_2$ of $\mathbb{D}[N_{\mathrm{num}}]$ are approximately normally distributed,

$$\left.\begin{aligned}\sqrt{m}\left(\hat{D}_1\big((\mathcal{N}_j, L_j)_{1\le j\le m}\big) - \mathbb{D}[N_{\mathrm{num}}]\right) &\xrightarrow{d} \mathcal{N}\big(0, {\sigma_{\mathrm{iod}}}^2\big) \\ \sqrt{m}\left(\hat{D}_2\big((\mathcal{N}_j, L_j)_{1\le j\le m}\big) - \mathbb{D}[N_{\mathrm{num}}]\right) &\xrightarrow{d} \mathcal{N}\big(0, {\tau_{\mathrm{iod}}}^2\big)\end{aligned}\right. \quad \text{for } m \to \infty$$

with

$$\sigma_{\mathrm{iod}}{}^2 := \frac{\mathbb{E}\big[\frac{1}{L^3}\big]}{\mathbb{E}\big[\frac{1}{L}\big]^2}\left(\frac{\kappa_4[N_{\mathrm{num}}]}{\mathbb{E}[N_{\mathrm{num}}]^2} - \frac{2\,\kappa_3[N_{\mathrm{num}}]\;\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]}{\mathbb{E}[N_{\mathrm{num}}]^3} + \frac{\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^3}{\mathbb{E}[N_{\mathrm{num}}]^4}\right) + 2\,\frac{\mathbb{E}\big[\frac{1}{L^2}\big]}{\mathbb{E}\big[\frac{1}{L}\big]^2}\,\frac{\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^2}{\mathbb{E}[N_{\mathrm{num}}]^2},$$

and

$$\tau_{\mathrm{iod}}{}^2 := \mathbb{E}\big[\tfrac{1}{L}\big]\left(\frac{\kappa_4[N_{\mathrm{num}}]}{\mathbb{E}[N_{\mathrm{num}}]^2} - \frac{2\,\kappa_3[N_{\mathrm{num}}]\;\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]}{\mathbb{E}[N_{\mathrm{num}}]^3} + \frac{\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^3}{\mathbb{E}[N_{\mathrm{num}}]^4}\right) + 2\,\frac{\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^2}{\mathbb{E}[N_{\mathrm{num}}]^2},$$

with the third- and fourth-order cumulants κ_3 *and* κ_4*.*

Proof. Define

$$V_1 := \frac{1}{m-1}\sum_{j=1}^{m}\left(\frac{N_j}{L_j} - \frac{1}{m}\sum_{i=1}^{m}\frac{N_i}{L_i}\right)^2, \qquad E_1 := \frac{1}{m}\sum_{j=1}^{m}\frac{N_j}{L_j{}^2},$$

$$V_2 := \frac{1}{m}\sum_{j=1}^{m}\frac{N_j{}^2}{L_j} - \frac{1}{m}\,\frac{\left(\sum_{j=1}^{m} N_j\right)^2}{\sum_{j=1}^{m} L_j}, \qquad E_2 := \frac{1}{m}\sum_{j=1}^{m}\frac{N_j}{L_j},$$

so that $\hat{D}_1 = {}^{V_1}/_{E_1}$ and $\hat{D}_2 = {}^{V_2}/_{E_2}$. Next, consider the function

$$f\colon \mathbb{R}_{>0}{}^2 \to \mathbb{R}\colon\;\; (x,y) \longmapsto \frac{x}{y}.$$

For both $i \in \{1,2\}$ a result of Cramér [Cra62, pp. 366–367] provides the asymptotic normality of $f(V_i, E_i)$ if m is large,

$$\hat{D}_i = \frac{V_i}{E_i} = f(V_i, E_i) \overset{a}{\sim} \mathcal{N}(\mu_i, \sigma_i),$$

with mean

$$\mu_i = f\big(\mathbb{E}[V_i]\,,\mathbb{E}[E_i]\,\big)\Big|_{\mathcal{O}(m^{-1})} = \frac{\mathbb{E}[V_i]}{\mathbb{E}[E_i]}\bigg|_{\mathcal{O}(m^{-1})}$$

and variance

$$\sigma_i{}^2 = \Bigg[\Big(\tfrac{\partial f}{\partial x}\big(\mathbb{E}[V_i]\,,\mathbb{E}[E_i]\,\big)\Big)^2\,\mathbb{V}\mathrm{ar}[V_i] + \Big(\tfrac{\partial f}{\partial y}\big(\mathbb{E}[V_i]\,,\mathbb{E}[E_i]\,\big)\Big)^2\,\mathbb{V}\mathrm{ar}[E_i] + 2\,\tfrac{\partial f}{\partial x}\big(\mathbb{E}[V_i]\,,\mathbb{E}[E_i]\,\big)\;\tfrac{\partial f}{\partial y}\big(\mathbb{E}[V_i]\,,\mathbb{E}[E_i]\,\big)\;\mathbb{C}\mathrm{ov}[V_i, E_i]\Bigg|_{\mathcal{O}(m^{-3/2})},$$

where $|_{\mathcal{O}(m^{-1})}$ and $|_{\mathcal{O}(m^{-3/2})}$ mean that only those parts of the corresponding terms are considered which are of a smaller order than m^{-1} and $m^{-3/2}$ respectively. Analyzing the partial derivatives yields

$$\sigma_i{}^2 = \left[\frac{\mathbb{V}\mathrm{ar}[V_i]}{\mathbb{E}[E_i]^2} + \frac{\mathbb{E}[V_i]^2\,\mathbb{V}\mathrm{ar}[E_i]}{\mathbb{E}[E_i]^4} - \frac{2\,\mathbb{E}[V_i]\,\mathbb{C}\mathrm{ov}[V_i,E_i]}{\mathbb{E}[E_i]^3} \right|_{\mathcal{O}(m^{-3/2})}.$$

The calculations of the expectations, variances and covariances are quite complex. They can be looked up in the appendix (see Lemma A.4). However, it holds

$$\begin{aligned}
\mathbb{E}[E_1] &= \mathbb{E}\left[\tfrac{1}{L}\right]\mathbb{E}[N_{\mathrm{num}}]\,,\\
\mathbb{E}[E_2] &= \mathbb{E}[N_{\mathrm{num}}]\,,\\
\mathbb{E}[V_1] &= \mathbb{E}\left[\tfrac{1}{L}\right]\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]\,,\\
\mathbb{E}[V_2] &= \mathbb{V}\mathrm{ar}[N_{\mathrm{num}}] - \frac{1}{m}\,\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]\,,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[E_1] &= \frac{1}{m}\left(\mathbb{E}\left[\tfrac{1}{L^3}\right]\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}] + \mathbb{V}\mathrm{ar}\left[\tfrac{1}{L}\right]\mathbb{E}[N_{\mathrm{num}}]^2\right),\\
\mathbb{V}\mathrm{ar}[E_2] &= \frac{1}{m}\,\mathbb{E}\left[\tfrac{1}{L}\right]\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]\,,\\
\mathbb{V}\mathrm{ar}[V_1] &= \frac{1}{m}\left(\mathbb{E}\left[\tfrac{1}{L^3}\right]\kappa_4[N_{\mathrm{num}}] + \left(3\,\mathbb{E}\left[\tfrac{1}{L^2}\right] - \mathbb{E}\left[\tfrac{1}{L}\right]^2\right)\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^2\right) + \mathcal{O}\big(m^{-2}\big)\,,\\
\mathbb{V}\mathrm{ar}[V_2] &= \frac{1}{m}\left(\mathbb{E}\left[\tfrac{1}{L}\right]\kappa_4[N_{\mathrm{num}}] + 2\,\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^2\right) + \mathcal{O}\big(m^{-2}\big)\,,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[V_1,E_1] &= \frac{1}{m}\Big(\mathbb{E}\left[\tfrac{1}{L^3}\right]\kappa_3[N_{\mathrm{num}}] + \mathbb{V}\mathrm{ar}\left[\tfrac{1}{L}\right]\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]\,\mathbb{E}[N_{\mathrm{num}}]\Big) + \mathcal{O}\big(m^{-2}\big)\,,\\
\mathbb{C}\mathrm{ov}[V_2,E_2] &= \frac{1}{m}\,\mathbb{E}\left[\tfrac{1}{L}\right]\kappa_3[N_{\mathrm{num}}] + \mathcal{O}\big(m^{-2}\big)\,.
\end{aligned}$$

So, it follows

$$\sigma_1{}^2 = \frac{\sigma_{\mathrm{iod}}{}^2}{m}, \quad \sigma_2{}^2 = \frac{\tau_{\mathrm{iod}}{}^2}{m} \quad \text{and} \quad \mu_i = \frac{\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]}{\mathbb{E}[N_{\mathrm{num}}]} = \mathbb{D}[N_{\mathrm{num}}] \quad \forall i \in \{1,2\},$$

and therefore

$$\hat{D}_1 \overset{a}{\sim} \mathcal{N}\big(\mathbb{D}[N_{\mathrm{num}}]\,,\tfrac{\sigma_{\mathrm{iod}}{}^2}{m}\big) \qquad \text{and} \qquad \hat{D}_2 \overset{a}{\sim} \mathcal{N}\big(\mathbb{D}[N_{\mathrm{num}}]\,,\tfrac{\tau_{\mathrm{iod}}{}^2}{m}\big),$$

which is just a transformation of the proposition. □

The asymptotic variances ${\sigma_{\mathrm{iod}}}^2$ and ${\tau_{\mathrm{iod}}}^2$ look much alike, but because of

$$\frac{\mathbb{E}\left[\frac{1}{L^3}\right]}{\mathbb{E}\left[\frac{1}{L}\right]^2} = \mathbb{E}\left[\tfrac{1}{L}\right] \frac{\mathbb{E}\left[\frac{1}{L^3}\right]}{\mathbb{E}\left[\frac{1}{L}\right]^3} \geq \mathbb{E}\left[\tfrac{1}{L}\right] \qquad \text{and} \qquad \frac{\mathbb{E}\left[\frac{1}{L^2}\right]}{\mathbb{E}\left[\frac{1}{L}\right]^2} \geq 1,$$

the asymptotic variance of $\hat{D}_1$, ${\sigma_{\mathrm{iod}}}^2$, is larger than the asymptotic variance of $\hat{D}_2$, ${\tau_{\mathrm{iod}}}^2$. In this sense, $\hat{D}_2$ is a better estimator than the intuitive estimator $\hat{D}_1$.

However, the variances ${\sigma_{\mathrm{iod}}}^2$ and ${\tau_{\mathrm{iod}}}^2$ depend on unknown distribution parameters. The estimation of them needs estimates of the first three (non-central) moments of $1/L$ and of the first four cumulants of N_{num} (remember that expectation and variance are equal to the first and the second order cumulant respectively, see Section 2.4.6).

Generally, if $X = (X_1, \dots, X_m)$ is a random vector consisting of statistically independent and identically distributed random variables, the sample moments $\hat{\mu}_n(X)$,

$$\hat{\mu}_n(X) := \frac{1}{m} \sum_{j=1}^{m} {X_j}^n \qquad \forall n \in \mathbb{N},$$

are unbiased and consistent estimators of the (non-central) moments $\mathbb{E}[X^n]$ [Cra62, p. 346]. On the other hand, the statistics $\hat{k}_1, \dots, \hat{k}_4$ defined by

$$\begin{aligned}
\hat{k}_1(X) &:= \hat{\mu}_1(X), \\
\hat{k}_2(X) &:= \frac{m}{m-1} \left(\hat{\mu}_2(X) - \hat{\mu}_1(X)^2\right), \\
\hat{k}_3(X) &:= \frac{m^2}{(m-1)(m-2)} \Big(\hat{\mu}_3(X) - 3\,\hat{\mu}_2(X)\,\hat{\mu}_1(X) + 2\,\hat{\mu}_1(X)^3\Big), \\
\hat{k}_4(X) &:= \frac{m^3}{(m-1)(m-2)(m-3)} \Big(\tfrac{m+1}{m}\,\hat{\mu}_4(X) - 4\,\tfrac{m+1}{m}\,\hat{\mu}_3(X)\,\hat{\mu}_1(X) \\
&\qquad - 3\,\tfrac{m-1}{m}\,\hat{\mu}_2(X)^2 + 12\,\hat{\mu}_2(X)\,\hat{\mu}_1(X)^2 - 6\,\hat{\mu}_1(X)^4\Big)
\end{aligned} \tag{3.7}$$

are unbiased and consistent estimators of the first four cumulants [Fis29, Cra62, p. 352] (Kenney and Keeping [KK65, pp. 189–190] also give standard errors of these so called ***k*-statistics**).

Since the variate N_{num} is not observable, the cumulants of N_{num} must be expressed in terms of N^{*L}_{num} and L, first (similar to the index of dispersion itself, see Section 3.5.1). For this purpose, let us use the variate N^{*L}_{num}/L, because $\kappa_1\left[N^{*L}_{\mathrm{num}}/L \,\middle|\, L\right] = \mathbb{E}\left[N^{*L}_{\mathrm{num}}/L \,\middle|\, L\right] = \mathbb{E}[N_{\mathrm{num}}]$ is almost sure constant (see proof of Lemma 3.5.1). Thus, it is straightforward to calculate the cumulants from their conditional versions: homogeneity and additivity of the cumulants (see Section

2.4.6) ensure

$$\kappa_n\Big[\tfrac{N_{\text{num}}^{*L}}{L}\Big|\,L\Big] = \frac{1}{L^n}\sum_{i=1}^{L}\kappa_n[N_{\text{num}}|L] = \frac{1}{L^{n-1}}\,\kappa_n[N_{\text{num}}] \qquad \forall n\in\{1,2,3,4\},$$

and so the law of total cumulants (see Section 2.4.6) yields

$$\begin{aligned}\kappa_n\Big[\tfrac{N_{\text{num}}^{*L}}{L}\Big] &= \mathbb{E}\big[\tfrac{1}{L^{n-1}}\big]\,\kappa_n[N_{\text{num}}] \qquad \forall n\in\{1,2,3\},\\ \kappa_4\Big[\tfrac{N_{\text{num}}^{*L}}{L}\Big] &= \mathbb{E}\big[\tfrac{1}{L^3}\big]\,\kappa_4[N_{\text{num}}] + 3\,\mathbb{V}\text{ar}\big[\tfrac{1}{L}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}]^2 .\end{aligned} \tag{3.8}$$

Conversely, it must hold

$$\kappa_n[N_{\text{num}}] = \frac{\kappa_n\Big[\tfrac{N_{\text{num}}^{*L}}{L}\Big]}{\mathbb{E}\big[\tfrac{1}{L^{n-1}}\big]} - 3\,\mathbb{V}\text{ar}\Big[\tfrac{N_{\text{num}}^{*L}}{L}\Big]^2\,\frac{\mathbb{V}\text{ar}\big[\tfrac{1}{L}\big]}{\mathbb{E}\big[\tfrac{1}{L^3}\big]\,\mathbb{E}\big[\tfrac{1}{L}\big]^2}\,\mathbb{1}_{\{4\}}(n) \quad \forall n\in\{1,2,3,4\}.$$

These expressions for the cumulants of N_{num} can replace the respective terms in ${\sigma_{\text{iod}}}^2$ and ${\tau_{\text{iod}}}^2$ as defined in Theorem 3.5.2. Then, the cumulants of N_{num}^{*L}/L and the moments of $1/L$ can be estimated by the estimators $\hat{\mu}_n$ and $\hat{k}_n$ as defined above. As a result, the terms

$$\begin{aligned}{\hat{\sigma}_{\text{iod}}}^2 &= \frac{\hat{k}_4(U) - 3\,\frac{\hat{k}_2(K)}{\hat{\mu}_1(K)^2}\,\hat{k}_2(U)^2}{\hat{\mu}_1(K)^2\,\hat{\mu}_1(U)^2} - \frac{2\,\hat{\mu}_3(K)\,\hat{k}_3(U)\,\hat{k}_2(U)}{\hat{\mu}_2(K)\,\hat{\mu}_1(K)^3\,\hat{\mu}_1(U)^3} + \frac{\hat{\mu}_3(K)\,\hat{k}_2(U)^3}{\hat{\mu}_1(K)^5\,\hat{\mu}_1(U)^4}\\ &\qquad + \frac{2\,\hat{\mu}_2(K)\,\hat{k}_2(U)^2}{\hat{\mu}_1(K)^4\,\hat{\mu}_1(U)^2},\\ {\hat{\tau}_{\text{iod}}}^2 &= \frac{\hat{\mu}_1(K)\,\hat{k}_4(U) - 3\,\frac{\hat{k}_2(K)}{\hat{\mu}_1(K)}\,\hat{k}_2(U)^2}{\hat{\mu}_3(K)\,\hat{\mu}_1(U)^2} - \frac{2\,\hat{k}_3(U)\,\hat{k}_2(U)}{\hat{\mu}_2(K)\,\hat{\mu}_1(U)^3} + \frac{\hat{k}_2(U)^3}{\hat{\mu}_1(K)^2\,\hat{\mu}_1(U)^4}\\ &\qquad + \frac{2\,\hat{k}_2(U)^2}{\hat{\mu}_1(K)^2\,\hat{\mu}_1(U)^2},\end{aligned}$$

are consistent estimators of ${\sigma_{\text{iod}}}^2$ and ${\tau_{\text{iod}}}^2$ respectively, where K and U are random vectors,

$$K := \big(\tfrac{1}{L_1},\ldots,\tfrac{1}{L_m}\big) \qquad \text{and} \qquad U := \big(\tfrac{\mathcal{N}_1}{L_1},\ldots,\tfrac{\mathcal{N}_m}{L_m}\big),$$

consisting of the variates $(L_j)_{j\in\mathbb{N}}$ and $(\mathcal{N}_j)_{j\in\mathbb{N}}$ from Theorem 3.5.2.

Finally, Theorem 3.5.2 provides approximate confidence intervals of $\mathbb{D}[N_{\text{num}}]$ based on the estimators $\hat{D}_1$, ${\hat{\sigma}_{\text{iod}}}^2$ and $\hat{D}_2$, ${\hat{\tau}_{\text{iod}}}^2$. According to this theorem and the consistency of ${\hat{\sigma}_{\text{iod}}}^2$, it holds

$$\begin{aligned}1-\alpha &= \lim_{m\to\infty}\mathbb{P}\Big(-q_{1-\alpha/2} \le \sqrt{\tfrac{m}{{\hat{\sigma}_{\text{iod}}}^2}}\,\Big(\hat{D}_1 - \mathbb{D}[N_{\text{num}}]\Big) \le q_{1-\alpha/2}\Big)\\ &= \lim_{m\to\infty}\mathbb{P}\Big(\hat{D}_1 - \tfrac{\hat{\sigma}_{\text{iod}}\,q_{1-\alpha/2}}{\sqrt{m}} \le \mathbb{D}[N_{\text{num}}] \le \hat{D}_1 + \tfrac{\hat{\sigma}_{\text{iod}}\,q_{1-\alpha/2}}{\sqrt{m}}\Big)\end{aligned}$$

if $q_{1-\alpha/2}$ denotes the $\left(1-\frac{\alpha}{2}\right)100\,\%$ quantile of the standard normal distribution ($\alpha \in (0,1)$). The same relation holds for $\hat{D}_2$ and $\hat{\tau}_{\mathrm{iod}}{}^2$. This yields the approximate confidence intervals

$$\left[\hat{D}_1 - \frac{\hat{\sigma}_{\mathrm{iod}}\, q_{1-\alpha/2}}{\sqrt{m}},\ \hat{D}_1 + \frac{\hat{\sigma}_{\mathrm{iod}}\, q_{1-\alpha/2}}{\sqrt{m}}\right], \quad \left[\hat{D}_2 - \frac{\hat{\tau}_{\mathrm{iod}}\, q_{1-\alpha/2}}{\sqrt{m}},\ \hat{D}_2 + \frac{\hat{\tau}_{\mathrm{iod}}\, q_{1-\alpha/2}}{\sqrt{m}}\right]$$

with confidence level $(1-\alpha)$. Again, since $\hat{\sigma}_{\mathrm{iod}}{}^2$ is larger than $\hat{\tau}_{\mathrm{iod}}{}^2$, the second confidence interval is smaller than the first one.

3.5.3. Hypothesis Test for Poisson Approach

Theorem 3.5.2 provides the asymptotic distributions of $\hat{D}_1$ and $\hat{D}_2$, the estimators of the index of dispersion of N_{num} (see Equation (3.6) on page 40). Since the variances of this asymptotic distributions, $\sigma_{\mathrm{iod}}{}^2$ and $\tau_{\mathrm{iod}}{}^2$, are expressed in terms of the variates N_{num}, it is easy to design a significance test concerning the null hypothesis 'N_{num} is Poisson distributed'. For this purpose, the next corollary provides expressions for the variances $\sigma_{\mathrm{iod}}{}^2$ and $\tau_{\mathrm{iod}}{}^2$ in the Poisson case.

3.5.3 Corollary. *Let the situation be as in Theorem 3.5.2. If N_{num} is Poisson distributed, the variances of the limiting normal distributions are*

$$\sigma_{\mathrm{iod}}{}^2 = \frac{2\,\mathbb{E}\left[\frac{1}{L^2}\right]}{\mathbb{E}\left[\frac{1}{L}\right]^2} = 2 + 2\,\mathbb{CV}\left[\tfrac{1}{L}\right]^2 \qquad \text{and} \qquad \tau_{\mathrm{iod}}{}^2 = 2.$$

Proof. All the cumulants of a Poisson random variable are equal to the mean [LC98, p. 30]. □

The last corollary provides that $\tau_{\mathrm{iod}}{}^2$ is not only smaller than $\sigma_{\mathrm{iod}}{}^2$, but it does not depends on the mileages as long as N_{num} is Poisson distributed. Under the hypothesis that N_{num} is Poisson distributed, the estimator

$$\hat{D}_2 = \hat{D}_2\big((n_j, l_j)_{1\le j\le m}\big) := \frac{\frac{1}{m}\sum_{j=1}^m \frac{n_j{}^2}{l_j} - \frac{1}{m}\frac{\left(\sum_{j=1}^m n_j\right)^2}{\sum_{j=1}^m l_j}}{\frac{1}{m}\sum_{j=1}^m \frac{n_j}{l_j}}$$

is, according to Theorem 3.5.2, (approximately) normally distributed with mean 1 and variance $2/m$, because the index of dispersion $\mathbb{D}[N_{\mathrm{num}}]$ is equal to 1 and the variance $\tau_{\mathrm{iod}}{}^2$ is equal to 2 (see Definition 2.4.2 and Corollary 3.5.3). In other words:

$$\lim_{m\to\infty} \mathbb{P}\Big(-q_{1-\alpha/2} \le \sqrt{\tfrac{m}{2}}\big(\hat{D}_2 - 1\big) \le q_{1-\alpha/2}\Big) = 1-\alpha,$$

where again $q_{1-\alpha/2}$ is the $\left(1-\frac{\alpha}{2}\right)100\,\%$ quantile of the standard normal distribution. All these facts result in the following significance test:

Hypothesis test for Poisson approach

1. The observations $(n_1, l_1), \ldots, (n_m, l_m) \in \mathbb{N}_0 \times \mathbb{N}$ are realizations of $\big(N^{*L}_{\text{num}}, L\big)$.
2. Calculate the test statistic $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ based on the observation.
3. If it is
$$\left|\sqrt{\tfrac{m}{2}}\big(\hat{D}_2 - 1\big)\right| > q_{1-\alpha/2},$$
reject the null hypothesis that N_{num} is Poisson distributed with significance level $1-\alpha$.

3.5.4. Overdispersion

If the hypothesis of a Poisson distributed N_{num} is rejected, an alternative is needed. It is important to distinguish two cases: $\mathbb{V}\text{ar}[N_{\text{num}}]$ is significant larger or smaller than $\mathbb{E}[N_{\text{num}}]$.

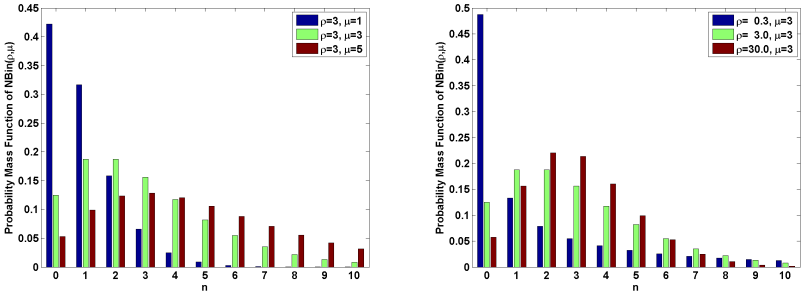
If the variance of N_{num} exceeds its expectation, N_{num} is called to be **overdispersed** [IJ07]. Many authors [IJ07, JZ05, YHAV07, LSWY02, LWI05, and references therein] propose the generalized Poisson distribution or the negative binomial distribution as an alternative to the Poisson distribution when dealing with overdispersion. Both distributions are related to the Poisson model. They are so-called **Poisson mixture distributions** [GY20, JZ05]. Poisson mixture means the following: suppose, there are random variables Y and W such that $Y|W\!=\!\omega$ is a Poisson variate with mean ω ($\omega \in \mathbb{R}_{>0}$). Then, the distribution of Y is called to be a Poisson mixture distribution. If f_W denotes the probability density function of W, the distribution of Y is specified through

$$\mathbb{P}(Y=n) = \int_0^\infty \mathbb{P}(Y=n \,|\, W=\omega)\ f_W(\omega)\ \mathrm{d}\omega = \int_0^\infty \mathrm{e}^{-\omega}\ \frac{\omega^n}{n!}\ f_W(\omega)\ \mathrm{d}\omega$$

for all $n \in \mathbb{N}_0$.

The generalized Poisson distribution as given in Definition 2.4.2 was first introduced by Consul and Jain [CJ73] and studied extensively by Consul [Con89]. Its benefit is that the generalized Poisson model not only deals with overdispersion but with underdispersion, too. Consul [Con89, p. 3] declares that *"the variance of this generalized Poisson distribution model is greater than, equal to, or less than the mean according to whether the second parameter, λ, is positive, zero, or negative, and both mean and variance tend to increase or decrease in value as θ increases or decreases"*. Here, in Definition 2.4.2, the generalized Poisson distribution is not used for negative values of λ. Nelson [Nel75] gives some reasons

Figure 3.2.: *Probability mass functions of negative binomial distribution.*



for that. He indicates a cautious handling with negative values of λ, because the distribution is truncated then and, more importantly, the generalized Poisson model does not satisfy some properties of a distribution function, e. g., except for a negligible proportion of choices for $\lambda \in [-1, 0)$ and θ, the probabilities do not add up to 1.

The reason for disqualifying the generalized Poisson distribution in context of overdispersion, too, is the fact that the distribution of the number of events per any range and mileage, $Z_{l,A}$, is not in the same family as the distribution of N_{num} if N_{num} is generalized Poisson distributed (see Example 3.2.2). In this case, $Z_{l,A}$ is rather distributed according to a compound generalized Poisson distribution as proposed by Ambagaspitiya and Balakrishnan [AB94]. But this distribution family is very unwieldy in the sense that the probability mass function is not expressible in closed form.

The situation with the negative binomial distribution is quite different. As shown in Example 3.2.2, if N_{num} is negative binomial, then $Z_{l,A}$ is negative binomial, too. In contrast to the generalized Poisson distribution, where Joe and Zhu [JZ05] proofed the Poisson mixture property but could not find the mixing distribution, Greenwood and Yules [GY20] show that the negative binomial distribution is a Poisson mixture where the mixing distribution of the mean is a gamma distribution. More precisely, they suppose f_W to be the probability density function of a gamma distribution with parameters[1] ϱ and $\frac{\mu}{\varrho}$ $(\varrho, \mu \in \mathbb{R}_{>0})$, $\Gamma\big(\varrho, \frac{\mu}{\varrho}\big)$, and they calculate (with $Y|W\!=\!\omega \sim \text{Poi}(\omega)$ as defined above)

$$\mathbb{P}(Y = n) = \int_0^\infty \mathrm{e}^{-\omega}\,\frac{\omega^n}{n!}\,\frac{\omega^{\varrho-1}\,\mathrm{e}^{-\frac{\varrho}{\mu}\omega}}{\left(\frac{\mu}{\varrho}\right)^{\varrho}\Gamma(\varrho)}\,\mathrm{d}\omega = \frac{\Gamma(\varrho+n)}{n!\;\Gamma(\varrho)}\left(\frac{\varrho}{\varrho+\mu}\right)^{\varrho}\left(\frac{\mu}{\varrho+\mu}\right)^{n}$$

for all $n \in \mathbb{N}_0$. Hence, Y is negative binomially distributed.

[1] Greenwood and Yules [GY20] use a different parametrization

3.5.5. Underdispersion

Underdispersion, where the variance is less than the expectation, is more unusual than overdispersion. Also in the present situation, it is more expectable that the whole population consists of several subpopulations, which increases the variance and so the dispersion. Ridout and Besbeas [RB04] give a short overview of several models for underdispersed count data and compare them to their exponentially weighted Poisson model. They mention other weighted Poisson models (Poisson polynomial distribution, power law weighted Poisson distribution), the double Poisson distribution, the changing birth rate distribution and the COM-Poisson distribution. All these distributions are suitable for modeling underdispersion. However, these models become more and more complicated and difficult. Here, the complexity would only increase as the counts are divided into subgroups (see Proposition 3.2.1). In face of the fact that underdispersion is not expected, for the sake of completeness a simple model for underdispersion shall be noticed here, and that is the binomial model.

The binomial distribution is simple, the variance is always less than the expectation, Example 3.2.2 shows that $Z_{l,A}$ inherits the binomial distribution from N_{num}, and the binomial distribution is related to the Poisson distribution since the Poisson Limit Theorem [Als05, p. 131] says that the $\text{Bin}(r, q)$ distribution approaches the $\text{Poi}(\lambda)$ distribution if r approaches ∞ and q approaches 0 while the term rq remains fixed at λ.

3.5.6. Binomial, Poisson and Negative Binomial Approach for High Mileages

Let $Z^{\text{Bin}}_{l,A}$, $Z^{\text{Poi}}_{l,A}$ and $Z^{\text{NBin}}_{l,A}$ be as distributed as $Z_{l,A}$ ($l \in \mathbb{N}$, $A \in \mathfrak{S}$) under the constraint of N_{num} being binomially, Poisson and negative binomially distributed, respectively. It is well-known that the binomial and the negative binomial distribution are very similar to the Poisson distribution if certain conditions hold. The Poisson Limit Theorem [Als05, p. 131] says that

$$\lim_{\substack{r\to\infty \\ rq\to\lambda}} \binom{r}{n} q^n (1-q)^{r-n} = \mathrm{e}^{-\lambda}\,\frac{\lambda^n}{n!} \qquad \forall n \in \mathbb{N}_0.$$

On the other hand, the negative binomial distribution tends to the Poisson distribution if the exponent tends to ∞ [FCW43],

$$\lim_{\substack{\varrho\to\infty \\ \mu\to\lambda}} \frac{\Gamma(\varrho+n)}{n!\,\Gamma(\varrho)} \left(\frac{\varrho}{\varrho+\mu}\right)^{\varrho} \left(\frac{\mu}{\varrho+\mu}\right)^{n} = \mathrm{e}^{-\lambda}\,\frac{\lambda^n}{n!} \qquad \forall n \in \mathbb{N}_0.$$

The question is whether the distributions of $Z^{\text{Bin}}_{l,A}$, $Z^{\text{Poi}}_{l,A}$ and $Z^{\text{NBin}}_{l,A}$ differ significantly from each other if the mileage l is high. If, for instance, $Z^{\text{Poi}}_{l,A}$ and

$Z_{l,A}^{\text{NBin}}$ became too similar to each other for large mileages, the negative binomial distribution would not be an adequate model for overdispersion, because the overdispersion of the data does not vanish for high mileages. The same holds true for the binomial approach and underdispersion. Since the index of dispersion of $Z_{l,A}$ is independent of the mileage though,

$$\mathbb{D}[Z_{l,A}] = 1 + p_A\big(\mathbb{D}[N_{\text{num}}] - 1\big)$$

(see Proposition 3.2.1), the binomial and the negative binomial distributions are still appropriate approaches for modeling underdispersion and overdispersion, respectively, even if the mileage l is high.

However, the situation is different with the maximum SOLE M_{sev}^{*l}. Let M_{Bin}^{*l}, M_{Poi}^{*l} and M_{NBin}^{*l} be random variables which are distributed according to M_{sev}^{*l} under the constraint of N_{num} being binomially, Poisson and negative binomially distributed, respectively, and suppose that there are series $(a_l)_{l\in\mathbb{N}} \subseteq \mathbb{R}_{>0}$ and $(b_l)_{l\in\mathbb{N}} \subseteq \mathbb{R}$ such that

$$\mathbb{P}\Big(\tfrac{M_{\text{Poi}}^{*l} - b_l}{a_l} \leq t\Big) \xrightarrow{l\to\infty} H(t)$$

for each continuity point of the non-degenerated cumulative distribution function H. The Fisher–Tippett Theorem (see Theorem 2.4.3) provides that in this situation H must be the rescaled cumulative distribution function of the generalized extreme value distribution with a specific parameter ξ. If the mileage is high, the number of SOLEs is large, too, regardless of the variance of N_{num}. Hence, the behavior of M_{sev}^{*l} is mostly determined by the distribution of the SOLEs, F_{sev}. Therefore, it is expectable that $(M_{\text{Bin}}^{*l} - b_l)/a_l$ and $(M_{\text{NBin}}^{*l} - b_l)/a_l$ converge in distribution to the same generalized extreme value distribution.

The following theorem shows that, in fact, the distributions of M_{Bin}^{*l} and M_{NBin}^{*l} can be approximated by the distribution of M_{Poi}^{*l} if the mileage l is high.

3.5.4 Theorem. *Let be $l, r \in \mathbb{N}$, $q \in (0,1)$ and $\lambda, \varrho, \mu \in \mathbb{R}_{>0}$. Suppose, M_{Bin}^{*l}, M_{Poi}^{*l} and M_{NBin}^{*l} are random variables with the following distributions:*

- M_{Bin}^{*l} *is as distributed as* M_{sev}^{*l} *under the constraint* $N_{\text{num}} \sim \text{Bin}(r, q)$,
- M_{Poi}^{*l} *is as distributed as* M_{sev}^{*l} *under the constraint* $N_{\text{num}} \sim \text{Poi}(\lambda)$,
- M_{NBin}^{*l} *is as distributed as* M_{sev}^{*l} *under the constraint* $N_{\text{num}} \sim \text{NBin}(\varrho, \mu)$.

Then it holds:

1. *If $\lambda = rq$, then*

$$0 \leq \mathbb{P}\Big(M_{\text{Poi}}^{*l} \leq t\Big) - \mathbb{P}\Big(M_{\text{Bin}}^{*l} \leq t\Big) \leq \frac{1}{rl\mathrm{e}} \qquad \forall t \in \mathbb{R}.$$

2. If $\lambda = \mu$, then

$$0 \le \mathbb{P}\Big(M^{*l}_{\mathrm{NBin}} \le t\Big) - \mathbb{P}\Big(M^{*l}_{\mathrm{Poi}} \le t\Big) \le \frac{1}{\varrho l \mathrm{e}} \qquad \forall t \in \mathbb{R}.$$

3. If $\mu = rq$, then

$$0 \le \mathbb{P}\Big(M^{*l}_{\mathrm{NBin}} \le t\Big) - \mathbb{P}\Big(M^{*l}_{\mathrm{Bin}} \le t\Big) \le \left(\frac{1}{r} + \frac{1}{\varrho}\right)\frac{1}{l\mathrm{e}} \qquad \forall t \in \mathbb{R}.$$

Proof. Let us have a look at the functions $f\colon [0,1] \to \mathbb{R}$ and $g\colon \mathbb{R}_{\ge 0} \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} \mathrm{e}^{-rlx} - \mathrm{e}^{rl\log(1-x)}, & \text{if } x \in [0,1) \\ \mathrm{e}^{-rl}, & \text{if } x = 1. \end{cases} \quad \text{and} \quad g(x) := \mathrm{e}^{-\varrho l \log(1+x)} - \mathrm{e}^{-\varrho l x}.$$

From

$$\lim_{x \nearrow 1} f(x) = \mathrm{e}^{-rl} = f(1), \quad g(0) = 0 = \lim_{x\to\infty} g(x), \quad x \ge \log(1+x) \quad \forall x \in \mathbb{R}_{>-1}$$

it follows that both f and g are nonnegative and continuous, and it follows that there are points $x_0 \in [0,1]$, $y_0 \in \mathbb{R}_{\ge 0}$ such that

$$f(x_0) = \max_{x\in[0,1]} f(x) \qquad \text{and} \qquad g(y_0) = \max_{x \in \mathbb{R}_{\ge 0}} g(x).$$

For all x in the domains of f and g respectively, the following two equivalences hold:

$$\Big[\tfrac{\mathrm{d}f}{\mathrm{d}x}(x) = 0 \quad \Leftrightarrow \quad x = 1 - \mathrm{e}^{rl(\log(1-x)+x)}\Big],$$
$$\Big[\tfrac{\mathrm{d}g}{\mathrm{d}x}(x) = 0 \quad \Leftrightarrow \quad x = \mathrm{e}^{-\varrho l(\log(1+x)-x)} - 1\Big].$$

This implies

$$f(x_0) = \left\{\begin{matrix} \Big(1 - \mathrm{e}^{rl(\log(1-x_0)+x_0)}\Big)\,\mathrm{e}^{-rlx_0}, & \text{if } x_0 \in [0,1) \\ \mathrm{e}^{-rl}, & \text{if } x_0 = 1 \end{matrix}\right\} = x_0\,\mathrm{e}^{-rlx_0},$$
$$g(y_0) = \Big(\mathrm{e}^{-\varrho l(\log(1+y_0)-y_0)} - 1\Big)\,\mathrm{e}^{-\varrho l y_0} = y_0\,\mathrm{e}^{-\varrho l y_0}.$$

The maximizer of the continuous function $x \mapsto x\,\mathrm{e}^{-rlx}$ is $\frac{1}{rl}$, which can be verified by discovering its derivative $x \mapsto (1-rlx)\,\mathrm{e}^{-rlx}$.

All in all, both f and g are nonnegative and do not exceed $(rl\mathrm{e})^{-1}$ and $(\varrho l\mathrm{e})^{-1}$ respectively. Eventually, consequences of Proposition 3.4.2 and Example 3.4.3 are

$$\mathbb{P}\Big(M^{*l}_{\mathrm{Poi}} \le t\Big) - \mathbb{P}\Big(M^{*l}_{\mathrm{Bin}} \le t\Big) = f\big(q(1-F_{\mathrm{sev}}(t))\big)\,\mathbb{1}_{\mathbb{R}_{\ge 0}}(t) \qquad \forall t \in \mathbb{R}$$

if $\lambda = rq$, and

$$\mathbb{P}\Big(M^{*l}_{\mathrm{NBin}} \le t\Big) - \mathbb{P}\Big(M^{*l}_{\mathrm{Poi}} \le t\Big) = g\Big(\tfrac{\mu}{\varrho}(1 - F_{\mathrm{sev}}(t))\Big)\ \mathbb{1}_{\mathbb{R}_{\ge 0}}(t) \qquad \forall t \in \mathbb{R}$$

if $\lambda = \mu$. The third statement follows from the first two results if it is chosen $\lambda = \mu = rq$. □

With the results of Theorem 3.5.4 it is evident that both $(M^{*l}_{\mathrm{Bin}} - b_l)/a_l$ and $(M^{*l}_{\mathrm{NBin}} - b_l)/a_l$ converge in distribution to the same generalized extreme value distribution as $(M^{*l}_{\mathrm{Poi}} - b_l)/a_l$ does. Of course, the exact extreme value parameter ξ of the limiting distribution and the correct selection of the series $(a_l)_{l\in\mathbb{N}}$ and $(b_l)_{l\in\mathbb{N}}$ depend on the distribution F_{sev} of the SOLEs.

As an example, let the shifted SOLE $(S_{\mathrm{sev}} - u_{\mathrm{sev}})$ be generalized Pareto distributed,

$$F_{\mathrm{sev}}(t) = F_{\mathrm{GPar}(\xi,\beta)}(t - u_{\mathrm{sev}}) \qquad \forall t \in \mathbb{R},$$

where $F_{\mathrm{GPar}(\xi,\beta)}$ is the cumulative distribution function of the generalized Pareto distribution with shape ξ and scale β ($\xi \in \mathbb{R}$, $\beta \in \mathbb{R}_{>0}$, motivation for this example see Section 3.6). For each $l \in \mathbb{N}$ define

$$a_l := \beta(\lambda l)^{\xi} \qquad \text{and} \qquad b_l := \begin{cases} u_{\mathrm{sev}} + \beta\,\frac{(\lambda l)^{\xi} - 1}{\xi}, & \text{if } \xi \ne 0,\\ u_{\mathrm{sev}} + \beta \log(\lambda l)\,, & \text{if } \xi = 0,\end{cases}$$

where λ is the expectation value of N_{num}. With these definitions it holds

$$F_{\mathrm{sev}}(a_l t + b_l) = F_{\mathrm{GPar}(\xi,\beta)}(a_l t + b_l - u_{\mathrm{sev}}) = \begin{cases} 1 - \frac{(1+\xi t)^{-\frac{1}{\xi}}}{\lambda l}, & \text{if } \xi \ne 0\\ 1 - \frac{\mathrm{e}^{-t}}{\lambda l}, & \text{if } \xi = 0.\end{cases}$$

With help of Example 3.4.3 this yields

$$\begin{aligned}\mathbb{P}\Big(\tfrac{M^{*l}_{\mathrm{Poi}} - b_l}{a_l} \le t\Big) &= \mathrm{e}^{\lambda l (F_{\mathrm{sev}}(a_l t + b_l) - 1)}\ \mathbb{1}_{\mathbb{R}_{\ge 0}}(a_l t + b_l)\\ &= \mathrm{e}^{-\lambda l}\ \mathbb{1}_{[0,u_{\mathrm{sev}})}(a_l t + b_l) + F_{\mathrm{GEV}(\xi)}(t)\ \mathbb{1}_{\mathbb{R}_{\ge u_{\mathrm{sev}}}}(a_l t + b_l)\\ &= \mathrm{e}^{-\lambda l}\ \mathbb{1}_{\left[c_l - \frac{u_{\mathrm{sev}}}{a_l}, c_l\right)}(t) + F_{\mathrm{GEV}(\xi)}(t)\ \mathbb{1}_{\mathbb{R}_{\ge c_l}}(t)\\ &\xrightarrow{l\to\infty} F_{\mathrm{GEV}(\xi)}(t) \qquad \forall t \in \mathbb{R},\end{aligned}$$

where

$$c_l := \begin{cases} \frac{(\lambda l)^{-\xi} - 1}{\xi}, & \text{if } \xi \ne 0,\\ -\log(\lambda l)\,, & \text{if } \xi = 0.\end{cases}$$

Eventually, Theorem 3.5.4 provides

$$F_{\mathrm{GEV}(\xi)}(t) - \frac{1}{rl\mathrm{e}} \le \mathbb{P}\Big(\tfrac{M^{*l}_{\mathrm{Bin}} - b_l}{a_l} \le t\Big) \le F_{\mathrm{GEV}(\xi)}(t) \qquad \forall t \in \mathbb{R}_{\ge c_l}$$

and

$$F_{\mathrm{GEV}(\xi)}(t) \le \mathbb{P}\Big(\tfrac{M^{*l}_{\mathrm{NBin}}-b_l}{a_l} \le t\Big) \le F_{\mathrm{GEV}(\xi)}(t) + \frac{1}{\varrho l \mathrm{e}} \qquad \forall t \in \mathbb{R}_{\ge c_l}.$$

Consequently, with regard to the maximum SOLE, N_{num} can be assumed to be Poisson distributed if l is so large that $(rl\mathrm{e})^{-1}$ and $(\varrho l\mathrm{e})^{-1}$ are negligible small, where r and ϱ are the number of trials and the exponent of a possible alternative binomial and negative binomial distribution, respectively.

3.5.7. Conclusion

The distribution of N_{num} will be selected in the following way: A priori, N_{num} is assumed to be Poisson distributed, $N_{\mathrm{num}} \sim \mathrm{Poi}(\lambda)$. According to the hypothesis test in Section 3.5.3, evaluate the estimator $\hat{D}_2$ (see Equation (3.6) on page 40) and take a decision as follows:

$$\begin{aligned}
\left|\sqrt{\tfrac{m}{2}}\left(\hat{D}_2 - 1\right)\right| &\le q_{1-\alpha/2} &&\leadsto && N_{\mathrm{num}} \text{ remains Poisson distributed,}\\
\sqrt{\tfrac{m}{2}}\left(\hat{D}_2 - 1\right) &< -q_{1-\alpha/2} &&\leadsto && N_{\mathrm{num}} \text{ is underdispersed,}\\
\sqrt{\tfrac{m}{2}}\left(\hat{D}_2 - 1\right) &> q_{1-\alpha/2} &&\leadsto && N_{\mathrm{num}} \text{ is overdispersed,}
\end{aligned}$$

where $q_{1-\alpha/2}$ denotes the $\left(1-\frac{\alpha}{2}\right)100\,\%$ quantile of the standard normal distribution ($\alpha \in (0,1)$). If N_{num} is identified to be underdispersed, it shall be binomially distributed, $N_{\mathrm{num}} \sim \mathrm{Bin}(r,q)$. Otherwise, if N_{num} is identified to be overdispersed, choose a negative binomial distribution, $N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu)$.

There are three factual reasons to prefer the estimator $\hat{D}_2$ to $\hat{D}_1$ (see Equation (3.6) on page 40):

1. The asymptotic variance of $\hat{D}_2$ is smaller than the asymptotic variance of $\hat{D}_1$ (see Theorem 3.5.2).
2. In contrast to $\hat{D}_1$, under the null hypothesis the asymptotic distribution of $\hat{D}_2$ is independent of the distribution of the mileage L (see Corollary 3.5.3).
3. If the negative binomial distribution is chosen due to the validity of the relation $\sqrt{\frac{m}{2}}\left(\hat{D}_2 - 1\right) > q_{1-\alpha/2}$, then the maximum likelihood estimators of the distribution parameters exist (see Theorem 4.3.7).

3.6. Selecting the Distribution of Severity of SOLEs

The selection of a convenient model for the severity of a SOLE is motivated by the Pickands–Balkema–de Haan Theorem 2.4.4. By definition, a SOLE S_{sev} is a random variable in a peaks-over-threshold model. The codomain of S_{sev}, $\mathcal{S}$, consists of all real values above the severity threshold u_{sev} (see Definition 3.1.1). Since the threshold u_{sev} marks events with very large severities, the Pickands–Balkema–de Haan Theorem is applicable to the distribution of S_{sev}. For this reason the SOLE S_{sev}, or rather the shifted one, $S_{\text{sev}} - u_{\text{sev}}$, is chosen to be generalized Pareto distributed.

Actually, the generalized Pareto distribution consists of three different types of distributions. The types are characterized by $\xi \in \mathbb{R}_{>0}$, $\xi = 0$ and $\xi \in \mathbb{R}_{<0}$. This classification is based on the connection to the generalized extreme value distribution which goes back to the Pickands–Balkema–de Haan Theorem. The generalized extreme value distribution consists of a Fréchet type ($\xi \in \mathbb{R}_{>0}$), a Gumbel type ($\xi = 0$) and a Weibull type ($\xi \in \mathbb{R}_{<0}$) (see Section 2.4.7).

De Haan and Ferreira [HF06, p. 19] present a criterion to check whether a distribution is in the Weibull domain of attraction. According to that, a cumulative distribution function F is in the domain of attraction of the extreme value distribution $\text{GEV}(\xi)$ with $\xi \in \mathbb{R}_{<0}$ if and only if

$$x_F < \infty \qquad \text{and} \qquad \lim_{t \searrow 0} \frac{1 - F(x_F - tx)}{1 - F(x_F - t)} = x^{-\frac{1}{\xi}} \quad \forall x \in \mathbb{R}_{>0},$$

where $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$. This means that F_{sev} cannot be in the Weibull domain of attraction if

$$\mathbb{P}(S_{\text{sev}} \geq t) > 0 \qquad \forall t \in \mathbb{R}.$$

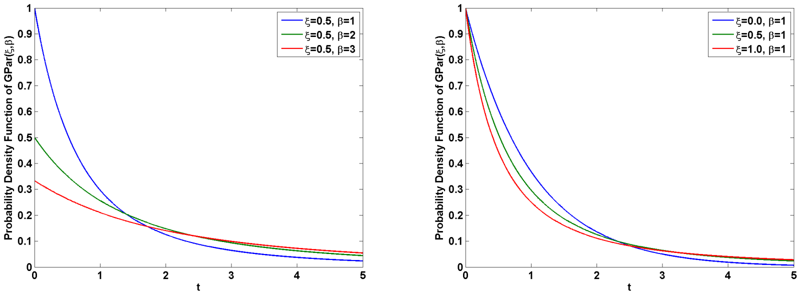
In fact, any load that occur in traffic and has an impact on the vehicle and its components is bounded from above, at last from the total energy of the universe. However, the load magnitudes which are counted here come from maneuvers that (mostly) must not damage the car or its components. The situation is similar to a plant in a high greenhouse: The size of the plant is bounded from above by the rooftop of the greenhouse. Actually, the rooftop is far higher than the biggest plant in this greenhouse ever could be. So, it can be assumed that there are no external barriers for the plant growth.

Like the plant in the large greenhouse, the loads in this experiment do not come up to their upper border. $\mathbb{P}(S_{\text{sev}} \geq t)$ shall be positive for any border $t \in \mathbb{R}$. Hence, F_{sev} cannot be in the Weibull domain of attraction, and a negative ξ can be excluded.

Summarized, the cumulative distribution function of the SOLE S_{sev} is chosen to be

$$F_{\text{sev}}(t) = F_{\text{GPar}(\xi,\beta)}(t - u_{\text{sev}}) = \mathbb{1}_{\mathbb{R}_{\geq u_{\text{sev}}}}(t) \cdot \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}(t - u_{\text{sev}})\right)^{-\frac{1}{\xi}}, & \text{if } \xi > 0, \\ 1 - \mathrm{e}^{-\frac{1}{\beta}(t - u_{\text{sev}})}, & \text{if } \xi = 0, \end{cases}$$

Figure 3.3.: *Probability density functions of generalized Pareto distribution.*



for all $t \in \mathbb{R}$, where $F_{\mathrm{GPar}(\xi,\beta)}$ is the cumulative distribution function of the generalized Pareto distribution with nonnegative shape ξ and (positive) scale β ($\xi \in \mathbb{R}_{\geq 0}$, $\beta \in \mathbb{R}_{>0}$). Figure 3.3 shows some probability density functions $\frac{\mathrm{d}F_{\mathrm{GPar}(\xi,\beta)}}{\mathrm{d}t}$ of the generalized Pareto distribution.

4. Parameter Estimation

In order to use the parametric model presented in Chapter 3, the values of the model parameters must be quantified. Using the maximum likelihood method these parameters shall be estimated based on data as described in Chapter 2. Section 4.1 recapitulates the fundamental definitions from Chapter 3 and prepares them for this chapter. Section 4.2 creates the mathematical framework for this chapter. It introduces two statistical experiments on the basis: the counting model, which excludes the observation of the maximum SOLEs, and the counting-maximum model, which also takes account the maximum SOLEs. In addition, it calculates the corresponding likelihood functions and the Fisher information. In Section 4.3 the parameter estimation of the parameters concerning the number of SOLEs is done, while Section 4.4 presents the parameter estimation of the parameters concerning the severity of a SOLE in the counting model. Finally, Section 4.5 estimates the severity parameters in the counting-maximum model.

4.1. Preliminary

According to the experimental design described in Section 2.2, in this chapter m shall be the number of vehicles providing data as mentioned ($m \in \mathbb{N}$). Each of these vehicles has its own partitioning of the detection range $\mathcal{S} = \mathbb{R}_{>u_{\text{sev}}}$,

$$A_{j1} := (u_{\text{sev}}, t_{j1}], \quad A_{j2} := (t_{j1}, t_{j2}], \quad \dots, \quad A_{jd} := (t_{j,d-1}, \infty)$$

with class limits $u_{\text{sev}} = t_{j0} < t_{j1} < \ldots < t_{j,d-1} < t_{jd} = \infty$ for all $j \in \mathbb{N}_{\leq m}$. In the following, not only the absolute load magnitude but the load relative to the severity threshold u_{sev} is required. Therefore, define for every $j \in \mathbb{N}_{\leq m}$ the relative class limits

$$s_{jk} := t_{jk} - u_{\text{sev}} \quad \forall k \in \{0, \ldots, d-1\} \qquad \text{and} \qquad s_{jd} = \infty.$$

The data consists of the observation

$$\begin{pmatrix} l_1 & z_{11} & \cdots & z_{1d} & x_1 \\ l_2 & z_{21} & \cdots & z_{2d} & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ l_m & z_{m1} & \cdots & z_{md} & x_m \end{pmatrix}$$

where, for all $j \in \mathbb{N}_{\leq m}$ and $k \in \mathbb{N}_{\leq d}$, l_j is the mileage of vehicle j, z_{jk} is number of SOLEs observed within the interval $A_{jk} = (t_{j,k-1}, t_{jk}]$ by vehicle j, and x_j is the maximum SOLE observed by vehicle j.

To be able to analyze these data, let us recall the basic random variables defined in Definition 3.1.1, Definition 3.1.2, Definition 3.1.4 and Definition 3.4.1:

- supra operating load event (SOLE) S_{sev} with codomain $\mathcal{S} = \mathbb{R}_{>u_{\text{sev}}}$ and cumulative distribution function F_{sev} ($u_{\text{sev}} \in \mathbb{R}_{>0}$),
- number of SOLEs per kilometer N_{num} with cumulative distribution function F_{num},
- number of SOLEs during l kilometers N_{num}^{*l} and number of SOLEs in range A during l kilometers $Z_{l,A}$ ($l \in \mathbb{N}$, $A \subseteq \mathcal{S}$ measurable),
- maximum SOLE per kilometer M_{sev} and maximum SOLE during l kilometers M_{sev}^{*l} ($l \in \mathbb{N}$).

In this terminology, z_{jk} is a realization of the random variable $Z_{l_j,A_{jk}}$ and x_j is a realization of $M_{\text{sev}}^{*l_j}$. The observations of distinct vehicles are assumed to be independent of each other. However, as seen in Theorem 3.3.2, the number of SOLEs in one class is not in general independent of the number of SOLEs in another range. For this reason, define the random vectors[1] $\boldsymbol{Z}_{\text{v}(1)}, \ldots, \boldsymbol{Z}_{\text{v}(m)}$ as being statistically independent and distributed as follows:

$$\boldsymbol{Z}_{\text{v}(j)} := \left(Z_{\text{v}(j)_1}, \ldots, Z_{\text{v}(j)_d}\right) \sim \left(Z_{l_j,A_{j1}}, \ldots, Z_{l_j,A_{jd}}\right) \qquad \forall j \in \mathbb{N}_{\leq m}.$$

The observation of vehicle j, $(z_{j1}, \ldots, z_{jd})$, can be interpreted as realization of $\boldsymbol{Z}_{\text{v}(j)}$ ($j \in \mathbb{N}_{\leq m}$).

In the same way define the random variables[1] $M_{\text{v}(1)}, \ldots, M_{\text{v}(m)}$ as being statistically independent and distributed according to the maximum SOLE during the corresponding mileage,

$$M_{\text{v}(j)} \sim M_{\text{sev}}^{*l_j} \qquad \forall j \in \mathbb{N}_{\leq m},$$

so that the observation x_j is a realization of $M_{\text{v}(j)}$ ($j \in \{1, \ldots, m\}$). In addition, it shall hold that $\left(M_{\text{v}(j)}, \boldsymbol{Z}_{\text{v}(j)}\right) \sim \left(M_{\text{sev}}^{*l_j}, Z_{l_j,A_{j1}}, \ldots, Z_{l_j,A_{jd}}\right)$ for all $j \in \mathbb{N}_{\leq m}$.

Finally, the total number of observed SOLEs per vehicle[1] shall be

$$N_{\text{v}(j)} := \sum_{k=1}^{d} Z_{\text{v}(j)_k} \sim N_{\text{num}}^{*l_j} \qquad \forall j \in \mathbb{N}_{\leq m}.$$

[1] The notation v(j) indicates *vehicle* j: $\boldsymbol{Z}_{\text{v}(j)}$, $M_{\text{v}(j)}$ and $N_{\text{v}(j)}$ are observations concerning vehicle j.

4.2. Statistical Experiment

The experiment on the basis can be described by two statistical models depending on whether the maximum values are part of the observation. The model that only take account the counts of the SOLEs is called **counting model**, while the model that also includes the maximum values shall be named **counting-maximum model**.

4.2.1. Counting Model

The first model only take account of the counts of SOLEs, $z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}}$. Since z is a realization of the random matrix $\big(\boldsymbol{Z}_{\mathrm{v}(j)}\big)_{1\le j\le m}$, the formal statistical experiment (according to Section 2.4.2) is

$$\mathcal{E}_{\mathrm{C}} := \Big(\mathbb{N}_0{}^{m\times d},\ \mathfrak{P}_0{}^{m\times d},\ \Big(\mathbb{P}_\vartheta\big((\boldsymbol{Z}_{\mathrm{v}(j)})_{1\le j\le m} \in \cdot\,\big)\Big)_{\vartheta\in\Theta}\Big).$$

The likelihood function of this statistical experiment is denoted by $\mathbb{L}_{\mathrm{C}}$, i. e.

$$\mathbb{L}_{\mathrm{C}}(\vartheta; z) := \mathbb{P}_\vartheta\Big(\bigcap_{j=1}^{m}\big\{\boldsymbol{Z}_{\mathrm{v}(j)} = (z_{j1},\dots,z_{jd})\big\}\Big) \qquad \forall\vartheta\in\Theta.$$

ℓ_{C} shall be the corresponding log-likelihood function,

$$\ell_{\mathrm{C}}(\vartheta; z) := \log(\mathbb{L}_{\mathrm{C}}(\vartheta; z)) \qquad \forall\vartheta\in\Theta.$$

The parameter space Θ consists of the parameters that specify the distributions of the number of SOLEs per kilometer, Θ_{num}, and of the severity of the SOLEs, Θ_{sev},

$$\Theta := \Theta_{\mathrm{num}} \times \Theta_{\mathrm{sev}}.$$

According to Section 3.6, the severity of a SOLE is assumed to be generalized Pareto distributed with nonnegative shape and positive scale. Hence, the parameter space Θ_{sev} is set to be

$$\Theta_{\mathrm{sev}} := \mathbb{R}_{\ge 0} \times \mathbb{R}_{>0}.$$

The structure of Θ_{num} depends on the chosen distribution of N_{num}. According to Section 3.5, possible models for N_{num} are the binomial, the Poisson and the negative binomial distribution. In the Poisson case the parameter is allowed to be positive and real-valued, in the negative binomial case both the exponent and the mean are positive and real-valued. In the general binomial case the number of trials is a natural number and the probability of success lies within the interval $(0, 1)$. However, the numerical results (see Section 5.7) provide a very small value for the observed number of SOLEs per kilometer, $\mathbb{E}[N_{\mathrm{num}}] \ll 1$. This fact is consistent with the assumption that SOLEs are accompanied by extreme high load magnitudes, and thus they are very rare. If N_{num} is binomially distributed with r trials and success probability q ($r \in \mathbb{N}$, $q \in (0,1)$), the expectation of N_{num} is

$$rq = \mathbb{E}[N_{\mathrm{num}}] \ll 1.$$

So, the higher the value of r, the smaller is the value of q. If q is too low, the index of dispersion of N_{num},

$$\mathbb{D}[N_{\mathrm{num}}] = \frac{\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]}{\mathbb{E}[N_{\mathrm{num}}]} = \frac{rq(1-q)}{rq} = 1 - q,$$

becomes (almost) equal to one. Consequently, N_{num} can just as well be assumed to be Poisson distributed. Therefore, the value of r should be rather low to be able to distinguish between the binomial and the Poisson approach. Moreover, the estimation of the trial parameter is not elementary. Olkin, Petkau and Zidek [OPZ81] point out that both the moment method estimator and the maximum likelihood estimator of r are not robust. The construction and evaluation of a more robust estimator is laborious anyway. In consideration of the fact that the binomial model only is a theoretical model (underdispersion is not expected, see Section 3.5.5), the search for an adequate estimator of r is disregarded. Instead, r shall be fixed as being as low as possible, which means $r = 1$. The binomial distribution with only one trial is also called **Bernoulli distribution** [JKK05]. Eventually, the parameter space of the distribution that specifies the number of SOLEs per kilometer is

$$\Theta_{\mathrm{num}} := \begin{cases} (0,1), & \text{if } N_{\mathrm{num}} \text{ is Bernoulli},\\ \mathbb{R}_{>0}, & \text{if } N_{\mathrm{num}} \text{ is Poisson},\\ \mathbb{R}_{>0} \times \mathbb{R}_{>0}, & \text{if } N_{\mathrm{num}} \text{ is negative binomial}. \end{cases}$$

All together, the parameter space is given by

$$\Theta = \Theta_{\mathrm{num}} \times \Theta_{\mathrm{sev}} = \begin{cases} (0,1) \times \big(\mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}\big), & \text{if } N_{\mathrm{num}} \text{ is Bernoulli},\\ \mathbb{R}_{>0} \times \big(\mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}\big), & \text{if } N_{\mathrm{num}} \text{ is Poisson},\\ \big(\mathbb{R}_{>0} \times \mathbb{R}_{>0}\big) \times \big(\mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}\big), & \text{if } N_{\mathrm{num}} \text{ is neg. binomial}. \end{cases}$$

Unless otherwise specified, an arbitrary element of Θ is denoted by ϑ. Each $\vartheta \in \Theta$ is put together of an element from Θ_{num} and an element from Θ_{sev}. Again, unless otherwise specified, the elements of Θ_{num} are named ν, and the elements of Θ_{sev} are denoted by ς. Furthermore, any $\varsigma \in \Theta_{\mathrm{sev}}$ consists of the shape ξ and the scale β of the generalized Pareto distribution, and any $\nu \in \Theta_{\mathrm{num}}$ stands for the mean parameter μ of the distribution of N_{num} and, if N_{num} is negative binomially distributed, for the exponent ϱ, too. Summarized, any parameter $\vartheta \in \Theta$ is build up as follows:

$$\vartheta = (\nu, \varsigma) = \begin{cases} \big(\mu, (\xi, \beta)\big), & \text{if } N_{\mathrm{num}} \sim \mathrm{Bin}(1,\mu),\ N_{\mathrm{num}} \sim \mathrm{Poi}(\mu),\\ \big((\varrho,\mu), (\xi,\beta)\big), & \text{if } N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu). \end{cases} \tag{4.1}$$

4.2.2. Counting-Maximum Model

This model also includes the observed maximum SOLEs $x = (x_j)_{1 \le j \le m}$. Here, the statistical experiment is

$$\mathcal{E}_{\mathrm{CM}} := \Big(\big(\mathbb{N}_0{}^d \times \mathcal{S}\big)^m,\ \big(\mathfrak{P}_0{}^d \otimes \mathfrak{S}\big)^m,\ \Big(\mathbb{P}_\vartheta\big((\boldsymbol{Z}_{\mathrm{v}(j)}, M_{\mathrm{v}(j)})_{1 \le j \le m} \in \cdot\ \big)\Big)_{\vartheta \in \Theta}\Big)$$

with the same parameter space Θ as in the counting model above. Hence, the likelihood function $\mathbb{L}_{\mathrm{CM}}$ of this model is

$$\mathbb{L}_{\mathrm{CM}}(\vartheta; z, x) := \tfrac{\mathrm{d}^m}{\mathrm{d}x_1 \dots \mathrm{d}x_m} \mathbb{P}_\vartheta \Big(\textstyle\bigcap_{j=1}^m \big\{ M_{\mathrm{v}(j)} \le x_j,\ \boldsymbol{Z}_{\mathrm{v}(j)} = (z_{j1}, \dots, z_{jd}) \big\} \Big)$$

for all $\vartheta \in \Theta$. The corresponding log-likelihood function is denoted by ℓ_{CM},

$$\ell_{\mathrm{CM}}(\vartheta; z, x) := \log(\mathbb{L}_{\mathrm{CM}}(\vartheta; z, x)) \qquad \forall \vartheta \in \Theta.$$

4.2.3. Log-Likelihood Functions

Both Section 3.5 and Section 3.6 establish reasonable approaches for the distribution of the total number of SOLEs, F_{num}, and for the distribution of the severity of any SOLE, F_{sev}. However, the parameters of these introduced distributions must be estimated based on data. The maximum likelihood method is an appropriate technique for this purpose. As mentioned in section 2.4.4, this method manages the circumstance that the data are interval censored, that the numbers of observed events in distinct ranges are not identically distributed and probably not statistically independent.

As the name suggests, the maximum likelihood method needs the (log-)likelihood function of the relevant statistical experiment. The following proposition identifies the log-likelihood functions of the counting model, ℓ_{C}, and of the counting-maximum model, ℓ_{CM}. Of particular note is that the parameters of F_{num} and F_{sev} can be segregated.

4.2.1 Proposition. *Suppose, (z, x) is a realization of $\big(\boldsymbol{Z}_{\mathrm{v}(j)}, M_{\mathrm{v}(j)}\big)_{1 \le j \le m}$,*

$$z = (z_{jk})_{\substack{1 \le j \le m \\ 1 \le k \le d}} \in \mathbb{N}_0{}^{m \times d} \qquad \text{and} \qquad x = (x_j)_{1 \le j \le m} \in (\{0\} \cup \mathbb{R}_{>u_{\mathrm{sev}}})^m.$$

Let $k_1, \dots, k_m$ be the classes containing the maximum SOLEs $x_1, \dots, x_m$,

$$k_j := \sum_{k=1}^d k\, \mathbb{1}_{A_{jk}}(x_j) \qquad \forall j \in \mathbb{N}_{\le m}.$$

Then, for all parameters $\vartheta = (\nu, \varsigma) \in \Theta_{\text{num}} \times \Theta_{\text{sev}}$, the log-likelihood functions of the counting model, ℓ_{C}, and of the counting-maximum model, ℓ_{CM}, are specified through

$$\ell_{\text{C}}(\vartheta; z) = \ell_{\text{num}}(\nu; z) + \ell^{\text{C}}_{\text{sev}}(\varsigma; z) + C(z),$$
$$\ell_{\text{CM}}(\vartheta; z, x) = \ell_{\text{num}}(\nu; z) + \ell^{\text{CM}}_{\text{sev}}(\varsigma; z, x) + D(z),$$

where

$$\ell_{\text{num}}(\nu; z) := \sum_{j=1}^{m} \log\Big(\mathbb{P}_\vartheta\Big(N^{*l_j}_{\text{num}} = \textstyle\sum_{k=1}^{d} z_{jk}\Big)\Big),$$

$$\ell^{\text{C}}_{\text{sev}}(\varsigma; z) := \sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk} \log\big(p_{A_{jk}}\big),$$

$$\ell^{\text{CM}}_{\text{sev}}(\varsigma; z, x) := \sum_{\substack{1\le j\le m\\ j:(k_j>0)}} \left(\sum_{k=1}^{k_j-1} z_{jk}\log\big(p_{A_{jk}}\big) + (z_{jk_j}-1)\log\Big(p_{(t_{j,k_j-1},x_j]}\Big) + \log\Big(\tfrac{\mathrm{d}F_{\text{sev}}}{\mathrm{d}x_j}(x_j)\Big)\right),$$

$$C(z) := \sum_{j=1}^{m} \log\left(\frac{(\sum_{k=1}^{d} z_{jk})!}{\prod_{k=1}^{d} z_{jk}!}\right), \qquad D(z) := C(z) + \sum_{j=1}^{m} \log\big(z_{jk_j}\big),$$

and, according to Definition 3.1.1,

$$p_{A_{jk}} = F_{\text{sev}}(t_{jk}) - F_{\text{sev}}(t_{j,k-1}) \quad \text{and} \quad p_{(t_{j,k_j-1},x_j]} = F_{\text{sev}}(x_j) - F_{\text{sev}}\big(t_{j,k_j-1}\big).$$

Proof. Since the random vectors $\boldsymbol{Z}_{\text{v}(1)}, \ldots, \boldsymbol{Z}_{\text{v}(d)}$ are statistically independent, the log-likelihood function ℓ_{C} of the counting model (see Section 4.2.1) looks like

$$\ell_{\text{C}}(\vartheta; z) = \log(\mathbb{L}_{\text{C}}(\vartheta; z)) = \sum_{j=1}^{m} \log\big(\mathbb{P}_\vartheta\big(Z_{l_j,A_{j1}} = z_{j1}, \ldots, Z_{l_j,A_{jd}} = z_{jd}\big)\big)$$

for all $\vartheta \in \Theta$. The rest follows directly from the special structure of the probability term, which can be looked up in Theorem 3.2.3.

The random vectors $\big(\boldsymbol{Z}_{\text{v}(1)}, M_{\text{v}(1)}\big), \ldots, \big(\boldsymbol{Z}_{\text{v}(m)}, M_{\text{v}(m)}\big)$ are statistically independent, too, and so the logarithm of the likelihood function $\mathbb{L}_{\text{CM}}$ of the counting-maximum model (see Section 4.2.2) is

$$\begin{aligned}\ell_{\text{CM}}(\vartheta; z, x) &= \log(\mathbb{L}_{\text{CM}}(\vartheta; z, x))\\ &= \sum_{j=1}^{m} \log\Big(\tfrac{\mathrm{d}}{\mathrm{d}x_j}\mathbb{P}_\vartheta\Big(M^{*l_j}_{\text{sev}} \le x_j,\ Z_{l_j,A_{j1}} = z_{j1}, \ldots, Z_{l_j,A_{jd}} = z_{jd}\Big)\Big)\end{aligned}$$

for all $\vartheta \in \Theta$. The results of Theorem 3.4.4 and Theorem 3.2.3 finish the proof. □

The result of the last proposition shows that the parameters ν and ς can be estimated separately from each other. The maximum likelihood estimator of ν is the maximizer of $\ell_{\text{num}}(\,\cdot\,; z)$ in both the counting model and the counting-maximum model. On the other hand, just look for the maximizer of $\ell^{\text{C}}_{\text{sev}}(\,\cdot\,; z)$ or $\ell^{\text{C}}_{\text{sev}}(\,\cdot\,; z, x)$ and the maximum likelihood estimator of ς is found.

4.2.4. Fisher Information of the Counting Model

Section 2.4.3 introduces the Fisher information of a statistical experiment and illustrates the link to the variances of estimators. Since the distributions of some estimators cannot be identified exactly, the Fisher information is an appropriate tool to approximate the variance of at least asymptotically efficient estimators.

I_{C} shall denote the Fisher information matrix in the counting model $\mathcal{E}_{\text{C}}$. According to the conventional notation of the parameters $\vartheta \in \Theta$ in Section 4.2.1 (see Equation (4.1) on page 58), with $\boldsymbol{Z} := \left(\boldsymbol{Z}_{\text{v}(j)}\right)_{1 \le j \le m}$ the Fisher information is defined by

$$I_{\text{C}}(\vartheta) = \begin{pmatrix} \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \varrho}\frac{\partial \ell_{\text{C}}}{\partial \varrho}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \varrho}\frac{\partial \ell_{\text{C}}}{\partial \mu}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \varrho}\frac{\partial \ell_{\text{C}}}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \varrho}\frac{\partial \ell_{\text{C}}}{\partial \beta}(\vartheta; \boldsymbol{Z})\right] \\ \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \mu}\frac{\partial \ell_{\text{C}}}{\partial \varrho}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \mu}\frac{\partial \ell_{\text{C}}}{\partial \mu}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \mu}\frac{\partial \ell_{\text{C}}}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \mu}\frac{\partial \ell_{\text{C}}}{\partial \beta}(\vartheta; \boldsymbol{Z})\right] \\ \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \xi}\frac{\partial \ell_{\text{C}}}{\partial \varrho}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \xi}\frac{\partial \ell_{\text{C}}}{\partial \mu}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \xi}\frac{\partial \ell_{\text{C}}}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \xi}\frac{\partial \ell_{\text{C}}}{\partial \beta}(\vartheta; \boldsymbol{Z})\right] \\ \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \beta}\frac{\partial \ell_{\text{C}}}{\partial \varrho}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \beta}\frac{\partial \ell_{\text{C}}}{\partial \mu}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \beta}\frac{\partial \ell_{\text{C}}}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] & \mathbb{E}_\vartheta\left[\frac{\partial \ell_{\text{C}}}{\partial \xi}\frac{\partial \ell_{\text{C}}}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] \end{pmatrix}$$

if N_{num} is negative binomially distributed[2]. Otherwise, just omit the first line and the first row. The following theorem introduces the Fisher information in detail. Take also note of Remark 4.2.3 following the theorem.

4.2.2 Theorem. *For all $p \in [0, 1]$, $x, y \in \mathbb{R}_{>0}$, the term $J_p(x, y)$ shall denote the (regularized) incomplete beta function [AS65, p. 263] evaluated at p, x, y,*

$$J_p(x, y) := \frac{\int_0^p t^{x-1}(1-t)^{y-1}\,\mathrm{d}t}{\int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t}.$$

Suppose, for all $i \in \{1, 2\}$, $x, t \in \mathbb{R}_{\ge 0}$, the term $\varphi_i(x, t)$ means

$$\varphi_i(x, t) := \mathbb{1}_{\{2\}}(i) + \mathbb{1}_{\{1\}}(i) \cdot \begin{cases} \frac{1}{x}\left(\log(1 + xt)\left(1 + \frac{1}{xt}\right) - 1\right), & \text{if } xt > 0, \\ \frac{t}{2}, & \text{if } xt = 0, \end{cases}$$

and it shall be

$$a_{ijk}(\xi, \beta) := s_{jk}\left(1 + \frac{\xi}{\beta}\, s_{j,k-1}\right)\varphi_i\!\left(\tfrac{\xi}{\beta}, s_{jk}\right) - s_{j,k-1}\left(1 + \frac{\xi}{\beta}\, s_{jk}\right)\varphi_i\!\left(\tfrac{\xi}{\beta}, s_{j,k-1}\right)$$

$$b_{jk}(\xi, \beta) := \begin{cases} \frac{\left(1+\frac{\xi}{\beta}s_{jk}\right)^{\frac{1}{\xi}} - \left(1+\frac{\xi}{\beta}s_{j,k-1}\right)^{\frac{1}{\xi}}}{(\beta+\xi s_{jk})^{-2}\,(\beta+\xi s_{j,k-1})^{-2}}, & \text{if } \xi > 0, \\ \left(\mathrm{e}^{\frac{1}{\beta}s_{jk}} - \mathrm{e}^{\frac{1}{\beta}s_{j,k-1}}\right)\beta^4, & \text{if } \xi = 0, \end{cases}$$

[2] at $\xi = 0$ $\frac{\partial}{\partial \xi}$ means the right partial derivative

and

$$c(\varrho,\mu) := \sum_{j=1}^{m}\sum_{n=1}^{\infty}\left(\sum_{x=0}^{n-1}\frac{1}{\varrho+\frac{x}{l_j}}\right)^2\frac{\Gamma(\varrho l_j+n)}{n!\,\Gamma(\varrho l_j)}\left(\frac{\varrho}{\varrho+\mu}\right)^{\varrho l_j}\left(\frac{\mu}{\varrho+\mu}\right)^{n}$$
$$-\sum_{j=1}^{m}\left(\sum_{x=0}^{\infty}\frac{J_{\frac{\mu}{\varrho+\mu}}(x+1,\varrho l_j)}{\varrho+\frac{x}{l_j}}\right)^2$$
$$+\left(\sum_{j=1}^{m}\left(\sum_{x=0}^{\infty}\frac{J_{\frac{\mu}{\varrho+\mu}}(x+1,\varrho l_j)}{\varrho+\frac{x}{l_j}}+l_j\log\left(\frac{\varrho}{\varrho+\mu}\right)\right)\right)^2.$$

The Fisher information matrix in the counting model $\mathcal{E}_\mathrm{C}$ is

$$I_\mathrm{C}(\vartheta) = \begin{pmatrix} I_\mathrm{num}(\nu) & 0 \\ 0 & I_\mathrm{sev}(\mu,\xi,\beta)\end{pmatrix} \qquad \forall\vartheta\in\Theta,$$

where

$$I_\mathrm{sev}(\mu,\xi,\beta) = \sum_{j=1}^{m}\sum_{k=1}^{d-1}\frac{\mu l_j}{b_{jk}(\xi,\beta)}\begin{pmatrix} a_{1jk}(\xi,\beta)^2 & a_{1jk}(\xi,\beta)\; a_{2jk}(\xi,\beta) \\ a_{1jk}(\xi,\beta)\; a_{2jk}(\xi,\beta) & a_{2jk}(\xi,\beta)^2\end{pmatrix}$$

and

$$I_\mathrm{num}(\nu) = \begin{cases} \frac{\sum_{j=1}^m l_j}{\mu(1-\mu)}, & \text{if } N_\mathrm{num}\sim \mathrm{Bin}(1,\mu),\\ \frac{\sum_{j=1}^m l_j}{\mu}, & \text{if } N_\mathrm{num}\sim \mathrm{Poi}(\mu),\\ \begin{pmatrix} c(\varrho,\mu)-\frac{\mu\sum_{j=1}^m l_j}{\varrho(\varrho+\mu)} & 0 \\ 0 & \frac{\varrho\sum_{j=1}^m l_j}{\mu(\varrho+\mu)}\end{pmatrix}, & \text{if } N_\mathrm{num}\sim \mathrm{NBin}(\varrho,\mu).\end{cases}$$

Proof. • calculation of $\mathbb{E}_\vartheta\left[\left(\frac{\partial\ell_\mathrm{C}}{\partial\mu}(\vartheta;\boldsymbol{Z})\right)^2\right]$:
Whether N_num is $\mathrm{Bin}(1,\mu)$, $\mathrm{Poi}(\mu)$ or $\mathrm{NBin}(\varrho,\mu)$ distributed, Proposition 4.2.1 and Lemma 4.3.1 yield

$$\frac{\partial\ell_\mathrm{C}}{\partial\mu}(\vartheta;\boldsymbol{Z}) = \frac{\partial\ell_\mathrm{num}}{\partial\mu}(\nu;\boldsymbol{Z}) = \frac{\mu}{\mathbb{V}\mathrm{ar}_\vartheta[N_\mathrm{num}]}\sum_{j=1}^{m}\left(\frac{N_{\mathrm{v}(j)}}{\mu}-l_j\right).$$

This means that

$$\mathbb{E}_\vartheta\left[\frac{\partial\ell_\mathrm{C}}{\partial\mu}(\vartheta;\boldsymbol{Z})\right] = \frac{\mu}{\mathbb{V}\mathrm{ar}_\vartheta[N_\mathrm{num}]}\sum_{j=1}^{m}\left(\frac{\mu l_j}{\mu}-l_j\right) = 0,$$

and hence

$$\mathbb{E}_\vartheta\left[\left(\frac{\partial\ell_\mathrm{C}}{\partial\mu}(\vartheta;\boldsymbol{Z})\right)^2\right] = \mathbb{V}\mathrm{ar}_\vartheta\left[\frac{\partial\ell_\mathrm{num}}{\partial\mu}(\nu;\boldsymbol{Z})\right] = \frac{\mathbb{V}\mathrm{ar}_\vartheta\left[\sum_{j=1}^m N_{\mathrm{v}(j)}\right]}{\mathbb{V}\mathrm{ar}_\vartheta[N_\mathrm{num}]^2} = \frac{\sum_{j=1}^m l_j}{\mathbb{V}\mathrm{ar}_\vartheta[N_\mathrm{num}]}.$$

- calculation of $\mathbb{E}_\vartheta\Big[\big(\frac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\big)^2\Big]$:

In case of $N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu)$, again, Proposition 4.2.1 and Lemma 4.3.1 provide

$$\begin{aligned}\tfrac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z}) &= \tfrac{\partial \ell_{\mathrm{num}}}{\partial \varrho}(\varrho,\mu;\boldsymbol{Z})\\ &= \sum_{j=1}^{m}\left(\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}} + l_j\log\Big(\tfrac{\varrho}{\varrho+\mu}\Big) - \frac{\mu}{\varrho+\mu}\left(\frac{N_{\mathrm{v}(j)}}{\mu}-l_j\right)\right).\end{aligned}$$

According to this, the sought-after expectation is

$$\begin{aligned}\mathbb{E}_\vartheta\Big[\big(\tfrac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\big)^2\Big] &= \mathbb{E}_\vartheta\Big[\tfrac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\Big]^2 + \mathbb{V}\mathrm{ar}_\vartheta\Big[\tfrac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\Big]\\ &= \left(\sum_{j=1}^{m}\left(\mathbb{E}_\vartheta\Big[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\Big] + l_j\log\Big(\tfrac{\varrho}{\varrho+\mu}\Big)\right)\right)^2\\ &\qquad + \sum_{j=1}^{m}\left(\mathbb{V}\mathrm{ar}_\vartheta\Big[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\Big] + \mathbb{V}\mathrm{ar}_\vartheta\Big[\tfrac{N_{\mathrm{v}(j)}}{\varrho+\mu}\Big]\right)\\ &\qquad - 2\sum_{j=1}^{m}\mathbb{C}\mathrm{ov}_\vartheta\Big[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}},\ \tfrac{N_{\mathrm{v}(j)}}{\varrho+\mu}\Big].\end{aligned}$$

The (regularized) incomplete beta function from above satisfies

$$\sum_{n=0}^{x}\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big) = 1 - J_{\frac{\mu}{\varrho+\mu}}(x+1,\varrho l_j) \qquad \forall x\in\mathbb{N}_0$$

[AS65, p. 945]. Hence, the expectation of the random sum is

$$\begin{aligned}\mathbb{E}_\vartheta\Big[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\Big] &= \sum_{n=1}^{\infty}\sum_{x=0}^{n-1}\frac{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}} = \sum_{x=0}^{\infty}\frac{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}>x\big)}{\varrho+\frac{x}{l_j}}\\ &= \sum_{x=0}^{\infty}\frac{J_{\frac{\mu}{\varrho+\mu}}(x+1,\varrho l_j)}{\varrho+\frac{x}{l_j}},\end{aligned}$$

and the variance of this term is

$$\begin{aligned}\mathbb{V}\mathrm{ar}_\vartheta\Big[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\Big] &= \mathbb{E}_\vartheta\left[\Big(\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\Big)^2\right] - \mathbb{E}_\vartheta\Big[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\Big]^2\\ &= \sum_{n=1}^{\infty}\left(\sum_{x=0}^{n-1}\frac{1}{\varrho+\frac{x}{l_j}}\right)^2\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big) - \left(\sum_{x=0}^{\infty}\frac{J_{\frac{\mu}{\varrho+\mu}}(x+1,\varrho l_j)}{\varrho+\frac{x}{l_j}}\right)^2.\end{aligned}$$

The other variance term obviously is

$$\mathbb{V}\mathrm{ar}_\vartheta\left[\tfrac{N_{\mathrm{v}(j)}}{\varrho+\mu}\right] = \frac{\mu l_j \left(1+\frac{\mu}{\varrho}\right)}{(\varrho+\mu)^2} = \frac{\mu l_j}{\varrho(\varrho+\mu)}.$$

Finally, the covariance term is also equal to ${}^{\mu l_j}/_{\varrho(\varrho+\mu)}$, because on the one hand

$$\begin{aligned}\mathbb{E}_\vartheta\left[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{N_{\mathrm{v}(j)}}{\varrho+\frac{x}{l_j}}\right] &= \sum_{n=1}^{\infty}\sum_{x=0}^{n-1}\frac{n\,\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}} \\ &= \mu l_j\sum_{x=0}^{\infty}\frac{1-\sum_{n=0}^{x}\frac{n}{\mu l_j}\,\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}},\end{aligned}$$

and on the other hand

$$\begin{aligned}\mathbb{E}_\vartheta\big[N_{\mathrm{v}(j)}\big]\,\mathbb{E}_\vartheta\left[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}\right] &= \mu l_j\sum_{n=1}^{\infty}\sum_{x=0}^{n-1}\frac{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}} \\ &= \mu l_j\sum_{x=0}^{\infty}\frac{1-\sum_{n=0}^{x}\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}},\end{aligned}$$

which means

$$\mathbb{C}\mathrm{ov}_\vartheta\left[\textstyle\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}},\frac{N_{\mathrm{v}(j)}}{\varrho+\mu}\right] = \frac{\mu l_j}{\varrho+\mu}\sum_{x=0}^{\infty}\sum_{n=0}^{x}\frac{\left(1-\frac{n}{\mu l_j}\right)\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}}.$$

All what remains to be done is to verify that the double sum on the right-hand side is equal to ${}^1/_\varrho$. For this purpose, define

$$f(x) := \frac{1}{\varrho+\frac{x}{l_j}}\sum_{n=0}^{x}\left(1-\frac{n}{\mu l_j}\right)\frac{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=x\big)} \qquad \forall x\in\mathbb{N}_0.$$

Obviously, it is $f(0)={}^1/_\varrho$, and, provided that $f(x_0)={}^1/_\varrho$ for an $x_0\in\mathbb{N}_0$,

$$\begin{aligned}f(x_0+1) &= f(x_0)\,\frac{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=x\big)\left(\varrho+\frac{x}{l_j}\right)}{\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=x+1\big)\left(\varrho+\frac{x+1}{l_j}\right)} + \frac{1}{\varrho+\frac{x+1}{l_j}}\left(1-\frac{x+1}{\mu l_j}\right) \\ &= \frac{(\varrho+\mu)(x+1)}{\varrho\mu(\varrho l_j+x+1)} + \frac{\mu l_j-(x+1)}{\mu(\varrho l_j+x+1)} = \frac{1}{\varrho}.\end{aligned}$$

So, in fact, it holds $f\equiv{}^1/_\varrho$, and this leads to

$$\frac{1}{\varrho} = \sum_{x=0}^{\infty}f(x)\,\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=x\big) = \sum_{x=0}^{\infty}\sum_{n=0}^{x}\frac{\left(1-\frac{n}{\mu l_j}\right)\mathbb{P}_\vartheta\big(N_{\mathrm{v}(j)}=n\big)}{\varrho+\frac{x}{l_j}}.$$

- calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z})\,\frac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\right]$:

When calculating the expectation of $\left(\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z})\right)^2$ above, it emerged that the expectation of $\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z})$ vanishes. This means

$$\begin{aligned}\mathbb{E}_\vartheta\left[\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z})\,\frac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\right] &= \mathbb{C}\mathrm{ov}_\vartheta\left[\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z}),\,\frac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\right]\\ &= \sum_{j=1}^{m}\mathbb{C}\mathrm{ov}_\vartheta\left[\frac{\varrho N_{\mathrm{v}(j)}}{\mu(\varrho+\mu)},\ \sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}}-\frac{N_{\mathrm{v}(j)}}{\varrho+\mu}\right]\\ &= \frac{\varrho}{\mu}\sum_{j=1}^{m}\left(\mathbb{C}\mathrm{ov}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}},\,\frac{N_{\mathrm{v}(j)}}{\varrho+\mu}\right]-\mathbb{V}\mathrm{ar}_\vartheta\left[\frac{N_{\mathrm{v}(j)}}{\varrho+\mu}\right]\right)\end{aligned}$$

When calculating the expectation of $\left(\frac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\right)^2$ above, both the covariance term and the variance term are turned out to be equal to $\mu l_j/\varrho(\varrho+\mu)$. Hence,

$$\mathbb{E}_\vartheta\left[\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z})\,\frac{\partial \ell_{\mathrm{C}}}{\partial \varrho}(\vartheta;\boldsymbol{Z})\right]=0.$$

- calculation of $\mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})\right)^2\right]$:

The partial derivative of ℓ_{C} with respect to ξ is according to Proposition 4.2.1 given by

$$\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})=\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \xi}(\varsigma;\boldsymbol{Z})=\sum_{j=1}^{m}\sum_{k=1}^{d}Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}.$$

Remember that the intervals A_{jk} span the whole severity space $\mathcal{S}$ (see Section 4.1), and so the probabilities $p_{A_{jk}}$ add up to one, $\sum_{k=1}^{d}p_{A_{jk}}=1$ for all $j\in\mathbb{N}_{\leq m}$. With this fact and the expectation value of $Z_{\mathrm{v}(j)_k}$ (see Proposition 3.2.1) it is easy to see that the expectation of $\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})$ vanishes,

$$\mathbb{E}_\vartheta\left[\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})\right]=\sum_{j=1}^{m}\sum_{k=1}^{d}\mu l_j\,\frac{\partial}{\partial \xi}p_{A_{jk}}=\mu\sum_{j=1}^{m}l_j\,\frac{\partial}{\partial \xi}\sum_{k=1}^{d}p_{A_{jk}}=\mu\sum_{j=1}^{m}l_j\,\frac{\partial}{\partial \xi}1=0.$$

Hence, the expectation of $\left(\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})\right)^2$ is equal to the variance of $\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})$,

$$\begin{aligned}&\mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})\right)^2\right]=\mathbb{V}\mathrm{ar}_\vartheta\left[\frac{\partial \ell_{\mathrm{C}}}{\partial \xi}(\vartheta;\boldsymbol{Z})\right]\\ &=\sum_{j=1}^{m}\sum_{k=1}^{d}\left(\mathbb{V}\mathrm{ar}_\vartheta\left[Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right]+2\sum_{i=k+1}^{d}\mathbb{C}\mathrm{ov}_\vartheta\left[Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}},\ Z_{\mathrm{v}(j)_i}\frac{\frac{\partial}{\partial \xi}p_{A_{ji}}}{p_{A_{ji}}}\right]\right).\end{aligned}$$

Proposition 3.2.1 and Lemma 3.3.1 provide expressions for the variance term,

$$\mathbb{V}\mathrm{ar}_\vartheta\left[Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right]=\left(\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right)^2\left(\mu l_j\,p_{A_{jk}}+l_j\,{p_{A_{jk}}}^2\left(\mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]-\mu\right)\right),$$

and for the covariance term,

$$\mathbb{C}\mathrm{ov}_\vartheta\left[Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial\xi}p_{A_{jk}}}{p_{A_{jk}}},\, Z_{\mathrm{v}(j)_i}\frac{\frac{\partial}{\partial\xi}p_{A_{ji}}}{p_{A_{ji}}}\right] = \tfrac{\partial}{\partial\xi}p_{A_{jk}}\,\tfrac{\partial}{\partial\xi}p_{A_{ji}}\,l_j\big(\mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]-\mu\big),$$

respectively. Hence, with the relation $\sum_{k=1}^d \frac{\partial}{\partial\xi}p_{A_{jk}} = 0$ it follows

$$\mathbb{E}_\vartheta\left[\left(\tfrac{\partial\ell_{\mathrm{C}}}{\partial\xi}(\vartheta;\boldsymbol{Z})\right)^2\right] = \sum_{j=1}^m\sum_{k=1}^d \mu l_j\,\frac{\left(\frac{\partial}{\partial\xi}p_{A_{jk}}\right)^2}{p_{A_{jk}}}.$$

The last expression can be transformed into a more convenient form without a derivation. For that purpose, remember that

$$\tfrac{\partial}{\partial\xi}p_{A_{jk}} = \frac{\partial}{\partial\xi}\big(F_{\mathrm{sev}}(t_{jk})-F_{\mathrm{sev}}(t_{j,k-1})\big) = \frac{\partial}{\partial\xi}\big(1-F_{\mathrm{sev}}(t_{j,k-1})\big)-\frac{\partial}{\partial\xi}\big(1-F_{\mathrm{sev}}(t_{jk})\big),$$

where F_{sev} is the cumulative probability function of a shifted generalized Pareto distribution (see Section 3.6),

$$1-F_{\mathrm{sev}}(t_{jk}) = \begin{cases}\left(1+\frac{\xi}{\beta}(t_{jk}-u)\right)^{-\frac{1}{\xi}} = \left(1+\frac{\xi}{\beta}s_{jk}\right)^{-\frac{1}{\xi}}, & \text{if } \xi>0,\\ \mathrm{e}^{-\frac{1}{\beta}(t_{jk}-u)} = \mathrm{e}^{-\frac{1}{\beta}s_{jk}}, & \text{if } \xi=0.\end{cases}$$

The derivative of this term with respect to ξ can be found in the appendix (see Lemma A.1),

$$\frac{\partial}{\partial\xi}\big(1-F_{\mathrm{sev}}(t_{jk})\big) = \underbrace{\big(1-F_{\mathrm{sev}}(t_{jk})\big)}_{f_{jk}}\ \underbrace{\frac{s_{jk}}{\beta^2+\xi\beta s_{jk}}\ \varphi_1\!\left(\tfrac{\xi}{\beta},s_{jk}\right)}_{h_{jk}}.$$

Since $f_{jd}=0$ for all $j\in\mathbb{N}_{\le m}$, with this notation it holds

$$\begin{aligned}\sum_{k=1}^d\frac{\left(\frac{\partial}{\partial\xi}p_{A_{jk}}\right)^2}{p_{A_{jk}}} &= \sum_{k=1}^{d-1}\frac{(f_{j,k-1}\,h_{j,k-1}-f_{jk}\,h_{jk})^2}{f_{j,k-1}-f_{jk}} + \frac{(f_{j,d-1}\,h_{j,d-1})^2}{f_{j,d-1}}\\ &= \sum_{k=1}^{d-1}\frac{(h_{jk}-h_{j,k-1})^2}{\frac{1}{f_{jk}}-\frac{1}{f_{j,k-1}}}\end{aligned}$$

with

$$\begin{aligned}h_{jk}-h_{j,k-1} &= \frac{a_{1jk}(\xi,\beta)}{(\beta+\xi s_{jk})\,(\beta+\xi s_{j,k-1})},\\ \frac{1}{f_{jk}}-\frac{1}{f_{j,k-1}} &= \frac{b_{jk}(\xi,\beta)}{(\beta+\xi s_{jk})^2\,(\beta+\xi s_{j,k-1})^2},\end{aligned}$$

where $a_{ijk}(\xi, \beta)$ and $b_{jk}(\xi, \beta)$ are defined in the proposition of this theorem. Eventually, one gets

$$\mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right)^2\right] = \mu \sum_{j=1}^{m} l_j \sum_{k=1}^{d} \frac{\left(\frac{\partial}{\partial \xi} p_{A_{jk}}\right)^2}{p_{A_{jk}}} = \mu \sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{1jk}(\xi, \beta)^2}{b_{jk}(\xi, \beta)}.$$

- calculation of $\mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right)^2\right]$:

The calculation of the expectation of $\left(\frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right)^2$ works very similar to the calculation of the expectation of $\left(\frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right)^2$ above. All that needs to be done is change $\frac{\partial}{\partial \xi}$ to $\frac{\partial}{\partial \beta}$ in the calculations. This results in changing φ_1 to φ_2 and, eventually, changing $a_{1jk}(\xi, \beta)$ to $a_{2jk}(\xi, \beta)$.

- calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right]$:

Also here, the calculation works very similar to the calculation of the expectation of $\left(\frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right)^2$. With this strategy one gets

$$\begin{aligned}\mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right)^2\right] &= \mu \sum_{j=1}^{m} l_j \sum_{k=1}^{d} \frac{\frac{\partial}{\partial \xi} p_{A_{jk}}\, \frac{\partial}{\partial \beta} p_{A_{jk}}}{p_{A_{jk}}} \\ &= \mu \sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{1jk}(\xi, \beta)\, a_{2jk}(\xi, \beta)}{b_{jk}(\xi, \beta)}.\end{aligned}$$

- calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right]$:

When calculating the expectation of $\left(\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})\right)^2$ above, it turns out that the expectation of $\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})$ vanishes. Hence,

$$\begin{aligned}\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] &= \mathbb{C}\mathrm{ov}_\vartheta\left[\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z}), \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] \\ &= \sum_{j=1}^{m} \sum_{k=1}^{d} \frac{\frac{\partial}{\partial \xi} p_{A_{jk}}}{p_{A_{jk}}} \frac{\mathbb{C}\mathrm{ov}_\vartheta\left[N_{\mathrm{v}(j)}, Z_{\mathrm{v}(j)_k}\right]}{\mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]}.\end{aligned}$$

By definition, $N_{\mathrm{v}(j)}$ is the sum of the random variables $Z_{\mathrm{v}(j)_1}, \ldots, Z_{\mathrm{v}(j)_d}$. Since the variances and covariances of the $Z_{\mathrm{v}(j)_k}$ are known (see Proposition 3.2.1 and Lemma 3.3.1), for a fixed k the covariance of $N_{\mathrm{v}(j)}$ and $Z_{\mathrm{v}(j)_k}$ is

$$\mathbb{C}\mathrm{ov}_\vartheta\left[N_{\mathrm{v}(j)}, Z_{\mathrm{v}(j)_k}\right] = \sum_{i=1}^{d} \mathbb{C}\mathrm{ov}_\vartheta\left[Z_{\mathrm{v}(j)_i}, Z_{\mathrm{v}(j)_k}\right] = l_j\, p_{A_{jk}} \mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}].$$

Consequently, it holds

$$\sum_{j=1}^{m} \sum_{k=1}^{d} \frac{\frac{\partial}{\partial \xi} p_{A_{jk}}}{p_{A_{jk}}} \frac{\mathbb{C}\mathrm{ov}_\vartheta\left[N_{\mathrm{v}(j)}, Z_{\mathrm{v}(j)_k}\right]}{\mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]} = \sum_{j=1}^{m} l_j \sum_{k=1}^{d} \frac{\partial}{\partial \xi} p_{A_{jk}} = \sum_{j=1}^{m} l_j\, \frac{\partial}{\partial \xi} 1 = 0.$$

- calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right]$:

Just change the operator $\frac{\partial}{\partial \xi}$ to $\frac{\partial}{\partial \beta}$ in the calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right]$ above.

- calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \varrho}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right]$:

When calculating the expectation of $\left(\frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right)^2$ above, it emerged that the expectation of $\frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})$ is equal to 0. This provides the equality of

$$\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \varrho}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right] = \mathbb{C}\mathrm{ov}_\vartheta\left[\frac{\partial \ell_C}{\partial \varrho}(\vartheta; \boldsymbol{Z}),\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})\right]$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{d}\left(\mathbb{C}\mathrm{ov}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}},\; Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right] - \mathbb{C}\mathrm{ov}_\vartheta\left[\frac{N_{\mathrm{v}(j)}}{\varrho+\mu},\; Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right]\right).$$

When calculating the expectation of $\frac{\partial \ell_C}{\partial \mu}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})$ above, it can be seen that the covariance of $N_{\mathrm{v}(j)}$ and $Z_{\mathrm{v}(j)_k}$ is equal to $l_j\, p_{A_{jk}} \mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]$. Therefore, the second covariance expression vanishes,

$$\sum_{k=1}^{d}\mathbb{C}\mathrm{ov}_\vartheta\left[\frac{N_{\mathrm{v}(j)}}{\varrho+\mu},\; Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right] = \frac{l_j\, \mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]}{\varrho+\mu}\,\frac{\partial}{\partial \xi}\sum_{k=1}^{d}p_{A_{jk}} = 0.$$

Also the first covariance terms add up to 0. To establish this, remember that $Z_{\mathrm{v}(j)_k}$ given $N_{\mathrm{v}(j)}$ is binomially distributed with $N_{\mathrm{v}(j)}$ trials and success probability $p_{A_{jk}}$ (see proof of Proposition 3.2.1). Therefore

$$\mathbb{E}_\vartheta\left[Z_{\mathrm{v}(j)_k}\right] = \mathbb{E}_\vartheta\left[\mathbb{E}_\vartheta\left[\,Z_{\mathrm{v}(j)_k}\,\middle|\,N_{\mathrm{v}(j)}\right]\right] = p_{A_{jk}}\mathbb{E}_\vartheta\left[N_{\mathrm{v}(j)}\right]$$

and

$$\mathbb{E}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{Z_{\mathrm{v}(j)_k}}{\varrho+\frac{x}{l_j}}\right] = \mathbb{E}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{\mathbb{E}_\vartheta\left[\,Z_{\mathrm{v}(j)_k}\,\middle|\,N_{\mathrm{v}(j)}\right]}{\varrho+\frac{x}{l_j}}\right] = p_{A_{jk}}\mathbb{E}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{N_{\mathrm{v}(j)}}{\varrho+\frac{x}{l_j}}\right].$$

All this results in

$$\sum_{k=1}^{d}\mathbb{C}\mathrm{ov}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}},\, Z_{\mathrm{v}(j)_k}\frac{\frac{\partial}{\partial \xi}p_{A_{jk}}}{p_{A_{jk}}}\right] = \mathbb{C}\mathrm{ov}_\vartheta\left[\sum_{x=0}^{N_{\mathrm{v}(j)}-1}\frac{1}{\varrho+\frac{x}{l_j}},\, N_{\mathrm{v}(j)}\right]\frac{\partial}{\partial \xi}\sum_{k=1}^{d}p_{A_{jk}}$$

which is equal to 0.

- calculation of $\mathbb{E}_\vartheta\left[\frac{\partial \ell_C}{\partial \varrho}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \beta}(\vartheta; \boldsymbol{Z})\right]$:

Very similar to the evaluation of the expectation of $\frac{\partial \ell_C}{\partial \varrho}(\vartheta; \boldsymbol{Z})\, \frac{\partial \ell_C}{\partial \xi}(\vartheta; \boldsymbol{Z})$, also here the sought-after expectation value is 0. □

The calculation of the Fisher information matrix in the last theorem turns out that the parameter group concerning the number of SOLEs, $\nu = ((\varrho), \mu)$, is orthogonal to the parameter group concerning the severity of a SOLE, $\varsigma = (\xi, \beta)$, i. e. the respective entries of the Fisher information matrix are 0 [CR87]. This is consistent with the fact that ν and ς can be estimated separately as shown in Proposition 4.2.1.

Besides, in the negative binomial case, the mean μ and the exponent ϱ are orthogonal, too. Consequently, if $(\hat{\varrho}, \hat{\mu})$ is an asymptotically efficient estimator of (ϱ, μ) in the sense of Section 2.4.3, then $\hat{\varrho}$ and $\hat{\mu}$ are asymptotically independent, because jointly normally distributed and uncorrelated random variables are statistically independent [Als05, p. 141].

The following remark gives an idea of how to simplify the term $c(\varrho, \mu)$ in the last theorem.

4.2.3 Remark. Numerical calculations support the assumption that the following relation holds:

$$\log(1-p) + \sum_{x=0}^{\infty} \frac{J_p(x+1,y)}{y+x} = 0 \qquad \forall p \in (0,1),\ \forall y \in \mathbb{R}_{>0}, \tag{4.2}$$

where $J_p(x+1, y)$ denotes the (regularized) incomplete beta function evaluated at p, $x+1$ and y (see Theorem 4.2.2). If this relation is true, the term $c(\varrho, \mu)$ in Theorem 4.2.2 can be written as

$$c(\varrho,\mu) = \sum_{j=1}^{m} \sum_{n=1}^{\infty} \left(\sum_{x=0}^{n-1} \frac{1}{\varrho + \frac{x}{l_j}} \right)^2 \frac{\Gamma(\varrho l_j + n)}{n!\, \Gamma(\varrho l_j)} \left(\frac{\varrho}{\varrho+\mu} \right)^{\varrho l_j} \left(\frac{\mu}{\varrho+\mu} \right)^n - \log\left(1 + \tfrac{\mu}{\varrho}\right)^2 \sum_{j=1}^{m} {l_j}^2.$$

In addition, one may use the relation

$$\sum_{x=0}^{n-1} \frac{1}{\varrho + \frac{x}{l_j}} = l_j \sum_{x=0}^{n-1} \frac{1}{\varrho l_j + x} = l_j \Big(\psi(\varrho l_j + n) - \psi(\varrho l_j) \Big)$$

with digamma function ψ, $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log(\Gamma(x))$ [AS65, p. 258].

At least for all $y \in \mathbb{N}$, Equation (4.2) can be verified by dint of the relation

$$\begin{aligned} \sum_{x=0}^{\infty} \frac{J_p(x+1,y)}{y+x} &= \sum_{x=0}^{\infty} \frac{1}{y+x} \sum_{n=0}^{y-1} \binom{x+y}{n} p^{x+y-n} (1-p)^n \\ &= \sum_{x=0}^{\infty} \left(\frac{p^{x+y}}{x+y} + \sum_{n=1}^{y-1} \binom{x+y}{n} \frac{p^{x+y-n}(1-p)^n}{x+y} \right) \end{aligned}$$

for all $p \in (0,1)$ and $y \in \mathbb{N}$ [AS65, p. 944]. By comparison of the coefficients of the p^x it can be shown that

$$\sum_{x=0}^{\infty}\sum_{n=1}^{y-1}\binom{x+y}{n}\frac{p^{x+y-n}\,(1-p)^n}{x+y}=\sum_{x=1}^{y-1}\frac{p^x}{x}\qquad \forall p\in(0,1),\ \forall y\in\mathbb{N}.$$

Therefore, it follows for all $p \in (0,1)$ and $y \in \mathbb{N}$ that

$$\sum_{x=0}^{\infty}\frac{J_p(x+1,y)}{y+x}=\sum_{x=0}^{\infty}\frac{p^{x+y}}{x+y}+\sum_{x=1}^{y-1}\frac{p^x}{x}=\sum_{x=1}^{\infty}\frac{p^x}{x}=-\log(1-p)\,,$$

where the last equation sign holds since $\sum_{x=1}^{\infty} -p^x/x$ is a series expansion of $\log(1-p)$ [AS65, p. 68].

4.3. Estimating the Number of SOLEs per Kilometer

In Section 4.2.3 it is argued that the distribution of the number of SOLEs per kilometer can be estimated without considering their severities. According to Proposition 4.2.1, the maximum likelihood estimator of the parameter of F_{num}, ν, is the maximizer of the function

$$\ell_{\text{num}}(\nu;z)=\sum_{j=1}^{m}\log\Big(\mathbb{P}_\vartheta\Big(N_{\text{num}}^{*l_j}=\textstyle\sum_{k=1}^{d} z_{jk}\Big)\Big)\,.$$

In case of a binomial, Poisson and negative binomially distributed N_{num}, the variate $N_{\text{num}}^{*l_j}$ is binomial, Poisson and negative binomial, too,

$$\begin{aligned} & N_{\text{num}} \sim \text{Bin}(1,\mu),\ \text{Poi}(\mu),\ \text{NBin}(\varrho,\mu) \\ \Rightarrow\quad & N_{\text{num}}^{*l_j} \sim \text{Bin}(l_j,\mu),\ \text{Poi}(\mu l_j),\ \text{NBin}(\varrho l_j,\mu l_j), \end{aligned}$$

because $N_{\text{num}}^{*l_j}$ is just the sum of l_j statistically independent random variables which are all distributed according to N_{num} (see Definition 3.1.4). In these three cases it is easy to write out the function ℓ_{num}. The next lemma collects the particular versions of ℓ_{num}.

4.3.1 Lemma. *Let $z=(z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in {\mathbb{N}_0}^{m\times d}$ be a realization of $\big(\boldsymbol{Z}_{\text{v}(j)}\big)_{1\le j\le m}$. Apply the shortcuts*

$$n_j:=\sum_{k=1}^{d} z_{jk},\quad n:=\sum_{j=1}^{m} n_j\qquad\text{and}\qquad l:=\sum_{j=1}^{m} l_j.$$

1. *If N_{num} is Bernoulli distributed with success probability μ ($\mu \in (0,1)$), then*
$$\ell_{\text{num}}(\mu; z) = \sum_{j=1}^{m} \log\left(\binom{l_j}{n_j}\right) + n\log(\mu) + (l-n)\log(1-\mu)\,,$$
and therefore
$$\tfrac{\partial \ell_{\text{num}}}{\partial \mu}(\mu; z) = \frac{1}{1-\mu}\left(\frac{n}{\mu} - l\right).$$

2. *If N_{num} is Poisson distributed with mean μ ($\mu \in \mathbb{R}_{>0}$), then*
$$\ell_{\text{num}}(\mu; z) = -\mu l + n\log(\mu) + \sum_{j=1}^{m} \log\left(\tfrac{{l_j}^{n_j}}{n_j!}\right),$$
and therefore
$$\tfrac{\partial \ell_{\text{num}}}{\partial \mu}(\mu; z) = \frac{n}{\mu} - l.$$

3. *If N_{num} is negative binomially distributed with exponent ϱ and mean μ ($\varrho, \mu \in \mathbb{R}_{>0}$), then*
$$\ell_{\text{num}}(\varrho, \mu; z) = \sum_{j=1}^{m} \log\left(\tfrac{\Gamma(\varrho l_j + n_j)}{n_j!\ \Gamma(\varrho l_j)}\right) + \varrho l \log(\varrho) + n\log(\mu) - (\varrho l + n)\log(\varrho + \mu)\,,$$
and therefore
$$\begin{aligned}\tfrac{\partial \ell_{\text{num}}}{\partial \varrho}(\varrho, \mu; z) &= \sum_{j=1}^{m}\sum_{x=0}^{n_j-1} \frac{1}{\varrho + \frac{x}{l_j}} + l\log\left(\tfrac{\varrho}{\varrho+\mu}\right) - \frac{\mu}{\varrho+\mu}\left(\frac{n}{\mu} - l\right)\\ &= \sum_{j=1}^{m} l_j\big(\psi(\varrho l_j + n_j) - \psi(\varrho l_j)\big) + l\log\left(\tfrac{\varrho}{\varrho+\mu}\right) - \frac{\mu}{\varrho+\mu}\left(\frac{n}{\mu} - l\right)\end{aligned}$$
$$\tfrac{\partial \ell_{\text{num}}}{\partial \mu}(\varrho, \mu; z) = \frac{\varrho}{\varrho+\mu}\left(\frac{n}{\mu} - l\right)$$
with digamma function ψ, $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\log(\Gamma(x))$, [AS65, p. 258].

Proof. Just substitute the probability mass functions of binomial, Poisson and negative binomial distribution as given in Definition 2.4.2 into the definition of ℓ_{num} from Theorem 4.2.1.

In case of the negative binomial distribution consider that [AS65, p. 258]
$$\psi(\varrho l_j + n_j) = \psi(\varrho l_j) + \sum_{x=0}^{n_j-1} \frac{1}{\varrho l_j + x} \qquad \forall j \in \mathbb{N}_{\leq m}.$$
□

In the following two sections 4.3.1 and 4.3.2 let us concretize form and characteristics of the maximum likelihood estimators of μ and ϱ.

4.3.1. Estimating the Mean μ of the Number of SOLEs per Kilometer

Remember, the number of SOLEs per kilomter N_{num} is assumed to be Bernoulli, Poisson or negative binomially distributed (see Section 3.5 and Section 4.2.1), $N_{\text{num}} \sim \text{Bin}(1,\mu)$, $\text{Poi}(\mu)$, $\text{NBin}(\varrho,\mu)$, with $\mu, \varrho \in \mathbb{R}_{>0}$, except in the binomial case where $\mu \in (0,1)$. In all three cases the parametrization is chosen such that μ denotes the expectation of N_{num}. The intuition suggests that the total number of observed SOLEs divided by the total number of kilometers is an adequate estimator of the mean number of possible events per kilometer. The following theorem confirms this intuition.

4.3.2 Theorem. *Let $z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in {\mathbb{N}_0}^{m\times d}$ be a realization of $\left(\boldsymbol{Z}_{\text{v}(j)}\right)_{1\le j\le m}$ with $z \not\equiv 0$. Suppose, N_{num} is either Bernoulli, Poisson or negative binomially distributed with unknown mean μ, more precisely*

$$N_{\text{num}} \sim \text{Bin}(1,\mu) \qquad \text{or} \qquad N_{\text{num}} \sim \text{Poi}(\mu) \qquad \text{or} \qquad N_{\text{num}} \sim \text{NBin}(\varrho,\mu).$$

Then, the maximum likelihood estimator $\hat{\mu}_m$ of μ based on z exists in case of the Poisson and negative binomial distribution, and it exists in case of the binomial distribution if and only if $\sum_{j=1}^m l_j > \sum_{j=1}^m \sum_{k=1}^d z_{jk}$. If the maximum likelihood estimator of μ based on z exists, it is given by

$$\hat{\mu}_m(z) = \frac{\sum_{j=1}^m \sum_{k=1}^d z_{jk}}{\sum_{j=1}^m l_j}.$$

Proof. Since Lemma 4.3.1 holds, the maximum likelihood estimator of μ must satisfy the following equivalent equations:

$$\left(\frac{\sum_{j=1}^m \sum_{k=1}^d z_{jk}}{\hat{\mu}_m} - \sum_{j=1}^m l_j\right) = 0 \qquad \Leftrightarrow \qquad \hat{\mu}_m = \frac{\sum_{j=1}^m \sum_{k=1}^d z_{jk}}{\sum_{j=1}^m l_j}.$$

Furthermore, the sign of the derivative of ℓ_{num} with respect to the mean μ changes from positive to negative at this point. Hence, $\hat{\mu}_m$ as given in the proposition of this theorem must be the maximum likelihood estimator of μ.

If the number of observed SOLEs, $\sum_{j=1}^m \sum_{k=1}^d z_{jk}$, exceeds the number of kilometers, $\sum_{j=1}^m l_j$, the equation above does not have a solution satisfying $\hat{\mu}_m \in (0,1)$. Hence, in this situation there does not exist a maximum likelihood estimator in the binomial case. □

The last Theorem 4.3.1 reveals the possibility that the maximum likelihood estimator of μ does not exist if N_{num} is Bernoulli distributed. However, this is only a theoretical problem. In Section 4.2.1 it is argued that the Bernoulli approach is chosen, because the mean number of SOLEs per kilometer is expected to be always by many orders of magnitude below 1. Practically, the number of observed SOLEs will not exceed the number of kilometers. Section 5.7 gives a quantitative confirmation of this assumption.

Since the structure of the maximum likelihood estimator $\hat{\mu}_m$ of μ is quite simple, it is easy to verify some characteristics of it. The next theorem provides that $\hat{\mu}_m$ is a consistent, efficient and uniformly minimum-variance unbiased estimator.

4.3.3 Theorem. *Let the situation be as in Theorem 4.3.2 with the conventional notation of the parameters according to Section 4.2.1 (see Equation (4.1) on page 58) and* $\boldsymbol{Z} := \left(\boldsymbol{Z}_{\mathrm{v}(j)}\right)_{1\le j\le m}$*, then it holds:*

1. $\hat{\mu}_m$ *is an unbiased estimator of* μ*,*
$$\mathbb{E}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] = \mu \qquad \forall\vartheta\in\Theta.$$
2. *The variance of* $\hat{\mu}_m(\boldsymbol{Z})$ *is given by*
$$\mathbb{V}\mathrm{ar}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] = \frac{\mathbb{V}\mathrm{ar}_\vartheta[N_{\mathrm{num}}]}{\sum_{j=1}^m l_j} = \frac{\mu(1+w_\vartheta\,\mu)}{\sum_{j=1}^m l_j} \qquad \forall\vartheta\in\Theta,$$
where
$$w_\vartheta := \begin{cases} -1, & \text{if } N_{\mathrm{num}} \sim \mathrm{Bin}(1,\mu),\\ 0, & \text{if } N_{\mathrm{num}} \sim \mathrm{Poi}(\mu),\\ \frac{1}{\varrho}, & \text{if } N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu).\end{cases}$$
3. $\hat{\mu}_m$ *is a consistent estimator of* μ*,*
$$\hat{\mu}_m(\boldsymbol{Z}) \xrightarrow{P} \mu \qquad \text{for } m\to\infty \quad \forall\vartheta\in\Theta.$$
4. $\hat{\mu}_m$ *is an uniformly minimum-variance unbiased estimator (UMVUE) of* μ*, i. e. it is unbiased and for any other unbiased estimator* $\tilde{\mu}$ *of* μ *it holds*
$$\mathbb{V}\mathrm{ar}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] \le \mathbb{V}\mathrm{ar}_\vartheta[\tilde{\mu}(\boldsymbol{Z})] \qquad \forall\vartheta\in\Theta.$$
5. $\hat{\mu}_m$ *is an efficient estimator of* μ*, i. e. it is unbiased and it achieves equality on the information inequality (see Section 4.2.4) in the reduced model where all parameters are kept constant but the mean* μ*.*

Proof. **1./2.:** For all $j\in\mathbb{N}_{\le m}$, the sum $N_{\mathrm{v}(j)} = \sum_{k=1}^d Z_{\mathrm{v}(j)_k}$ is distributed according to the total number of events during l_j kilometers, $N_{\mathrm{v}(j)} \sim N_{\mathrm{num}}^{*l_j}$. Since the observations of different vehicles are mutually statistically independent, it follows

$$\sum_{j=1}^m\sum_{k=1}^d Z_{\mathrm{v}(j)_k} \sim \begin{cases} \mathrm{Bin}\left(\sum_{j=1}^m l_j, \mu\right), & \text{if } N_{\mathrm{num}} \sim \mathrm{Bin}(1,\mu),\\ \mathrm{Poi}\left(\mu\sum_{j=1}^m l_j\right), & \text{if } N_{\mathrm{num}} \sim \mathrm{Poi}(\mu),\\ \mathrm{NBin}\left(\varrho\sum_{j=1}^m l_j, \mu\sum_{j=1}^m l_j\right), & \text{if } N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu).\end{cases}$$

Hence,

$$\mathbb{E}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] = \frac{\mathbb{E}_\vartheta\left[\sum_{j=1}^m \sum_{k=1}^d Z_{\mathrm{v}(j)_k}\right]}{\sum_{j=1}^m l_j} = \frac{\mu \sum_{j=1}^m l_j}{\sum_{j=1}^m l_j} = \mu,$$

and

$$\mathbb{V}\mathrm{ar}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] = \frac{\mathbb{V}\mathrm{ar}_\vartheta\left[\sum_{j=1}^m \sum_{k=1}^d Z_{\mathrm{v}(j)_k}\right]}{\left(\sum_{j=1}^m l_j\right)^2} = \frac{\mu(1+w_\vartheta\,\mu)}{\sum_{j=1}^m l_j}.$$

3.: Chebyshev's inequality [UC11] in combination with the result of the second statement in this theorem ensures that for all $\varepsilon > 0$ it holds

$$\mathbb{P}_\vartheta\left(\left|\hat{\mu}_m(\boldsymbol{Z}) - \mu\right| > \varepsilon\right) \leq \frac{\mathbb{V}\mathrm{ar}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})]}{\varepsilon^2} = \frac{\mu(1+w_\vartheta\,\mu)}{\varepsilon^2 \sum_{j=1}^m l_j} \xrightarrow{m\to\infty} 0.$$

4.: With the notation $n_j := \sum_{k=1}^d z_{jk}$ for all $j \in \mathbb{N}_{\leq m}$ and the statistical independence of the $N_{\mathrm{v}(j)}$ one gets

$$\begin{aligned}\mathbb{P}_\vartheta\left(\bigcap_{j=1}^m \{N_{\mathrm{v}(j)} = n_j\}\right) &= \prod_{j=1}^m \mathbb{P}_\vartheta\left(N_{\mathrm{v}(j)} = n_j\right) \\ &= B(\nu)\ \mathrm{e}^{Q(\nu)\,T(n_1,\ldots,n_m)}\ h(n_1,\ldots,n_m,(\varrho)),\end{aligned}$$

where $T(n_1,\ldots,n_m) := \sum_{j=1}^m n_j$ and

$$B(\nu) := \begin{cases} (1-\mu)^{\sum_{j=1}^m l_j}, \\ \mathrm{e}^{-\mu\sum_{j=1}^m l_j}, \\ \left(\frac{\varrho}{\varrho+\mu}\right)^{\varrho\sum_{j=1}^m l_j}, \end{cases} \qquad Q(\nu) := \begin{cases} \log\left(\frac{\mu}{1-\mu}\right), & \text{if } F_{\mathrm{num}} \sim \mathrm{Bin}(1,\mu), \\ \log(\mu), & \text{if } F_{\mathrm{num}} \sim \mathrm{Poi}(\mu), \\ \log\left(\frac{\mu}{\varrho+\mu}\right), & \text{if } F_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu), \end{cases}$$

$$h(n_1,\ldots,n_m,(\varrho)) := \begin{cases} \prod_{j=1}^m \binom{l_j}{n_j}, & \text{if } F_{\mathrm{num}} \sim \mathrm{Bin}(1,\mu), \\ \prod_{j=1}^m \frac{l_j^{\,n_j}}{n_j!}, & \text{if } F_{\mathrm{num}} \sim \mathrm{Poi}(\mu), \\ \prod_{j=1}^m \frac{\Gamma\left(\varrho l_j + n_j\right)}{n_j!\ \Gamma(\varrho l_j)}, & \text{if } F_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu). \end{cases}$$

As can be seen in Section 4.2.4, even if N_{num} is assumed to be negative binomially distributed, μ and ϱ are orthogonal, and so the parameters can be estimated separately [CR87]. Therefore, ϱ can be treated as known and constant. Hence, the underlying family of distributions is a one-dimensional exponential family, which ensures that T is a complete and sufficient statistic [LC98, pp. 39–42]. With the chosen notation, $\hat{\mu}_m(\boldsymbol{Z})$ is a function of T,

$$\hat{\mu}_m(\boldsymbol{Z}) = \frac{1}{\sum_{j=1}^m l_j}\ T\left(\left(\sum_{k=1}^d Z_{\mathrm{v}(j)_k}\right)_{1\leq j\leq m}\right),$$

and $\hat{\mu}_m(\boldsymbol{Z})$ is indeed an UMVUE of μ [LC98, p. 88].

5.: Theorem 4.2.2 provides that the Fisher information with regard to the parameter μ is

$$\mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_{\mathrm{C}}}{\partial \mu}(\vartheta;\boldsymbol{Z})\right)^2\right] = \mathbb{E}_\vartheta\left[\left(\frac{\partial \ell_{\mathrm{CM}}}{\partial \mu}(\vartheta;\boldsymbol{Z})\right)^2\right] = \frac{\sum_{j=1}^m l_j}{\mu(1+w_\vartheta\,\mu)}.$$

But the term on the right-hand side is exactly the inverse of the variance of $\hat{\mu}_m$ as can be seen in the second statement of this theorem. ☐

In other words, the last theorem teaches that the maximum likelihood estimator $\hat{\mu}_m$ estimates the true mean μ correctly on average, that the variance of $\hat{\mu}_m$ is inversely proportional to the total mileage, that $\hat{\mu}_m$ tends to the correct mean μ in probability, and that there is no other unbiased estimator with an uniformly smaller variance. In this sense $\hat{\mu}_m$ is an optimal estimator.

For quantifying the accuracy of estimate by $\hat{\mu}_m$, a confidence interval of μ based on $\hat{\mu}_m$ is needed. An interval is called **confidence interval of parameter μ with level** $(1-\alpha)$ $(\alpha \in (0,1))$ if it contains the real value μ with probability $(1-\alpha)$ [UC11]. Since the sample size m is rather large, it is sufficient to calculate an approximate confidence interval. As can be seen in the proof of the next theorem, the central limit theorem [LC98, p. 58] ensures that $\hat{\mu}_m$ is approximately normally distributed. This fact yields an approximate confidence interval.

4.3.4 Theorem. *Let the situation be as in Theorem 4.3.2 with the conventional notation of the parameters according to Section 4.2.1 (see Equation (4.1) on page 58) and* $\boldsymbol{Z} := \left(\boldsymbol{Z}_{\mathrm{v}(j)}\right)_{1\le j\le m}$. *For a given* $\alpha \in (0,1)$, *let* $q_{1-\alpha/2}$ *be the* $\left(1-\frac{\alpha}{2}\right)100\,\%$ *quantile of the standard normal distribution. Then, for all* $\vartheta \in \Theta$ *and* $\alpha \in (0,1)$, *the maximum likelihood estimator* $\hat{\mu}_m$ *of* μ *satisfies*

$$\lim_{m\to\infty} \mathbb{P}_\vartheta\left(\left|\hat{\mu}_m(\boldsymbol{Z})-\mu\right| \le q_{1-\alpha/2}\sqrt{\frac{\hat{\mu}_m(\boldsymbol{Z})\,(1+\hat{w}_m(\boldsymbol{Z})\,\hat{\mu}_m(\boldsymbol{Z}))}{\sum_{j=1}^m l_j}}\right) = 1-\alpha,$$

where

$$\hat{w}_m := \begin{cases} -1, & \text{if } N_{\mathrm{num}} \sim \mathrm{Bin}(1,\mu),\\ 0, & \text{if } N_{\mathrm{num}} \sim \mathrm{Poi}(\mu),\\ \frac{1}{\hat{\varrho}_m}, & \text{if } N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu), \end{cases}$$

and $\hat{\varrho}_m$ *denotes the maximum likelihood estimator of* ϱ *(see Theorem 4.3.7).*

Proof. For the following triangular array construction

$$\begin{array}{cccccccc}
N_{11} & N_{12} & \dots & N_{1l_1} & & & & \\
N_{21} & N_{22} & \dots & N_{2l_1} & N_{2,l_1+1} & \dots & N_{2,l_1+l_2} & \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \ddots \\
N_{m1} & N_{m2} & \dots & N_{ml_1} & N_{m,l_1+1} & & \dots & & N_{m,\sum_{j=1}^m l_j} \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \ddots
\end{array}$$

with statistically independent random variables $(N_{ji})_{j,i\in\mathbb{N}}$, which are all distributed according to N_{num}, the Central Limit Theorem holds [Als05, p. 194],

$$\frac{\sum_{i=1}^{\sum_{j=1}^m l_j} N_{mi} - \mathbb{E}_\vartheta\left[\sum_{i=1}^{\sum_{j=1}^m l_j} N_{mi}\right]}{\sqrt{\mathbb{V}\mathrm{ar}_\vartheta\left[\sum_{i=1}^{\sum_{j=1}^m l_j} N_{mi}\right]}} \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{for } m \to \infty.$$

Moreover, Theorem 4.3.2 verifies that

$$\hat{\mu}_m(\boldsymbol{Z}) = \frac{\sum_{j=1}^m N_{\mathrm{v}(j)}}{\sum_{j=1}^m l_j} \sim \frac{\sum_{i=1}^{\sum_{j=1}^m l_j} N_{mi}}{\sum_{j=1}^m l_j},$$

and therefore it holds

$$\lim_{m\to\infty} \mathbb{P}_\vartheta\left(-q_{1-\alpha/2} \le \frac{\hat{\mu}_m(\boldsymbol{Z}) - \mathbb{E}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})]}{\sqrt{\mathbb{V}\mathrm{ar}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})]}} \le q_{1-\alpha/2}\right) = 1 - \alpha.$$

Theorem 4.3.3 yields

$$\mathbb{E}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] = \mu \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}_\vartheta[\hat{\mu}_m(\boldsymbol{Z})] = \frac{\mu(1 + w_\vartheta\,\mu)}{\sum_{j=1}^m l_j}.$$

Theorem 4.3.3 and the consistency of $\hat{\varrho}_m$ (see Section 4.3.2) ensure that both $\hat{\mu}_m(\boldsymbol{Z})$ and $\hat{\varrho}_m(\boldsymbol{Z})$ tend in probability to the real parameters μ and ϱ respectively, which causes [Als05, p. 170]

$$\sqrt{\frac{\mu(1 + w_\vartheta\,\mu)}{\hat{\mu}_m(\boldsymbol{Z})\left(1 + \hat{w}_m(\boldsymbol{Z})\,\hat{\mu}_m(\boldsymbol{Z})\right)}} \xrightarrow{P} 1 \qquad \text{for } m \to \infty.$$

With all these facts and Slutsky's Theorem [Slu25, Cra62, pp. 254–255] one gets

$$\lim_{m\to\infty} \mathbb{P}_\vartheta\left(-q_{1-\alpha/2} \le \frac{\hat{\mu}_m(\boldsymbol{Z}) - \mu}{\sqrt{\frac{\mu(1+w_\vartheta\,\mu)}{\sum_{j=1}^m l_j}}}\ \sqrt{\frac{\mu(1+w_\vartheta\,\mu)}{\hat{\mu}_m(\boldsymbol{Z})\,(1+\hat{w}_m(\boldsymbol{Z})\,\hat{\mu}_m(\boldsymbol{Z}))}} \le q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

This is just a transcription of the proposition. □

The (approximate) confidence interval of μ introduced in Theorem 4.3.4 only depends on the estimates $\hat{\mu}_m$ and, in case of the negative binomial distribution, $\hat{\varrho}_m$, and on the total mileage $\sum_{j=1}^m l_j$. The interval becomes smaller if the mileage increases. Since the accuracy of estimate of μ can be quantified by the length of its confidence interval, this fact generates the obvious question how long the experiment must run so the confidence interval becomes sufficient small.

In the preliminary stage of the experiment this question cannot be answered, because, in the end, the number of observed SOLEs is responsible for the length of the confidence interval. To realize this define the actual version of the approximate confidence interval from Theorem 4.3.4,

$$C_\mu(\alpha, z) := \left[\hat{\mu}_m(z) - q_{1-\alpha/2}\sqrt{\frac{\hat{\mu}_m(z)\left(1 + \hat{w}_m(z)\,\hat{\mu}_m(z)\right)}{\sum_{j=1}^m l_j}},\right. \\ \left.\hat{\mu}_m(z) + q_{1-\alpha/2}\sqrt{\frac{\hat{\mu}_m(z)\left(1 + \hat{w}_m(z)\,\hat{\mu}_m(z)\right)}{\sum_{j=1}^m l_j}}\right], \qquad (4.3)$$

where $z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}}$ is a realization of $\boldsymbol{Z} = (\boldsymbol{Z}_{\mathrm{v}(j)})_{1\le j\le m}$. Since $C_\mu(\alpha, z)$ is centered around $\hat{\mu}_m(z)$ and the tolerable interval length should depend on the magnitude of μ, it make sense to express the length of $C_\mu(\alpha, z)$ by the relative deviation from the estimate $\hat{\mu}_m(z)$. In other words, find a $\delta \in \mathbb{R}_{>0}$ such that

$$C_\mu(\alpha, z) = \Big[(1-\delta)\,\hat{\mu}_m(z),\ (1+\delta)\,\hat{\mu}_m(z)\Big].$$

The next proposition shows how to find such a δ.

4.3.5 Proposition. *Let the situation be as in Theorem 4.3.2 and Theorem 4.3.4, and let be $\alpha, \delta \in (0, 1)$. Then the following statements are equivalent:*

(i) The radius of $C_\mu(\alpha, z)$, the actual approximate confidence interval of μ (see Equation (4.3) on page 77), is at most $100\,\delta\,\%$ of the estimated value $\hat{\mu}_m(z)$,

$$C_\mu(\alpha, z) \subseteq \Big[(1-\delta)\,\hat{\mu}_m(z),\ (1+\delta)\,\hat{\mu}_m(z)\Big].$$

(ii) The bound δ is chosen large enough, i. e.

$$\delta \ge q_{1-\alpha/2}\sqrt{\frac{1}{\sum_{j=1}^m \sum_{k=1}^d z_{jk}} + \frac{\hat{w}_m(z)}{\sum_{j=1}^m l_j}}.$$

(iii) It holds

$$\left(\frac{\delta}{q_{1-\alpha/2}}\right)^2 > \frac{\hat{w}_m(z)}{\sum_{j=1}^m l_j},$$

and the total number of observed SOLEs is large enough, i. e.

$$\sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk} \geq \frac{1}{\left(\frac{\delta}{q_{1-\alpha/2}}\right)^2 - \frac{\hat{w}_m(z)}{\sum_{j=1}^{m} l_j}}.$$

(iv) The total number of kilometers satisfies the inequality

$$\sum_{j=1}^{m} l_j \geq \left(\frac{1}{\hat{\mu}_m(z)} + \hat{w}_m(z)\right)\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2.$$

This proposition still holds if all '$\geq$' and '$\subseteq$' are replaced by '$=$'.

Proof. The following equivalences are a consequence of the definitions of $\hat{\mu}_m(z)$ and $C_\mu(\alpha, z)$ (see Theorem 4.3.2 and Equation (4.3) on page 77). They still hold if all '$\leq$' and '$\subseteq$' are replaced by '$=$':

$$\begin{aligned}
& C_\mu(\alpha, z) \subseteq \Big[(1-\delta)\,\hat{\mu}_m(z),\ (1+\delta)\,\hat{\mu}_m(z)\Big] \\
\Leftrightarrow\quad & q_{1-\alpha/2}\sqrt{\frac{\hat{\mu}_m(z)\left(1+\hat{w}_m(z)\,\hat{\mu}_m(z)\right)}{\sum_{j=1}^{m} l_j}} \leq \delta\hat{\mu}_m(z) \\
\Leftrightarrow\quad & \left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 \leq \frac{\hat{\mu}_m(z)}{1+\hat{w}_m(z)\,\hat{\mu}_m(z)}\sum_{j=1}^{m} l_j = \left(\frac{1}{\sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk}} + \frac{\hat{w}_m(z)}{\sum_{j=1}^{m} l_j}\right)^{-1}.
\end{aligned}$$

The first line equates to statement (i), the last line is a simple transformation of statements (ii), (iii) and (iv). □

The fourth statement in the last proposition ostensibly answers the question concerning a sufficient and necessary observation time. However, since $\hat{\mu}_m(z)$ equals the quotient of total number of events and total mileage (see Theorem 4.3.2), the inequation in the fourth statement of Proposition 4.3.5 means

$$\sum_{j=1}^{m} l_j \geq \left(\frac{\sum_{j=1}^{m} l_j}{\sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk}} + \hat{w}_m(z)\right)\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2.$$

The total mileage is found on both sides, and in the Poisson case, where $\hat{w}_m(z) = 0$, the total mileage can be canceled on both sides. The necessary mileage cannot be determined in this way.

At least, the necessary number of SOLEs that must be observed is given in the third statement of Proposition 4.3.5. If N_{num} is Bernoulli or Poisson distributed, it is sufficient to observe $({}^{q_{1-\alpha/2}}/_{\delta})^2$ SOLEs, because in this case it is $\hat{w}_m(z) \in \{-1, 0\}$, and so

$$\sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk} \geq \left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 \geq \frac{1}{\left(\frac{\delta}{q_{1-\alpha/2}}\right)^2 - \frac{\hat{w}_m(z)}{\sum_{j=1}^{m} l_j}}.$$

If N_{num} is Poisson distributed, this is even a necessary number of events. The next step is to collect some data and estimate roughly the mean number of SOLEs per kilometer. This allows to forecast the sufficient mileage for observing about $({}^{q_{1-\alpha/2}}/_{\delta})^2$ events.

Let us construct a typical example to illustrate the previous procedure.

4.3.6 Example. Let N_{num} be Poisson distributed, $N_{\mathrm{num}} \sim \mathrm{Poi}(\mu)$. Choose $\alpha = 0.05$ and $\delta = 0.1$ so that

$$\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 = \left(\frac{q_{0.975}}{0.1}\right)^2 \approx 384.1\,.$$

The symbol "$\approx$" means that the values are rounded. This symbol is used in the same way throughout this example. Proposition 4.3.5 provides that at least 385 SOLEs must be observed,

$$\sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk} \geq 385\,,$$

so the radius of the actual approximate confidence interval with confidence level 0.95 is less than ten percent of the estimated value $\hat{\mu}_m(z)$,

$$C_\mu(0.05, z) \subseteq \Big[0.9\,\hat{\mu}_m(z),\ 1.1\,\hat{\mu}_m(z)\Big].$$

In order to assess how long the experiment must run so 385 SOLEs can be observed, one needs an estimate of the number of SOLEs per kilometer. Suppose, the experiment runs since a while, and the estimated value for μ is

$$\hat{\mu}_m(z) = 2 \cdot 10^{-3}.$$

Then, again Proposition 4.3.5 indicates that the interval $C_\mu(\alpha, z)$ is as small as desired if the observation period is at least 192 072.9 kilometers,

$$\sum_{j=1}^{m} l_j \geq \frac{1}{\hat{\mu}_m(z)}\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 = 500 \cdot \left(\frac{q_{0.975}}{0.1}\right)^2 \approx 192\,072.9\,.$$

If, instead, N_{num} is negative binomially distributed, $N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho, \mu)$, one additionally needs a rough estimate of the exponent ϱ, because $\hat{w}_m(z) = {}^1/_{\hat{\varrho}_m(z)}$. Suppose, this estimate equals the estimate of μ,

$$\hat{\varrho}_m(z) = 2 \cdot 10^{-3} = \hat{\mu}_m(z).$$

Then a sufficient mileage is 384 145.9 kilometers

$$\sum_{j=1}^{m} l_j \geq \left(\frac{1}{\hat{\mu}_m(z)} + \frac{1}{\hat{\varrho}_m(z)}\right)\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 = 1000 \cdot \left(\frac{q_{0.975}}{0.01}\right)^2 \approx 384\,145.9\,.$$

Because, in the end, the total number of observed SOLEs determines the dimensions of the confidence interval, the necessary mileage obviously depends on the number of expectable events per kilometer: the greater μ, the smaller is the necessary mileage. This can also be seen in Proposition 4.3.5 and Example 4.3.6.

However, the variance of N_{num} greatly influences the magnitude of the necessary mileage, too. If, according to Proposition 4.3.5, $l_{\text{nec}}^{\text{Poi}}$ and $l_{\text{nec}}^{\text{NBin}}$ denote the necessary mileages in cases of the Poisson and negative binomial distribution, respectively,

$$l_{\text{nec}}^{\text{Poi}} := \frac{1}{\hat{\mu}_m(z)}\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 \qquad \text{and} \qquad l_{\text{nec}}^{\text{NBin}} := \left(\frac{1}{\hat{\mu}_m(z)} + \frac{1}{\hat{\varrho}_m(z)}\right)\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2,$$

and if $N_{\text{num}}^{\text{Poi}} \sim \text{Poi}(\mu)$ and $N_{\text{num}}^{\text{NBin}} \sim \text{NBin}(\varrho, \mu)$, then it holds

$$\frac{l_{\text{nec}}^{\text{NBin}}}{l_{\text{nec}}^{\text{Poi}}} = 1 + \frac{\hat{\mu}_m(z)}{\hat{\varrho}_m(z)} \qquad \text{and} \qquad \frac{\mathbb{V}\text{ar}_\vartheta\left[N_{\text{num}}^{\text{NBin}}\right]}{\mathbb{V}\text{ar}_\vartheta[N_{\text{num}}^{\text{Poi}}]} = 1 + \frac{\mu}{\varrho}.$$

This shows that the necessary mileage increases linearly with the (estimated) variance of N_{num}.

4.3.2. Estimating the Exponent ϱ of the Number of SOLEs per Kilometer

Let the number of SOLEs per kilometer be negative binomially distributed, $N_{\text{num}} \sim \text{NBin}(\varrho, \mu)$, with exponent ϱ and mean μ ($\varrho, \mu \in \mathbb{R}_{>0}$). The estimate of μ is discussed in Section 4.3.1 already. Here, the maximum likelihood estimator of ϱ shall be found.

Anscombe [Ans50] presumed that for $l_1 = \ldots = l_m = 1$ the maximum likelihood estimator of ϱ exists if and only if the (biased) sample variance is larger than the sample mean,

$$\frac{1}{m}\sum_{j=1}^{m}\left(n_j - \frac{1}{m}\sum_{i=1}^{m} n_i\right)^2 > \frac{1}{m}\sum_{j=1}^{m} n_j,$$

where $n_1, \ldots, n_m$ are statistically independent realizations of N_{num}. This condition is in line with the fact that a negative binomially distributed variate is always overdispersed, i. e. the variance is larger than the expectation (see Definition 2.4.2). Aragón *et al.* [AEE92] tried to proof Anscombe's conjecture, but according to Wang [Wan96] their proof is partly wrong. Wang [Wan96] points to Levin and Reeds [LR77] for an overlooked proof of Anscombe's conjecture.

The following theorem provides a similar condition when the mileages l_j vary.

4.3.7 Theorem. *Let* $z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in \mathbb{N}_0{}^{m\times d}$ *be a realization of* $\big(\boldsymbol{Z}_{\mathrm{v}(j)}\big)_{1\le j\le m}$ *with* $z \not\equiv 0$. *Apply the shortcuts*

$$n_j := \sum_{k=1}^{d} z_{jk}, \quad n := \sum_{j=1}^{m} n_j \qquad \text{and} \qquad l := \sum_{j=1}^{m} l_j .$$

Suppose, N_{num} *is negative binomially distributed,* $N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu)$. *Then, the maximum likelihood estimator* $(\hat\varrho_m,\hat\mu_m)$ *of* (ϱ,μ) *based on* z *exists if and only if*

$$\sum_{j=1}^{m} \frac{{n_j}^2}{l_j} - \frac{n^2}{l} > \sum_{j=1}^{m} \frac{n_j}{l_j}.$$

Moreover, if the maximum likelihood estimator exists, $\hat\varrho_m$ *is the unique solution of both of the following equivalent equations:*

$$\sum_{j=1}^{m} l_j\big(\psi(\varrho l_j+n_j)-\psi(\varrho l_j)\big) = l\log\Big(1+\tfrac{n}{\varrho l}\Big) \;\Leftrightarrow\; \sum_{j=1}^{m}\sum_{x=0}^{n_j-1} \frac{1}{\varrho+\frac{x}{l_j}} = l\log\Big(1+\tfrac{n}{\varrho l}\Big),$$

where ψ *is the digamma function,* $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\log(\Gamma(x))$, *[AS65, p. 258].* $\hat\mu_m$ *is given by*

$$\hat\mu_m(z) = \frac{n}{l}.$$

***Proof*.** The formula for $\hat\mu_m$ is provided by Theorem 4.3.2, and the maximum likelihood estimator $\hat\varrho_m$ based on z is the solution of the equation

$$\tfrac{\partial \ell_{\mathrm{num}}}{\partial\varrho}(\varrho,\hat\mu_m(z);z) = 0$$

(see Proposition 4.2.1). This derivative is given in Lemma 4.3.1,

$$\begin{aligned}\tfrac{\partial \ell_{\mathrm{num}}}{\partial\varrho}(\varrho,\hat\mu_m(z);z) &= \sum_{j=1}^{m} l_j\big(\psi(\varrho l_j+n_j)-\psi(\varrho l_j)\big) - l\log\Big(1+\tfrac{n}{\varrho l}\Big)\\ &= \sum_{j=1}^{m}\sum_{x=0}^{n_j-1}\frac{1}{\varrho+\frac{x}{l_j}} - l\log\Big(1+\tfrac{n}{\varrho l}\Big).\end{aligned}$$

With the reparametrization $\eta = \frac{\varrho l}{m}$ and the shortcuts $a_j := \frac{l}{m l_j}$ $(j \in \mathbb{N}_{\leq m})$ the derivative of ℓ_{num} looks like

$$\tfrac{\partial \ell_{\mathrm{num}}}{\partial \varrho}(\varrho, \hat{\mu}_m(z); z) = l \underbrace{\left(\frac{1}{m} \sum_{j=1}^{m} \sum_{x=0}^{n_j - 1} \frac{1}{\eta + x a_j} - \log\left(1 + \tfrac{n}{m\eta}\right) \right)}_{=: H(\eta)}.$$

It remains to be shown that H as function with respect to η has a root if and only if the term $\left(\sum_{j=1}^{m} {n_j}^2/l_j - n^2/l\right)$ is larger than $\sum_{j=1}^{m} n_j/l_j$, that this root is unique if it exists, and that H changes its sign from positive to negative at this root.

In the standard case where $l_1 = \ldots = l_m = 1$ and, therefore, $a_1 = \ldots = a_m = 1$ it has already been shown that $H(\eta)$ has one or no root on $\mathbb{R}_{>0}$ [Wan96, LR77]. It can be expected that even for general factors $a_j \in \mathbb{R}_{>0}$ H has at most one root.

To decide whether H has a root or not, Aragón *et al.* [AEE92] use the reparametrization $r = \frac{1}{\eta}$. Here, this reparametrization leads to the definition

$$\begin{aligned} h(r) :=& \frac{1}{m} \sum_{j=1}^{m} \sum_{x=0}^{n_j - 1} \frac{r}{1 + x a_j r} - \log\left(1 + \tfrac{n}{m} r\right) \\ =& \frac{1}{m} \sum_{j=1}^{m} \sum_{x=1}^{n_j - 1} \frac{r}{1 + x a_j r} + r \left(\frac{m - m_0}{m} - \frac{\log\left(1 + \frac{n}{m} r\right)}{r} \right) \qquad \forall r \in \mathbb{R}_{\geq 0}, \end{aligned}$$

where m_0 is the number of observations with $n_j = 0$. Due to this definition it holds

$$h(r) = H\left(\tfrac{1}{r}\right) = \frac{1}{l} \tfrac{\partial \ell_{\mathrm{num}}}{\partial \varrho}\left(\tfrac{m}{rl}, \hat{\mu}_m(z); z\right) \qquad \forall r \in \mathbb{R}_{>0}.$$

Hence, if r_0 is a root of h on $\mathbb{R}_{>0}$, then $\hat{\varrho}_m(z) = \frac{m}{r_0 l}$.

On the one hand, $h(r)$ tends to ∞ if r approaches ∞, because

$$\lim_{r \to \infty} \sum_{j=1}^{m} \sum_{x=1}^{n_j - 1} \frac{r}{1 + x a_j r} = \sum_{j=1}^{m} \sum_{x=1}^{n_j - 1} \frac{1}{x a_j} \geq 0,$$

and, according to l'Hôpital's Rule [For04, p. 171],

$$\lim_{r \to \infty} \frac{\log\left(1 + \frac{n}{m} r\right)}{r} = \lim_{r \to \infty} \frac{n}{m + nr} = 0.$$

On the other hand, due to the structure of the derivatives of h,

$$\frac{\mathrm{d}h}{\mathrm{d}r}(r) = \frac{1}{m}\sum_{j=1}^{m}\sum_{x=0}^{n_j-1}\frac{1}{(1+xa_jr)^2} - \frac{n}{m+nr},$$

$$\frac{\mathrm{d}^2h}{\mathrm{d}r^2}(r) = -\frac{1}{m}\sum_{j=1}^{m}\sum_{x=0}^{n_j-1}\frac{2a_jx}{(1+xa_jr)^3} + \frac{n^2}{(m+nr)^2},$$

it holds

$$h(0) = 0 \qquad \text{and} \qquad \frac{\mathrm{d}h}{\mathrm{d}r}(0) = 0$$

and

$$\frac{\mathrm{d}^2h}{\mathrm{d}r^2}(0) = -\frac{1}{m}\sum_{j=1}^{m} n_j(n_j-1)a_j + \frac{n^2}{m^2} = \frac{l}{m^2}\left(\sum_{j=1}^{m}\frac{n_j}{l_j} - \sum_{j=1}^{m}\frac{{n_j}^2}{l_j} + \frac{n^2}{l}\right).$$

Hence, if $\left(\sum_{j=1}^{m} {n_j}^2/l_j - n^2/l\right)$ is larger (less) than $\sum_{j=1}^{m} n_j/l_j$, then it is $\frac{\mathrm{d}^2h}{\mathrm{d}r^2}(0) < 0$ $\left(\frac{\mathrm{d}^2h}{\mathrm{d}r^2}(0) > 0\right)$. In this case, h is strictly concave (convex) near 0 [For04, p. 166]. Moreover, the properties $h(0) = 0$, $\frac{\mathrm{d}h}{\mathrm{d}r}(0) = 0$ and $\lim_{r\to\infty} h(r) = \infty$ provide in this situation that there are $\varepsilon, \delta \in \mathbb{R}_{>0}$ such that

$$h(r) < 0 \ \ \big(h(r) > 0\big) \quad \forall r \in (0,\varepsilon) \qquad \text{and} \qquad h(r) > 0 \quad \forall r \in (\delta,\infty).$$

As a consequence h as function on $\mathbb{R}_{>0}$ must have an odd (even) number of roots. But, as found above, h has at most one root. Thus, if $\left(\sum_{j=1}^{m} {n_j}^2/l_j - n^2/l\right)$ is larger than $\sum_{j=1}^{m} n_j/l_j$, h and therefore H both have a unique root on $\mathbb{R}_{>0}$, and if $\left(\sum_{j=1}^{m} {n_j}^2/l_j - n^2/l\right)$ is less than $\sum_{j=1}^{m} n_j/l_j$, h and H both have no root on $\mathbb{R}_{>0}$. For continuity reasons h and H cannot have a root if $\left(\sum_{j=1}^{m} {n_j}^2/l_j - n^2/l\right)$ is equal to $\sum_{j=1}^{m} n_j/l_j$. □

The necessary and sufficient condition in Theorem 4.3.7 for the existence of the maximum likelihood estimator of ϱ reminds of $\hat{D}_2$, the estimator of the index of dispersion,

$$\hat{D}_2 = \hat{D}_2\big((n_j,l_j)_{1\le j\le m}\big) := \frac{\frac{1}{m}\sum_{j=1}^{m}\frac{{n_j}^2}{l_j} - \frac{1}{m}\frac{\left(\sum_{j=1}^{m} n_j\right)^2}{\sum_{j=1}^{m} l_j}}{\frac{1}{m}\sum_{j=1}^{m}\frac{n_j}{l_j}}$$

(see Section 3.5.1, Equation (3.6)). Theorem 4.3.7 provides that $\hat{\varrho}_m$ exists if and only if $(\hat{D}_2 - 1)$ is positive. As described in Section 3.5.7, N_{num} is only assumed to be negative binomially distributed if

$$(\hat{D}_2 - 1) > \sqrt{\tfrac{2}{m}}\, q_{1-\alpha/2},$$

where $q_{1-\alpha/2}$ denotes the $\left(1-\frac{\alpha}{2}\right)100\,\%$ quantile of the standard normal distribution ($\alpha \in (0,1)$). Since $q_{1-\alpha/2}$ is positive for any $\alpha \in (0,1)$, $(\hat{D}_2 - 1)$ must be positive then. Summarized, if N_{num} is chosen to be negative binomially distributed according to the hypothesis test in Section 3.5.3, then the maximum likelihood estimator $(\hat{\varrho}_m, \hat{\mu}_m)$ of (ϱ, μ) exists.

The maximum likelihood estimator $\hat{\varrho}_m$ cannot be unbiased, because Wang [Wan96] verified that an unbiased estimator of ϱ does not exist (his simple proof even works for arbitrary mileages). Anscombe [Ans50] even found (in case of $l_1 = \ldots = l_m = 1$) that $\hat{\varrho}_m$ does not have a proper distribution, because with positive probability the sample mean is larger than the sample variance. Also here, with positive probability the maximum likelihood estimator of ϱ does not exist. As mentioned above, $\hat{\varrho}_m$ exists if and only if $\hat{D}_2$ exceeds 1. Theorem 3.5.2 provides that $\hat{D}_2$ is approximately normally distributed with mean $\mathbb{D}[N_{\text{num}}]$ and variance ${\tau_{\text{iod}}}^2$. Hence, the probability that $\hat{\varrho}_m$ does not exist approximately is

$$\begin{aligned}\mathbb{P}_\vartheta\left(\hat{D}_2 \le 1\right) &= \mathbb{P}_\vartheta\left(\sqrt{\tfrac{m}{{\tau_{\text{iod}}}^2}}\big(\hat{D}_2 - \mathbb{D}[N_{\text{num}}]\big) \le -\sqrt{m}c_\vartheta\right)\\ &\approx \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\sqrt{m}c_\vartheta} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x,\end{aligned}$$

where $c_\vartheta := {(\mathbb{D}[N_{\text{num}}]-1)}/{\tau_{\text{iod}}}$. Since the index of dispersion $\mathbb{D}[N_{\text{num}}]$ is larger than 1 in the negative binomial case, the probability that $\hat{\varrho}_m$ does not exist tends to zero for $m \to \infty$. If someone is interested in the speed of convergence, he must calculate the constant c_ϑ. According to Anscombe [Ans50] the first four cumulants of a negative binomially distributed variate are

$$\kappa_1[N_{\text{num}}] = \mu, \quad \kappa_2[N_{\text{num}}] = \mu\left(1+\frac{\mu}{\varrho}\right), \quad \kappa_3[N_{\text{num}}] = \mu\left(1+\frac{\mu}{\varrho}\right)\left(1+\frac{2\mu}{\varrho}\right)$$
$$\kappa_4[N_{\text{num}}] = \mu\left(1+\frac{\mu}{\varrho}\right)\left(1+\frac{6\mu}{\varrho}+\frac{6\mu^2}{\varrho^2}\right).$$

Hence, numerator and denominator of c_ϑ are

$$\big(\mathbb{D}[N_{\text{num}}]-1\big) = \frac{\mu}{\varrho}, \qquad \tau_{\text{iod}} = \sqrt{\mathbb{E}\big[\tfrac{1}{L}\big]\left(\frac{2}{\varrho}+\frac{5\mu}{\varrho^2}+\frac{3\mu^2}{\varrho^3}\right)+2\left(1+\frac{\mu}{\varrho}\right)^2} \tag{4.4}$$

due to Theorem 3.5.2.

It is well-known (if $l_1 = \ldots = l_m = 1$) that the maximum likelihood estimator of (ϱ, μ) is asymptotically efficient [Hal41, Ans50, Law87, SZ06], i. e. $(\hat{\varrho}_m, \hat{\mu}_m)$ is asymptotically jointly normally distributed with mean (ϱ, μ) and the inverse of the Fisher information matrix as covariance matrix (see Section

2.4.3). Generally, this is true for all maximum likelihood estimators if some regularity conditions hold [LC98, pp. 449/463]. Even if the random variables are not identically distributed, in many situations the maximum likelihood estimator is asymptotically efficient (consider particularly the results of Inagaki [Ina73], also see Bradley [BG62] and Hoaley [Hoa71]; examples for inconsistent maximum likelihood estimators see Crowder [Cro86]). Therefore and with a reference to the numerical results in section 5.3.2, it can be assumed that even here $(\hat{\varrho}_m, \hat{\mu}_m)$ is asymptotically efficient. Theorem 4.2.2 provides under the assumption of Remark 4.2.3 that the inverse of the Fisher information matrix concerning ϱ and μ is

$$I_{\mathrm{num}}(\varrho,\mu)^{-1} = \begin{pmatrix} {\sigma_{\varrho,m}}^2 & 0 \\ 0 & \frac{\mu\left(1+\frac{\mu}{\varrho}\right)}{\sum_{j=1}^m l_j} \end{pmatrix},$$

where

$${\sigma_{\varrho,m}}^2 := \left(\sum_{j=1}^m {l_j}^2\, \mathbb{E}_\vartheta\left[h\left(\varrho l_j, N_{\mathrm{num}}^{*l_j}\right)\right] - \log\left(1+\tfrac{\mu}{\varrho}\right)^2 \sum_{j=1}^m {l_j}^2 - \frac{\mu}{\varrho(\varrho+\mu)}\sum_{j=1}^m l_j\right)^{-1},$$

$$h\colon {\mathbb{R}_{>0}}^2 \to \mathbb{R}_{>0}\colon\; (x,y) \longmapsto \left(\sum_{n=0}^{y-1} \frac{1}{x+n}\right)^2 = \big(\psi(x+y)-\psi(x)\big)^2.$$

The fact that $\hat{\mu}_m$ is asymptotically normally distributed with mean μ and variance $\mu\left(1+\frac{\mu}{\varrho}\right)/\sum_{j=1}^m l_j$ is already known from Theorem 4.3.3 and Theorem 4.3.4. The asymptotic efficiency of $\hat{\varrho}_m$ yields with $\boldsymbol{Z} := (\boldsymbol{Z}_{\mathrm{v}(j)})_{1\le j\le m}$

$$\begin{aligned} 1-\alpha &= \lim_{m\to\infty} \mathbb{P}_\vartheta\left(-q_{1-\alpha/2} \le \tfrac{\hat{\varrho}_m(\boldsymbol{Z})-\varrho}{\hat{\sigma}_{\varrho,m}(\boldsymbol{Z})} \le q_{1-\alpha/2}\right) \\ &= \lim_{m\to\infty} \mathbb{P}_\vartheta\left(\hat{\varrho}_m(\boldsymbol{Z}) - \hat{\sigma}_{\varrho,m}(\boldsymbol{Z})\, q_{1-\alpha/2} \le \varrho \le \hat{\varrho}_m(\boldsymbol{Z}) + \hat{\sigma}_{\varrho,m}(\boldsymbol{Z})\, q_{1-\alpha/2}\right), \end{aligned}$$

where $q_{1-\alpha/2}$ is the $\left(1-\frac{\alpha}{2}\right)100\,\%$ quantile of the standard normal distribution and $\hat{\sigma}_{\varrho,m}(\boldsymbol{Z})$ equates $\sigma_{\varrho,m} = \sqrt{{\sigma_{\varrho,m}}^2}$ with μ and ϱ replaced by the estimators $\hat{\mu}_m(\boldsymbol{Z})$ and $\hat{\varrho}_m(\boldsymbol{Z})$ respectively. Eventually, if z is a realization of $\boldsymbol{Z}$, an approximate actual confidence interval of ϱ with confidence level $(1-\alpha)$ $(\alpha \in (0,1))$ is

$$C_\varrho(\alpha, z) := \left[\hat{\varrho}_m(z) - \hat{\sigma}_{\varrho,m}(z)\, q_{1-\alpha/2},\; \hat{\varrho}_m(z) + \hat{\sigma}_{\varrho,m}(z)\, q_{1-\alpha/2}\right]. \tag{4.5}$$

It also should be noted that the entries of the anti-diagonal of the inverse Fisher information matrix above are all equal to 0. Since this is the asymptotic covariance matrix of $(\hat{\varrho}_m, \hat{\mu}_m)$, $\hat{\varrho}_m$ and $\hat{\mu}_m$ are asymptotically independent, because jointly normally distributed and uncorrelated random variables are statistically independent [Als05, p. 141].

Section 4.3.1 discusses the question how long the experiment must run so the confidence interval of μ is smaller than a default value (see Proposition 4.3.5 and

Example 4.3.6). It is argued that rough estimates of $\hat{\mu}_m$ and $\hat{\varrho}_m$ are needed to answer this question. Concerning ϱ, it is even more difficult to find an answer. However, since the (approximate) variance of $\hat{\varrho}_m$, ${\sigma_{\varrho,m}}^2$, surely increases if the mileages increase, it must hold

$$ {\sigma_{\varrho,m}}^2 \leq \left(m{l_0}^2\, \mathbb{E}_{\vartheta}\left[h\left(\varrho l_0, N^{*l_0}_{\text{num}}\right)\right] - m{l_0}^2 \log\left(1+\tfrac{\mu}{\varrho}\right)^2 - \frac{ml_0\mu}{\varrho(\varrho+\mu)}\right)^{-1}, $$

where $l_0 := \min\{l_j \mid 1 \leq j \leq m\}$. This results in the following basic transformations:

$$ m \geq \frac{\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2}{\hat{\varrho}_m(z)^2\left({l_0}^2\,\mathbb{E}_{\hat{\vartheta}_m}\left[h\left(\hat{\varrho}_m l_0, N^{*l_0}_{\text{num}}\right)\right] - {l_0}^2\log\left(1+\frac{\hat{\mu}_m}{\hat{\varrho}_m}\right)^2 - \frac{l_0\hat{\mu}_m}{\hat{\varrho}_m(\hat{\varrho}_m+\hat{\mu}_m)}\right)} \tag{4.6} $$

$$ \Rightarrow \qquad 1 \geq \frac{\hat{\sigma}_{\varrho,m}(z)^2}{\hat{\varrho}_m(z)^2}\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2 $$

$$ \Leftrightarrow \qquad C_\varrho(\alpha,z) \subseteq \left[(1-\delta)\,\hat{\varrho}_m(z),\ (1+\delta)\,\hat{\varrho}_m(z)\right] $$

Consequently, Equation (4.6) above yields a sufficient sample size of vehicles each with mileage l_0 so that the radius of $C_\varrho(\alpha,z)$ is at most $100\,\delta\,\%$ of the estimated value $\hat{\varrho}_m$.

4.3.8 Example. Let the situation be as in Example 4.3.6, i. e.

$$ \alpha = 0.05, \qquad \delta = 0.1, \qquad \hat{\varrho}_m(z) = 2\cdot 10^{-3} = \hat{\mu}_m(z). $$

Based on these values, the evaluation of Equation (4.6) on page 86 yields

$$ m \geq \frac{\left(\frac{q_{1-\alpha/2}}{\delta}\right)^2}{\hat{\varrho}_m(z)^2\left({l_0}^2\,\mathbb{E}_{\hat{\vartheta}_m}\left[h\left(\hat{\varrho}_m l_0, N^{*l_0}_{\text{num}}\right)\right] - {l_0}^2\log\left(1+\frac{\hat{\mu}_m}{\hat{\varrho}_m}\right)^2 - \frac{l_0\hat{\mu}_m}{\hat{\varrho}_m(\hat{\varrho}_m+\hat{\mu}_m)}\right)} \approx \begin{cases} 996\,912.5, & \text{if } l_0 = 1,\\ 3782.8, & \text{if } l_0 = 1000. \end{cases} $$

This means that both a sample size of 996 913 vehicles each with mileage 1 and a sample size of 3783 vehicles each with mileage 1000 are sufficient for $C_\varrho(0.05,z) \subseteq \left[0.9\,\hat{\varrho}_m(z),\ 1.1\,\hat{\varrho}_m(z)\right]$.

4.4. Estimating the Severity of a SOLE in the Counting Model

Section 3.6 provides that an arbitrary SOLE S_{sev} is assumed to have a shifted generalized Pareto distribution with positive shape,

$$F_{\text{sev}}(t) = F_{\text{sev}}(t;\xi,\beta) = \mathbb{1}_{\mathbb{R}_{>u_{\text{sev}}}}(t) \cdot \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}(t - u_{\text{sev}})\right)^{-\frac{1}{\xi}}, & \text{if } \xi > 0, \\ 1 - \mathrm{e}^{-\frac{1}{\beta}(t-u_{\text{sev}})}, & \text{if } \xi = 0 \end{cases}$$

for all $t \in \mathbb{R}$. The parameters ξ and β shall be estimated based on the counts $z = (z_{jk})_{\substack{1\le j\le m \\ 1\le k\le d}}$ as described in Section 4.1 and Section 4.2.1. According to Proposition 4.2.1, the maximum likelihood estimator of $\varsigma = (\xi,\beta)$ is the maximizer of

$$\ell^{\text{C}}_{\text{sev}}(\varsigma; z) = \sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk}\log\bigl(p_{A_{jk}}\bigr) = \sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk}\log(F_{\text{sev}}(t_{jk}) - F_{\text{sev}}(t_{j,k-1}))\,.$$

Before looking for a maximum likelihood estimator of ς in Section 4.4.2, the following Section 4.4.1 gives some conditions in which such an estimator does *not* exist. In doing so, it is not taken into account that F_{sev} is chosen to be the cumulative distribution function of a shifted generalized Pareto distribution but Theorem 4.4.1 holds for arbitrary severity distributions. Finally, Section 4.4.4 covers the case when the class limits are equidistant, i. e. the class lengths are all equal to each other,

$$t_{jk} - t_{j,k-1} = t_{j,k+1} - t_{jk} \qquad \forall j \in \mathbb{N}_{\le m},\ \forall k \in \mathbb{N}_{\le d-2}.$$

It is shown that a class length can be found which optimizes the accuracy of parameter estimate.

4.4.1. When does the Maximum Likelihood Estimator of (ξ,β) not Exist

Regardless of whether S_{sev} is shifted generalized Pareto distributed, there are situations where the maximum likelihood estimator of the parameters of F_{sev} does not exist. If, for instance, the observation reveals only events in the lowest class, one may believe that the severity of any SOLE lies almost sure within the lowest class. However, an absolutely continuous and strictly increasing cumulative distribution function gives positive probabilities to all classes. Depending on the structure of F_{sev}, it may happen that the likelihood function does not have a maximum. Kulldorff [Kul61] verifies this fact in case of the exponential distribution and the normal distribution. The following theorem proves this and some other facts in general.

4.4.1 Theorem. *Let* $z = (z_{jk})_{\substack{1 \le j \le m \\ 1 \le k \le d}} \in {\mathbb{N}_0}^{m \times d}$ *be a realization of* $\big(\boldsymbol{Z}_{\mathrm{v}(j)}\big)_{1 \le j \le m}$.

1. *Suppose* F_{sev} *has the following characteristics:*
   - $F_{\mathrm{sev}}(t; \varsigma) < 1 \qquad \forall t \in \mathbb{R},\ \forall \varsigma \in \Theta_{\mathrm{sev}},$
   - $\sup\limits_{\varsigma \in \Theta_{\mathrm{sev}}} F_{\mathrm{sev}}(t; \varsigma) = 1 \qquad \forall t \in \mathbb{R}_{>u_{\mathrm{sev}}}.$

   If $z_{jk} = 0$ *for all* $j \in \mathbb{N}_{\le m}$ *and* $k \in \{2, \ldots, d\}$, *then* $\varsigma \mapsto \ell^{\mathrm{C}}_{\mathrm{sev}}(\varsigma; z)$ *is identically zero or does not have a maximum. Thus, in this case, there does not exist a maximum likelihood estimator of* ς *based on* z.

2. *Suppose* F_{sev} *has the following characteristics:*
   - $F_{\mathrm{sev}}(t; \varsigma) > 0 \qquad \forall t \in \mathbb{R}_{>u_{\mathrm{sev}}},\ \forall \varsigma \in \Theta_{\mathrm{sev}},$
   - $\inf\limits_{\varsigma \in \Theta_{\mathrm{sev}}} F_{\mathrm{sev}}(t; \varsigma) = 0 \qquad \forall t \in \mathbb{R}.$

   If $z_{jk} = 0$ *for all* $j \in \mathbb{N}_{\le m}$ *and* $k \in \mathbb{N}_{\le d-1}$, *then the function* $\varsigma \mapsto \ell^{\mathrm{C}}_{\mathrm{sev}}(\varsigma; z)$ *is identically zero or does not have a maximum. Thus, in this case, there does not exist a maximum likelihood estimator of* ς *based on* z.

3. *Suppose* F_{sev} *has the following characteristics:*
   - $F_{\mathrm{sev}}(s; \varsigma) < F_{\mathrm{sev}}(t; \varsigma) \qquad \forall \varsigma \in \Theta_{\mathrm{sev}},\ \forall s, t \in \mathbb{R}_{\ge u_{\mathrm{sev}}}$ with $s < t$,
   - *for all* $a \in [0, 1]$ *there is a sequence* $(\varsigma_n)_{n \in \mathbb{N}} \subseteq \Theta_{\mathrm{sev}}$ *such that*
   $$\lim_{n \to \infty} F_{\mathrm{sev}}(t; \varsigma_n) = a \qquad \forall t \in \mathbb{R}_{>u_{\mathrm{sev}}}.$$

   Suppose it is $d \in \mathbb{N}_{\ge 3}$ *and the class limits are chosen such that*
   $$\max_{1 \le j \le m} t_{j1} < \min_{1 \le j \le m} t_{j,d-1}.$$
   If $z_{jk} = 0$ *for all* $j \in \mathbb{N}_{\le m}$ *and* $k \in \{2, \ldots, d-1\}$, *then the function* $\varsigma \mapsto \ell^{\mathrm{C}}_{\mathrm{sev}}(\varsigma; z)$ *is identically zero or does not have a maximum. Thus, in this case, there does not exist a maximum likelihood estimator of* ς *based on* z.

Proof. **1.:** If $z \equiv 0$, then $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,; z) \equiv 0$. Otherwise, it holds
$$\ell^{\mathrm{C}}_{\mathrm{sev}}(\varsigma; z) = \sum_{j=1}^{m} z_{j1} \log(F_{\mathrm{sev}}(t_{j1}; \varsigma)) < 0 \qquad \forall \varsigma \in \Theta_{\mathrm{sev}}.$$
Define $t_{\min} := \min\{t_{j1} | 1 \le j \le m\}$, then it can be shown that the supremum of $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,; z)$ vanishes,
$$\begin{aligned} 0 \ge \sup_{\varsigma \in \Theta_{\mathrm{sev}}} \ell^{\mathrm{C}}_{\mathrm{sev}}(\varsigma; z) &\ge \sup_{\varsigma \in \Theta_{\mathrm{sev}}} \log(F_{\mathrm{sev}}(t_{\min}; \vartheta)) \sum_{j=1}^{m} z_{j1} \\ &= \log\Bigg(\sup_{\varsigma \in \Theta_{\mathrm{sev}}} F_{\mathrm{sev}}(t_{\min}; \vartheta)\Bigg) \sum_{j=1}^{m} z_{j1} = 0 \end{aligned}$$

Thus, $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,;z)$ does not have a maximum.

2.: If $z \equiv 0$, then $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,;z) \equiv 0$. Otherwise, it holds

$$\ell_{\text{sev}}^{\text{C}}(\varsigma;z) = \sum_{j=1}^{m} z_{jd} \log(1 - F_{\text{sev}}(t_{j,d-1};\varsigma)) < 0 \qquad \forall \varsigma \in \Theta_{\text{sev}}.$$

However, the supremum of $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,;z)$ is equal to 0, because with the notation $t_{\max} := \max\{t_{j,d-1} | 1 \le j \le m\}$ it is

$$0 \ge \sup_{\varsigma \in \Theta_{\text{sev}}} \ell_{\text{sev}}^{\text{C}}(\varsigma;z) \ge \log\left(1 - \inf_{\varsigma \in \Theta_{\text{sev}}} F_{\text{sev}}(t_{\max};\varsigma)\right) \sum_{j=1}^{m} z_{jd} = 0.$$

Thus, $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,;z)$ does not have a maximum.

3.: Without loss of generality, let there be $j, i \in \mathbb{N}_{\le m}$ such that $z_{j1} > 0$ and $z_{id} > 0$, because otherwise the situation is described by the first or the second statement of this theorem.

Define $t^* := \max\{t_{j1} | 1 \le j \le m\}$, then for all $\varsigma \in \Theta_{\text{sev}}$ it holds

$$\begin{aligned} \ell_{\text{sev}}^{\text{C}}(\varsigma;z) &= \sum_{j=1}^{m} \Big(z_{j1} \log(F_{\text{sev}}(t_{j1};\varsigma)) + z_{jd} \log(1 - F_{\text{sev}}(t_{j,d-1};\varsigma)) \Big) \\ &< \log(F_{\text{sev}}(t^*;\varsigma)) \sum_{j=1}^{m} z_{j1} + \log(1 - F_{\text{sev}}(t^*;\varsigma)) \sum_{j=1}^{m} z_{jd}, \end{aligned}$$

because the assumptions of this statement include

$$F_{\text{sev}}(t_{j1};\varsigma) \le F_{\text{sev}}(t^*;\varsigma) \quad \text{and} \quad F_{\text{sev}}(t_{j,d-1};\varsigma) > F_{\text{sev}}(t^*;\varsigma) \qquad \forall j \in \mathbb{N}_{\le m}.$$

The continuous function $x \mapsto a\log(x) + b\log(1-x)$ on $(0,1)$ $(a, b \in \mathbb{R}_{>0})$ becomes its global maximum at $x = \frac{a}{a+b}$, which can be verified by discovering its derivative $x \mapsto \frac{1}{1-x}\left(\frac{a}{x} - (a+b)\right)$. This implies

$$\ell_{\text{sev}}^{\text{C}}(\varsigma;z) < \log\left(\frac{\sum_{j=1}^{m} z_{j1}}{\sum_{j=1}^{m}(z_{j1}+z_{jd})}\right) \sum_{j=1}^{m} z_{j1} + \log\left(\frac{\sum_{j=1}^{m} z_{jd}}{\sum_{j=1}^{m}(z_{j1}+z_{jd})}\right) \sum_{j=1}^{m} z_{jd} \quad \forall \varsigma \in \Theta_{\text{sev}}.$$

But the existence of a sequence $(\varsigma_n)_{n\in\mathbb{N}} \subseteq \Theta_{\text{sev}}$ with

$$\lim_{n\to\infty} F_{\text{sev}}(t;\varsigma_n) = \frac{\sum_{j=1}^{m} z_{j1}}{\sum_{j=1}^{m}(z_{j1}+z_{jd})} \qquad \forall t \in \mathbb{R}_{>u_{\text{sev}}}$$

ensures

$$\lim_{n\to\infty} \ell_{\text{sev}}^{\text{C}}(\varsigma_n;z) = \log\left(\frac{\sum_{j=1}^{m} z_{j1}}{\sum_{j=1}^{m}(z_{j1}+z_{jd})}\right) \sum_{j=1}^{m} z_{j1} + \log\left(\frac{\sum_{j=1}^{m} z_{jd}}{\sum_{j=1}^{m}(z_{j1}+z_{jd})}\right) \sum_{j=1}^{m} z_{jd}.$$

Thus, the maximum of $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,;z)$ does not exist. □

In Section 4.4.2, more precisely in Theorem 4.4.4, it is shown that the generalized Pareto distribution satisfies all the conditions mentioned in the last theorem. Moreover, in the situation of the third statement of Theorem 4.4.1 it even holds 'if and only if' as long as S_{sev} is (shifted) generalized Pareto distributed.

4.4.2. Maximum Likelihood Estimation of (ξ, β)

The common way to find a maximum likelihood estimator is looking for roots of the derivative(s) of the log-likelihood function. Since the function $\ell_{\mathrm{sev}}^{\mathrm{C}}$ shall be maximized, the gradient of $\ell_{\mathrm{sev}}^{\mathrm{C}}$ is needed. It is important to note that the gradient, which is given in the next lemma, is not directly related to the class limits t_{jk}, but, as it is $\ell_{\mathrm{sev}}^{\mathrm{C}}$, it depends on the relative class limits $s_{jk} = t_{jk} - u_{\mathrm{sev}}$ as defined in Section 4.1.

4.4.2 Lemma. *Let* $z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in {\mathbb{N}_0}^{m\times d}$ *be a realization of* $(\boldsymbol{Z}_{\mathrm{v}(j)})_{1\le j\le m}$. *Define for all* $t \in \mathbb{R}_{\ge u_{\mathrm{sev}}}$ *and* $x, a \in \mathbb{R}_{\ge 0}$

$$F_{\mathrm{sev}}^*(t) := 1 - F_{\mathrm{sev}}(t) = \begin{cases} \left(1 + \frac{\xi}{\beta}(t - u_{\mathrm{sev}})\right)^{-\frac{1}{\xi}}, & \text{if } \xi > 0, \\ \mathrm{e}^{-\frac{1}{\beta}(t-u_{\mathrm{sev}})}, & \text{if } \xi = 0, \end{cases}$$

$$\varphi_i(x, a) := \mathbb{1}_{\{2\}}(i) + \mathbb{1}_{\{1\}}(i) \cdot \begin{cases} \frac{1}{x}\left(\log(1 + xa)\left(1 + \frac{1}{xa}\right) - 1\right), & \text{if } xa > 0, \\ \frac{a}{2}, & \text{if } xa = 0, \end{cases}$$

$(i \in \{1, 2\})$. Then the gradient[3] *of* $\ell_{\mathrm{sev}}^{\mathrm{C}}(\,\cdot\,; z)$ *is*

$$\operatorname{grad}\left(\ell_{\mathrm{sev}}^{\mathrm{C}}(\xi, \beta; z)\right) := \left(\frac{\partial \ell_{\mathrm{sev}}^{\mathrm{C}}}{\partial \xi}(\xi, \beta; z), \frac{\partial \ell_{\mathrm{sev}}^{\mathrm{C}}}{\partial \beta}(\xi, \beta; z)\right) = \left(\Delta_1(\xi, \beta; z),\ \Delta_2(\xi, \beta; z)\right)$$

with

$$\begin{aligned} \Delta_i(\xi, \beta; z) &= \sum_{j=1}^{m}\sum_{k=1}^{d-1} \frac{\varphi_i\left(\frac{\xi}{\beta}, s_{jk}\right) s_{jk}}{\beta^2 + \xi\beta s_{jk}} \left(\frac{z_{j,k+1}}{1 - \frac{F_{\mathrm{sev}}^*(t_{j,k+1})}{F_{\mathrm{sev}}^*(t_{jk})}} - \frac{z_{jk}}{\frac{F_{\mathrm{sev}}^*(t_{j,k-1})}{F_{\mathrm{sev}}^*(t_{jk})} - 1} \right) \\ &= \sum_{j=1}^{m}\sum_{k=1}^{d-1} z_{jk} \left(\frac{\varphi_i\left(\frac{\xi}{\beta}, s_{j,k-1}\right) \frac{s_{j,k-1}}{\beta^2 + \xi\beta s_{j,k-1}}}{1 - \frac{F_{\mathrm{sev}}^*(t_{jk})}{F_{\mathrm{sev}}^*(t_{j,k-1})}} - \frac{\varphi_i\left(\frac{\xi}{\beta}, s_{jk}\right) \frac{s_{jk}}{\beta^2 + \xi\beta s_{jk}}}{\frac{F_{\mathrm{sev}}^*(t_{j,k-1})}{F_{\mathrm{sev}}^*(t_{jk})} - 1} \right) \\ &\qquad + \sum_{j=1}^{m} z_{jd}\ \varphi_i\left(\frac{\xi}{\beta}, s_{j,d-1}\right) \frac{s_{j,d-1}}{\beta^2 + \xi\beta s_{j,d-1}}, \end{aligned}$$

[3] at $\xi = 0$ $\frac{\partial}{\partial \xi}$ means the right partial derivative

Proof. The definition of $\ell_{\text{sev}}^{\text{C}}$ in Proposition 4.2.1 implies

$$\frac{\partial \ell_{\text{sev}}^{\text{C}}}{\partial \xi}(\xi, \beta; z) = \sum_{j=1}^{m} \sum_{k=1}^{d} z_{jk} \frac{\frac{\partial}{\partial \xi} p_{A_{jk}}}{p_{A_{jk}}}.$$

$p_{A_{jk}}$ denotes the probability that a SOLE falls into the interval $A_{jk} = (t_{j,k-1}, t_{jk}]$,

$$p_{A_{jk}} = p_{(t_{j,k-1}, t_{jk}]} = F_{\text{sev}}(t_{jk}) - F_{\text{sev}}(t_{j,k-1}) = F_{\text{sev}}^*(t_{j,k-1}) - F_{\text{sev}}^*(t_{jk}).$$

(see Definition 3.1.1). For all $j \in \mathbb{N}_{\leq m}$, the coefficient of z_{jk} in the derivative of $\ell_{\text{sev}}^{\text{C}}$ is

$$\frac{\frac{\partial}{\partial \xi} p_{A_{jd}}}{p_{A_{jd}}} = \frac{\frac{\partial}{\partial \xi} F_{\text{sev}}^*(t_{j,d-1})}{F_{\text{sev}}^*(t_{j,d-1})} \quad \text{and} \quad \frac{\frac{\partial}{\partial \xi} p_{A_{jk}}}{p_{A_{jk}}} = \frac{\frac{\frac{\partial}{\partial \xi} F_{\text{sev}}^*(t_{j,k-1})}{F_{\text{sev}}^*(t_{j,k-1})}}{1 - \frac{F_{\text{sev}}^*(t_{jk})}{F_{\text{sev}}^*(t_{j,k-1})}} - \frac{\frac{\frac{\partial}{\partial \xi} F_{\text{sev}}^*(t_{jk})}{F_{\text{sev}}^*(t_{jk})}}{\frac{F_{\text{sev}}^*(t_{j,k-1})}{F_{\text{sev}}^*(t_{jk})} - 1}$$

for all $k \in \mathbb{N}_{\leq d-1}$ (for $k = d$ consider that $F_{\text{sev}}^*(t_{jd}) = 0$). The (right) derivative of F_{sev}^* can be read out in the appendix (see Lemma A.1).

The derivative of $\ell_{\text{sev}}^{\text{C}}$ with respect to β can be verified in the exact same way. □

The partial derivatives of $\ell_{\text{sev}}^{\text{C}}$ with respect to ξ and β differ only in the terms φ_1 and φ_2 defined in the last lemma. While $\varphi_2 \equiv 1$, $\varphi_1(\,\cdot\,, a)$ is, for any fixed $a \in \mathbb{R}_{>0}$, continuous, positive, strictly decreasing and strictly convex, and it holds

$$\lim_{x \to 0} \varphi_1(x, a) = \frac{a}{2} \qquad \text{and} \qquad \lim_{x \to \infty} \varphi_1(x, a) = 0$$

(see lemma A.2 in the appendix; moreover, it can be conjectured that all the derivative terms $(-1)^n \frac{\partial^n \varphi_1}{\partial x^n}(\,\cdot\,, a)$ are positive, strictly decreasing, strictly convex, and they tend to 0 if x approaches ∞).

By means of the gradient of $\ell_{\text{sev}}^{\text{C}}$ a maximum likelihood estimator of (ξ, β) can be found. Every common root of the partial derivatives of $\ell_{\text{sev}}^{\text{C}}$ is a candidate. But questions raised are: 1. is there a common root, and 2. are there more than one common root. The next proposition gives answers to these questions under the condition that not all medium classes are empty (otherwise, the maximum likelihood estimator does not exist, see Theorem 4.4.1 and Theorem 4.4.4).

4.4.3 Proposition. *Let be $d \in \mathbb{N}_{\geq 3}$. For any $j \in \mathbb{N}_{\leq m}$ and $k \in \{2, \ldots, d-1\}$ define the function*

$$\beta_{jk}^* \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \colon \ \xi \longmapsto \begin{cases} \xi s_{jk} \dfrac{1 - \left(\frac{s_{j,k-1}}{s_{jk}}\right)^{\frac{1}{\xi+1}}}{\left(\frac{s_{jk}}{s_{j,k-1}}\right)^{\frac{\xi}{\xi+1}} - 1}, & \text{if } \xi > 0, \\[2ex] \dfrac{s_{jk} - s_{j,k-1}}{\log\left(\frac{s_{jk}}{s_{j,k-1}}\right)}, & \text{if } \xi = 0. \end{cases}$$

Given a realization $z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in \mathbb{N}_0{}^{m\times d}$ *of* $\big(\boldsymbol{Z}_{\mathrm{v}(j)}\big)_{1\le j\le m}$ *such that there is at least one* $(j_0,k_0) \in \mathbb{N}_{\le m}\times\{2,\dots,d-1\}$ *with* $z_{j_0k_0} > 0$*, it holds:*

1. *For every* $\xi\in\mathbb{R}_{\ge 0}$ *there exist unique roots* $\beta^\circ(\xi), \beta^*(\xi)\in\mathbb{R}_{>0}$ *such that*

$$\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi}(\xi,\beta^\circ(\xi);z) = 0 \qquad\text{and}\qquad \frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta}(\xi,\beta^*(\xi);z) = 0,$$

and both derivatives as function with respect to β *change their sign from positive to negative at the particular root.*

2. *If the upper classes are empty, i. e. there is a* $b\in\{2,\dots,d-1\}$ *such that* $z_{jk} = 0$ *for all* $j\in\mathbb{N}_{\le m}$ *and* $k\in\{b+1,\dots,d\}$*, then the root* $\beta^*(\xi)$ *of* $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta}(\xi,\cdot\,;z)$ *is bounded from above in the following way:*

$$\beta^*(\xi) \le \max_{\substack{1\le j\le m\\ 2\le k\le b}} \beta^*_{jk}(\xi) \le \max_{\substack{1\le j\le m\\ 2\le k\le b}} \frac{s_{jk}-s_{j,k-1}}{\log\Big(\frac{s_{jk}}{s_{j,k-1}}\Big)} < \max_{1\le j\le m} s_{jb}.$$

If the lower classes are empty, i. e. there is an $a\in\{2,\dots,d-1\}$ *such that* $z_{jk} = 0$ *for all* $j\in\mathbb{N}_{\le m}$ *and* $k\in\mathbb{N}_{\le a-1}$*, then the root* $\beta^*(\xi)$ *of* $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta}(\xi,\cdot\,;z)$ *is bounded from below in the following way:*

$$\beta^*(\xi) \ge \min_{\substack{1\le j\le m\\ a\le k\le d-1}} \beta^*_{jk}(\xi) \ge \min_{\substack{1\le j\le m\\ a\le k\le d-1}} \frac{\log\Big(\frac{s_{jk}}{s_{j,k-1}}\Big)}{\frac{1}{s_{j,k-1}}-\frac{1}{s_{jk}}} > \min_{1\le j\le m} s_{j,a-1}.$$

3. $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,;z)$ *has got at most one stationary point, i. e. there is at most one point* $(\xi_0,\beta_0)\in\Theta_{\mathrm{sev}}$ *satisfying*

$$\operatorname{grad}\Big(\ell^{\mathrm{C}}_{\mathrm{sev}}(\xi_0,\beta_0;z)\Big) = 0.$$

If such a stationary point exists, it is the global maximizer of $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,;z)$*.*

Proof. **1.:** At first, let us have a look at the partial derivatives of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ as given in Lemma 4.4.2. More precisely, analyze any single addend of this derivatives. For this purpose, define for $i\in\{1,2\}$

$$S_i(\xi,\beta,j,k) := \frac{\varphi_i\Big(\frac{\xi}{\beta}, s_{j,k-1}\Big)\,\frac{s_{j,k-1}}{\beta^2+\beta\xi s_{j,k-1}}}{1-\frac{F^*_{\mathrm{sev}}(t_{jk})}{F^*_{\mathrm{sev}}(t_{j,k-1})}} - \frac{\varphi_i\Big(\frac{\xi}{\beta}, s_{jk}\Big)\,\frac{s_{jk}}{\beta^2+\beta\xi s_{jk}}}{\frac{F^*_{\mathrm{sev}}(t_{j,k-1})}{F^*_{\mathrm{sev}}(t_{jk})}-1},$$

where $j\in\mathbb{N}_{\le m}$ and $k\in\{2,\dots,d-1\}$ (see Lemma 4.4.2 for the definitions of φ_i and F^*_{sev}). A transposition of S_i yields

$$S_i(\xi,\beta,j,k)$$
$$= \frac{\varphi_i\Big(\frac{\xi}{\beta}, s_{jk}\Big)\,\frac{s_{jk}}{\beta^2+\beta\xi s_{jk}}}{\frac{F^*_{\mathrm{sev}}(t_{j,k-1})}{F^*_{\mathrm{sev}}(t_{jk})}-1}\cdot\begin{cases}\frac{\varphi_i\big(\frac{\xi}{\beta},s_{j,k-1}\big)}{\varphi_i\big(\frac{\xi}{\beta},s_{jk}\big)}\,\frac{s_{j,k-1}}{s_{jk}}\left(\frac{\beta+\xi s_{jk}}{\beta+\xi s_{j,k-1}}\right)^{\frac{1}{\xi}+1}-1, & \text{if } \xi>0,\\[2ex] \frac{\varphi_i\big(0,s_{j,k-1}\big)}{\varphi_i\big(0,s_{jk}\big)}\,\frac{s_{j,k-1}}{s_{jk}}\,\mathrm{e}^{\frac{1}{\beta}\left(s_{jk}-s_{j,k-1}\right)}-1, & \text{if } \xi=0.\end{cases}$$

Because of $s_{jk} > s_{j,k-1}$ and $\varphi_2 \equiv 1$ it is equivalent

$$S_2(\xi,\beta,j,k) = 0 \quad\Leftrightarrow\quad \frac{s_{jk}}{s_{j,k-1}} = \begin{cases} \left(\frac{\beta+\xi s_{jk}}{\beta+\xi s_{j,k-1}}\right)^{\frac{1}{\xi}+1}, & \text{if } \xi > 0, \\ \mathrm{e}^{\frac{1}{\beta}\left(s_{jk}-s_{j,k-1}\right)}, & \text{if } \xi = 0. \end{cases}$$
$$\Leftrightarrow\quad \beta = \beta^*_{jk}(\xi),$$

where β^*_{jk} is the function defined in the proposition. This means that $S_2(\xi,\cdot,j,k)$ has a unique root for any fixed $\xi \in \mathbb{R}_{\geq 0}$, $j \in \mathbb{N}_{\leq m}$ and $k \in \{2,\ldots,d-1\}$.

The same applies to $S_1(\xi,\cdot,j,k)$. Due to $\varphi_1(0,a) = \frac{a}{2}$ for any $a \in \mathbb{R}_{\geq 0}$, it is equivalent

$$S_1(0,\beta,j,k) = 0 \quad\Leftrightarrow\quad \left(\frac{s_{jk}}{s_{j,k-1}}\right)^2 = \mathrm{e}^{\frac{1}{\beta}\left(s_{jk}-s_{j,k-1}\right)} \quad\Leftrightarrow\quad \beta = \frac{1}{2}\,\beta^*_{jk}(0).$$

Furthermore, in case of $\xi > 0$ the functions

$$\eta_{jk}\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}\colon\ x \longmapsto \frac{\log\left(\frac{1+xs_{jk}}{1+xs_{j,k-1}}\right)}{\log\left(\frac{\frac{1}{s_{j,k-1}}+x}{\frac{1}{s_{jk}}+x}\right) + \log\left(\frac{\varphi_1\left(x,s_{jk}\right)}{\varphi_1\left(x,s_{j,k-1}\right)}\right)}$$

are needed, because now it is equivalent

$$S_1(\xi,\beta,j,k) = 0 \quad\Leftrightarrow\quad \frac{\varphi_1\left(\frac{\xi}{\beta},s_{jk}\right)}{\varphi_1\left(\frac{\xi}{\beta},s_{j,k-1}\right)}\,\frac{s_{jk}}{s_{j,k-1}} = \left(\frac{1+\frac{\xi}{\beta}s_{jk}}{1+\frac{\xi}{\beta}s_{j,k-1}}\right)^{\frac{1}{\xi}+1}$$
$$\Leftrightarrow\quad \eta_{jk}\left(\frac{\xi}{\beta}\right) = \xi$$
$$\Leftrightarrow\quad \beta = \frac{\xi}{\eta^{-1}_{jk}(\xi)}.$$

The inverse function η^{-1}_{jk} of η_{jk} exists, because $\eta_{jk}(x) \xrightarrow{x\to 0} 0$, $\eta_{jk}(x) \xrightarrow{x\to\infty} \infty$, and the numerator of η_{jk} is positive and strictly increasing while both addends in the denominator are positive and strictly decreasing.

All this implies that $S_1(\xi,\cdot,j,k)$ and $S_2(\xi,\cdot,j,k)$ each have got a unique root for any fixed $\xi \in \mathbb{R}_{\geq 0}$, $j \in \mathbb{N}_{\leq m}$ and $k \in \{2,\ldots,d-1\}$. Furthermore, due to

$$\lim_{\beta\to 0}\left(\frac{\varphi_i\left(\frac{\xi}{\beta},s_{j,k-1}\right)}{\varphi_i\left(\frac{\xi}{\beta},s_{jk}\right)}\,\frac{s_{j,k-1}}{s_{jk}}\left(\frac{\beta+\xi s_{jk}}{\beta+\xi s_{j,k-1}}\right)^{\frac{1}{\xi}+1} - 1\right) = \left(\frac{s_{jk}}{s_{j,k-1}}\right)^{\frac{1}{\xi}} - 1 > 0,$$

the sign of $S_i(\xi,\cdot,j,k)$ must change from positive to negative at the unique root.

In addition to this, with $s_0 := \max\{s_{j,d-1} \,|\, 1 \le j \le m\}$ the term

$$\frac{\beta\, S_i(\xi,\beta,j,k)}{\varphi_i\left(\frac{\xi}{\beta}, s_0\right)}$$

is strictly decreasing as function with respect to β. Hence, also the sums

$$T_i(\xi,\beta;z) := \sum_{j=1}^{m}\sum_{k=2}^{d-1} z_{jk}\, \frac{\beta\, S_i(\xi,\beta,j,k)}{\varphi_i\left(\frac{\xi}{\beta}, s_0\right)}$$

as functions with respect to β each have a unique root and are strictly decreasing ($i \in \{1,2\}$). Together with the terms

$$Q_i(\xi,\beta;z) := -\sum_{j=1}^{m} z_{j1}\, \frac{\varphi_i\left(\frac{\xi}{\beta}, s_{j1}\right)}{\varphi_i\left(\frac{\xi}{\beta}, s_0\right)}\, \frac{s_{j1}}{\beta+\xi s_{j1}} \cdot \begin{cases} \frac{1}{\left(1+\frac{\xi}{\beta}s_{j1}\right)^{\frac{1}{\xi}}-1}, & \text{if } \xi>0,\\ \frac{1}{e^{\frac{1}{\beta}s_{j1}}-1}, & \text{if } \xi=0, \end{cases}$$

$$P_i(\xi,\beta;z) := \sum_{j=1}^{m} z_{jd}\, \frac{\varphi_i\left(\frac{\xi}{\beta}, s_{j,d-1}\right)}{\varphi_i\left(\frac{\xi}{\beta}, s_0\right)}\, \frac{s_{j,d-1}}{\beta+\xi s_{j,d-1}},$$

the partial derivatives of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ are

$$\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \xi}(\xi,\beta;z) = \frac{\varphi_1\left(\frac{\xi}{\beta}, s_0\right)}{\beta}\Big(T_1(\xi,\beta;z)+Q_1(\xi,\beta;z)+P_1(\xi,\beta;z)\Big),$$
$$\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi,\beta;z) = \frac{1}{\beta}\Big(T_2(\xi,\beta;z)+Q_2(\xi,\beta;z)+P_2(\xi,\beta;z)\Big)$$

(see Lemma 4.4.2). Since the terms Q_i as functions with respect to β are strictly decreasing and tend to 0 for $\beta \to 0$ (or they are identically zero), while the terms P_i are strictly decreasing and tend to 0 for $\beta \to \infty$ (or they are identically zero), the derivatives $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \xi}(\xi,\cdot;z)$ and $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi,\cdot;z)$ each have got a unique root, and their signs change from positive to negative at these roots.

2.: Again, have a look at the addends $S_i(\xi,\beta,j,k)$ of the partial derivatives of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ as defined in the proof of the first statement of this proposition. If there is a $b \in \{2,\ldots,d-1\}$ such that $\sum_{j=1}^{m}\sum_{k=b+1}^{d} z_{jk} = 0$, the partial derivative of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ with respect to β is (see Lemma 4.4.2)

$$\begin{aligned}\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi,\beta;z) &= \sum_{j=1}^{m}\left(\sum_{k=2}^{b} z_{jk}\, S_2(\xi,\beta,j,k) - z_{j1}\, \frac{\frac{s_{j1}}{\beta^2+\xi\beta s_{j1}}}{\frac{1}{F^*_{\mathrm{sev}}(t_{j1})}-1}\right)\\ &\le \sum_{j=1}^{m}\sum_{k=2}^{b} z_{jk}\, S_2(\xi,\beta,j,k).\end{aligned}$$

In the proof of the first statement above the term $\beta^*_{jk}(\xi)$ is identified to be the unique root of $S_2(\xi, \cdot, j, k)$. Consequently, it must be

$$\beta^*(\xi) \leq \max_{\substack{1 \leq j \leq m \\ 2 \leq k \leq b}} \beta^*_{jk}(\xi),$$

because if β is larger than the greatest single root $\beta^*_{jk}(\xi)$, all terms $S_2(\xi, \beta, j, k)$ are negative already. The $\beta^*_{jk}(\xi)$ can be rewritten as

$$\beta^*_{jk}(\xi) = \xi s_{jk} \, \frac{1 - \exp\Bigl(\frac{1}{\xi+1} \log\Bigl(\frac{s_{jk}}{s_{j,k-1}}\Bigr)\Bigr)}{\exp\Bigl(\frac{\xi}{\xi+1} \log\Bigl(\frac{s_{jk}}{s_{j,k-1}}\Bigr)\Bigr) - 1},$$

and results from the appendix (see Lemma A.3) ensure that

$$\beta^*_{jk}(\xi) \in \left[\frac{\log\Bigl(\frac{s_{jk}}{s_{j,k-1}}\Bigr)}{\frac{1}{s_{j,k-1}} - \frac{1}{s_{jk}}}, \; \frac{s_{jk} - s_{j,k-1}}{\log\Bigl(\frac{s_{jk}}{s_{j,k-1}}\Bigr)} \right] \subseteq [s_{j,k-1}, s_{jk}] \, .$$

If there is an $a \in \{2, \ldots, d-1\}$ such that $\sum_{j=1}^{m} \sum_{k=1}^{a-1} z_{jk} = 0$, the argumentation is the same, because here the partial derivative of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ with respect to β is (see Lemma 4.4.2)

$$\begin{aligned} \tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi, \beta; z) &= \sum_{j=1}^{m} \left(\sum_{k=a}^{d-1} z_{jk} \, S_2(\xi, \beta, j, k) + z_{jd} \, \frac{s_{j,d-1}}{\beta^2 + \xi \beta s_{j,d-1}} \right) \\ &\geq \sum_{j=1}^{m} \sum_{k=a}^{d-1} z_{jk} \, S_2(\xi, \beta, j, k). \end{aligned}$$

3.: The proof of this part is in line with Orme and Ruud [OR02] adapted to the special situation here. First of all, define for each $\xi \in \mathbb{R}_{\geq 0}$ the function

$$f_\xi \colon \mathbb{R}_{>0} \to \mathbb{R} \colon \; \beta \longmapsto \ell^{\mathrm{C}}_{\mathrm{sev}}(\xi, \beta; z).$$

The first statement of this proposition indicates that for each $\xi \in \mathbb{R}_{\geq 0}$ there is a unique point $\beta^*(\xi) \in \mathbb{R}_{>0}$ satisfying

$$\tfrac{\mathrm{d} f_\xi}{\mathrm{d} \beta}(\beta^*(\xi)) = \tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi, \beta^*(\xi); z) = 0.$$

The first statement also yields that the sign of f_ξ changes from positive to negative at $\beta^*(\xi)$. Hence, $\beta^*(\xi)$ is not only the unique stationary point of f_ξ, but it also is a maximizer of f_ξ and therefore a global maximizer.

Now, have a look at

$$g \colon \mathbb{R}_{\geq 0} \to \mathbb{R} \colon \; \xi \longmapsto f_\xi(\beta^*(\xi)) = \max_{\beta \in \mathbb{R}_{>0}} f_\xi(\beta).$$

The Implicit Function Theorem [For99, pp. 68–71] ensures that $\xi \mapsto \beta^*(\xi)$ is continuously differentiable. Thus, the first derivative[4] of g is

$$\tfrac{\mathrm{d}g}{\mathrm{d}\xi}(\xi) = \tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \xi}(\xi, \beta^*(\xi); z) + \tfrac{\mathrm{d}\beta^*}{\mathrm{d}\xi}(\xi)\, \tfrac{\mathrm{d} f_\xi}{\mathrm{d}\beta}(\beta^*(\xi)) = \tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \xi}(\xi, \beta^*(\xi); z).$$

Hence, ξ is a root of the derivative of g if and only if $(\xi, \beta^*(\xi))$ is a stationary point of $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,; z)$. Consequently, it holds

$$\Big\{(\xi,\beta) \in \Theta_{\mathrm{sev}} \,\Big|\, \mathrm{grad}\Big(\ell^{\mathrm{C}}_{\mathrm{sev}}(\xi,\beta;z)\Big) = 0\Big\} = \Big\{(\xi,\beta) \in \Theta_{\mathrm{sev}} \,\Big|\, \beta = \beta^*(\xi),\ \tfrac{\mathrm{d}g}{\mathrm{d}\xi}(\xi) = 0\Big\}.$$

The only thing remaining to be done is to show that $\frac{\mathrm{d}g}{\mathrm{d}\xi}$ has got at most one root.

For this goal, keep in mind that

$$\varphi_1\Big(\tfrac{\xi}{\beta}, a\Big)\, \frac{a}{\beta^2 + \xi\beta a} = \begin{cases} \frac{1}{\xi^2} \log\Big(1 + \frac{\xi}{\beta} a\Big) - \frac{\beta}{\xi}\, \frac{a}{\beta^2 + \xi\beta a}, & \text{if } \xi > 0, \\ \frac{a}{2}\, \frac{a}{\beta^2 + \xi\beta a}, & \text{if } \xi = 0. \end{cases}$$

Hence, with the definition

$$h(\xi,\beta;z) := \sum_{j=1}^{m} \sum_{k=1}^{d-1} z_{jk} \left(\frac{\log\Big(1 + \frac{\xi}{\beta} s_{j,k-1}\Big)}{1 - \frac{F^*_{\mathrm{sev}}(t_{jk})}{F^*_{\mathrm{sev}}(t_{j,k-1})}} - \frac{\log\Big(1 + \frac{\xi}{\beta} s_{jk}\Big)}{\frac{F^*_{\mathrm{sev}}(t_{j,k-1})}{F^*_{\mathrm{sev}}(t_{jk})} - 1} \right) + \sum_{j=1}^{m} z_{jd} \log\Big(1 + \tfrac{\xi}{\beta} s_{j,d-1}\Big)$$

the partial derivative of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ with respect to ξ is

$$\tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \xi}(\xi,\beta;z) = \frac{1}{\xi^2}\, h(\xi,\beta;z) - \frac{\beta}{\xi}\, \tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi,\beta;z) \qquad \forall \xi, \beta \in \mathbb{R}_{>0},$$

and therefore

$$\tfrac{\mathrm{d}g}{\mathrm{d}\xi}(\xi) = \frac{1}{\xi^2}\, h(\xi, \beta^*(\xi); z) \qquad \forall \xi, \beta \in \mathbb{R}_{>0}.$$

Let ξ_0 be a root of $\frac{\mathrm{d}g}{\mathrm{d}\xi}$. If $\xi_0 \in \mathbb{R}_{>0}$, the definitions of $\beta^*(\xi_0)$ and h ensure

$$\tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}(\xi_0, \beta^*(\xi_0); z) = 0 = h(\xi_0, \beta^*(\xi_0); z).$$

The terms $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}$ (or rather $\beta^2 \frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}$) and h are very similar to each other. They only differ in the terms

$$\frac{s_{jk}}{1 + \frac{\xi}{\beta} s_{jk}} \qquad \text{and} \qquad \log\Big(1 + \tfrac{\xi}{\beta} s_{jk}\Big).$$

[4] at $\xi = 0$ this means the right derivative

While $\frac{s_{jk}}{1+xs_{jk}}$ is strictly decreasing, $\log(1+xs_{jk})$ is strictly increasing (both as functions with respect to x). And $\frac{s_{jk}}{1+xs_{jk}}$ decreases even faster than $\frac{s_{j,k-1}}{1+xs_{j,k-1}}$, while $\log(1+xs_{jk})$ increases faster than $\log(1+xs_{j,k-1})$. A variation of ξ_0, $\xi_0 \to \xi_0 + \mathrm{d}\xi$, increases $\frac{\xi_0}{\beta^*(\xi_0)}$,

$$\frac{\xi_0}{\beta^*(\xi_0)} < \frac{\xi_0 + \mathrm{d}\xi}{\beta^*(\xi_0 + \mathrm{d}\xi)},$$

because $\beta^*(\xi)$ is decreasing as function with respect to ξ (all the $\beta^*_{jk}(\xi)$ are decreasing, see Lemma A.3 in the appendix). But it still keeps by definition

$$\tfrac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial \beta}\big(\xi_0 + \mathrm{d}\xi, \beta^*(\xi_0 + \mathrm{d}\xi); z\big) = 0.$$

At the same time, the different behavior of $\frac{s_{jk}}{1+\frac{\xi_0}{\beta^*(\xi_0)}s_{jk}}$ and $\log\Big(1 + \frac{\xi_0}{\beta^*(\xi_0)}s_{jk}\Big)$ must lead to

$$h\big(\xi_0 + \mathrm{d}\xi, \beta^*(\xi_0 + \mathrm{d}\xi); z\big) < 0.$$

The fact $h\big(\xi_0 - \mathrm{d}\xi, \beta^*(\xi_0 - \mathrm{d}\xi); z\big) > 0$ can be verified in the same way.

All this means that the sign of $\frac{\mathrm{d}g}{\mathrm{d}\xi}$ changes from positive to negative at all its positive roots. For continuity reasons this is even true if $\xi_0 = 0$. In other words, it holds

$$\tfrac{\mathrm{d}^2 g}{\mathrm{d}\xi^2}(\xi) < 0 \qquad \forall \xi \in \Big\{x \in \mathbb{R}_{\geq 0} \,\Big|\, \tfrac{\mathrm{d}g}{\mathrm{d}\xi}(x) = 0\Big\}.$$[5]

Hence, every stationary point of g automatically is a maximizer of g and the maximum is an isolated one. Assume that g has more than one stationary point and ξ_1 and ξ_2 are two of them. Then, due to the Theorem of Maximum and Minimum for Continuous Functions (sometimes Weierstrass' Theorem) [For04, p. 106] there is a $\xi_3 \in (\xi_1, \xi_2)$ such that $g(\xi_3) \leq g(\xi)$ for all $\xi \in [\xi_1, \xi_2]$. This ξ_3 must satisfy

$$\tfrac{\mathrm{d}g}{\mathrm{d}\xi}(\xi_3) = 0 \qquad \text{and} \qquad \tfrac{\mathrm{d}^2 g}{\mathrm{d}\xi^2}(\xi_3) \geq 0.$$

But it has been shown, that such a ξ_3 does not exist.

All in all, g have at most one stationary point, and therefore $\ell^{\mathrm{C}}_{\mathrm{sev}}$ only have at most one stationary point. This point $(\xi_0, \beta^*(\xi_0))$ satisfies (if it exists)

$$\ell^{\mathrm{C}}_{\mathrm{sev}}(\xi_0, \beta^*(\xi_0); z) = g(\xi_0) > g(\xi) = \ell^{\mathrm{C}}_{\mathrm{sev}}(\xi, \beta^*(\xi); z) \geq \ell^{\mathrm{C}}_{\mathrm{sev}}(\xi, \beta; z)$$

for all $\xi \in \mathbb{R}_{\geq 0} \setminus \{\xi_0\}$ and $\beta \in \mathbb{R}_{>0}$. Consequently, $(\xi_0, \beta^*(\xi_0))$ is the unique stationary point and the maximizer of $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,; z)$. □

The results of the last proposition are just a small step away from the statement that the maximum likelihood estimator of (ξ, β) exists (i. e. there is a

[5] at $\xi = 0$ $\frac{\mathrm{d}}{\mathrm{d}\xi}$ means the right derivative

unique global maximizer of $\ell_{\text{sev}}^{\text{C}}$) if there is at least one nonempty medium class. If someone has found a stationary point of $\ell_{\text{sev}}^{\text{C}}$, Proposition 4.4.3 ensures that this is the maximum likelihood estimator. On the other hand, if there is no stationary point, $\ell_{\text{sev}}^{\text{C}}$ still has a unique maximizer. This is the statement of the following theorem.

4.4.4 Theorem. *Let* $z = (z_{jk})_{\substack{1\le j\le m\\1\le k\le d}} \in {\mathbb{N}_0}^{m\times d}$ *be a realization of* $\big(\boldsymbol{Z}_{\mathrm{v}(j)}\big)_{1\le j\le m}$ *where* $d \in \mathbb{N}_{\ge 3}$.

1. *If there is a* $(j_0, k_0) \in \mathbb{N}_{\le m} \times \{2, \dots, d-1\}$ *such that* $z_{j_0 k_0} > 0$, *then the maximum likelihood estimator of* (ξ, β) *based on* z *exists.*
2. *If either* $\sum_{j=1}^{m}\sum_{k=2}^{d} z_{jk} = 0$ *or* $\sum_{j=1}^{m}\sum_{k=1}^{d-1} z_{jk} = 0$, *then the maximum likelihood estimator of* (ξ, β) *based on* z *does not exist, since the likelihood function as function with respect to* (ξ, β) *does not have a maximum.*
3. *Suppose, the class limits are chosen such that*

$$\max_{1\le j\le m} t_{j1} < \min_{1\le j\le m} t_{j,d-1}.$$

If $\sum_{j=1}^{m}\sum_{k=2}^{d-1} z_{jk} = 0$, *then the maximum likelihood estimator of* (ξ, β) *based on* z *does not exist, since the likelihood function as function with respect to* (ξ, β) *does not have a maximum.*

Proof. **1.:** Suppose, (ξ_0, β_0) is a stationary point of $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,; z)$ with $\xi_0 \in \mathbb{R}_{>0}$, i. e. it holds $\operatorname{grad}\big(\ell_{\text{sev}}^{\text{C}}(\xi_0, \beta_0; z)\big) = 0$. The third statement in Proposition 4.4.3 provides that this stationary point is unique and that it must be the maximum likelihood estimator of (ξ, β).

Suppose, $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,; z)$ does not have any stationary point in ${\mathbb{R}_{>0}}^2$. For all $(\xi_0, \beta_0) \in \big(\mathbb{R}_{\ge 0} \times \{0, \infty\}\big) \cup \big(\{\infty\} \times \mathbb{R}_{\ge 0}\big) \cup \big(\{\infty\} \times \{\infty\}\big)$ it holds according to Lemma A.1 in the appendix

$$\lim_{\substack{\xi\to\xi_0\\ \beta\to\beta_0}} \ell_{\text{sev}}^{\text{C}}(\xi, \beta; z) = \lim_{\substack{\xi\to\xi_0\\ \beta\to\beta_0}} \sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk} \log\big(p_{A_{jk}}\big) = -\infty.$$

This means that $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,; z)$ approaches $-\infty$ on the boundary of its domain. Hence, $\ell_{\text{sev}}^{\text{C}}(\,\cdot\,; z)$ must have a maximum. A maximizer cannot be an element of ${\mathbb{R}_{>0}}^2$, because there are no stationary points. Instead, it must be an element of $\{0\} \times \mathbb{R}_{>0}$. The function $\ell_{\text{sev}}^{\text{C}}(0, \cdot\,; z)$ has a unique maximizer $\beta^*(0)$ (see first statement of Proposition 4.4.3). It follows that $(0, \beta^*(0))$ is the maximum likelihood estimator based on z.

2./3.: The proof is established if it can be shown that the shifted generalized Pareto distribution satisfies all the conditions in Theorem 4.4.1.

A glance at the shifted generalized Pareto distribution at the beginning of this Section 4.4 reveals that F_{sev} is strictly increasing and

$$F_{\mathrm{sev}}(t;\xi,\beta) < 1 \quad \forall t \in \mathbb{R} \qquad \text{and} \qquad F_{\mathrm{sev}}(t;\xi,\beta) > 0 \quad \forall t \in \mathbb{R}_{>u_{\mathrm{sev}}}.$$

Moreover, Lemma A.1 in the appendix helps to verify that for any $a \in (0,1)$ it holds

$$\lim_{\xi\to\infty} F_{\mathrm{sev}}\big(t;\xi,a^{\xi}\big) = 1-a \qquad \forall t \in \mathbb{R}_{>u_{\mathrm{sev}}}$$

and

$$\lim_{\beta\to 0} F_{\mathrm{sev}}(t;\xi,\beta) = 1 \quad \forall t \in \mathbb{R}_{>u_{\mathrm{sev}}} \qquad \text{and} \qquad \lim_{\beta\to\infty} F_{\mathrm{sev}}(t;\xi,\beta) = 0 \quad \forall t \in \mathbb{R}.$$

□

Proposition 4.4.3 provides not only the existence of a maximum likelihood estimator (see Theorem 4.4.4), but it also supplies a way to localize it. Particularly, it gives a criterion when to choose the exponential model ($\xi = 0$) or the Pareto model ($\xi \in \mathbb{R}_{>0}$). The next corollary states this criterion.

4.4.5 Corollary. *Let the situation be as in Proposition 4.4.3 with the roots $\beta^\circ(\xi)$ and $\beta^*(\xi)$ from there. Suppose, $(\hat{\xi}_m, \hat{\beta}_m)$ is the maximum likelihood estimator of (ξ,β) based on z. If $\hat{\xi}_m(z) > 0$, then, for all $\xi \in \mathbb{R}_{\geq 0}$, it is equivalent*

$$\xi \lesseqgtr \hat{\xi}_m(z) \qquad \Leftrightarrow \qquad \beta^*(\xi) \lesseqgtr \beta^\circ(\xi).$$

Moreover, it is $\hat{\xi}_m(z) = 0$ if and only if $\beta^\circ(0) \leq \beta^*(0)$.

Proof. If $\hat{\xi}_m(z) > 0$, the maximum likelihood estimator $\big(\hat{\xi}_m(z), \hat{\beta}_m(z)\big)$ is a root of the gradient of $\ell^{\mathrm{C}}_{\mathrm{sev}}(\,\cdot\,;z)$ and therefore

$$\beta^\circ(\hat{\xi}_m(z)) = \hat{\beta}_m(z) = \beta^*(\hat{\xi}_m(z)).$$

The rest of the proof needs the function g which is defined in the proof of the third statement of Proposition 4.4.3,

$$g\colon \mathbb{R}_{\geq 0} \to \mathbb{R}\colon\ \xi \longmapsto \ell^{\mathrm{C}}_{\mathrm{sev}}(\xi,\beta^*(\xi);z) = \max_{\beta\in\mathbb{R}_{>0}} \ell^{\mathrm{C}}_{\mathrm{sev}}(\xi,\beta;z).$$

The chain rule from differential calculus [For99, p. 48] yields the derivatives of g,

$$\frac{\mathrm{d}g}{\mathrm{d}\xi}(\xi) = \frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi}(\xi,\beta^*(\xi);z) + \frac{\mathrm{d}\beta^*}{\mathrm{d}\xi}(\xi)\,\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta}(\xi,\beta^*(\xi);z) = \frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi}(\xi,\beta^*(\xi);z),$$
$$\frac{\mathrm{d}^2 g}{\mathrm{d}\xi^2}(\xi) = \frac{\partial^2 \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi^2}(\xi,\beta^*(\xi);z) + \frac{\mathrm{d}\beta^*}{\mathrm{d}\xi}(\xi)\,\frac{\partial^2 \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta\partial\xi}(\xi,\beta^*(\xi);z).$$

In the proof of the third statement of Proposition 4.4.3 it is argued that

$$\frac{\mathrm{d}^2 g}{\mathrm{d}\xi^2}(\hat{\xi}_m(z)) < 0.$$

Furthermore, from the first statement of Proposition 4.4.3 it follows

$$\frac{\partial^2 \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta\partial\xi}\left(\hat{\xi}_m(z), \beta^\circ(\hat{\xi}_m(z)); z\right) < 0.$$

All together it is

$$\frac{\mathrm{d}\beta^*}{\mathrm{d}\xi}(\hat{\xi}_m(z)) > -\frac{\frac{\partial^2 \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi^2}\left(\hat{\xi}_m(z), \beta^\circ(\hat{\xi}_m(z)); z\right)}{\frac{\partial^2 \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\beta\partial\xi}\left(\hat{\xi}_m(z), \beta^\circ(\hat{\xi}_m(z)); z\right)} = \frac{\mathrm{d}\beta^\circ}{\mathrm{d}\xi}(\hat{\xi}_m(z)).$$

The last equality comes from the Implicit Function Theorem [For99, pp. 68–71].

Summarized, it holds

$$\beta^*(\hat{\xi}_m(z)) - \beta^\circ(\hat{\xi}_m(z)) = 0 \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}\xi}\left(\beta^*(\hat{\xi}_m(z)) - \beta^\circ(\hat{\xi}_m(z))\right) > 0.$$

Hence, there is an ε such that for all $\xi \in \left[\hat{\xi}_m(z) - \varepsilon,\, \hat{\xi}_m(z) + \varepsilon\right]$ it is equivalent

$$\xi \lesseqgtr \hat{\xi}_m(z) \qquad \Leftrightarrow \qquad \beta^*(\xi) \lesseqgtr \beta^\circ(\xi).$$

Since $\ell^{\mathrm{C}}_{\mathrm{sev}}$ has no other stationary points (see Proposition 4.4.3), $\xi = \hat{\xi}_m(z)$ is the only value which satisfies $\beta^\circ(\xi) = \beta^*(\xi)$. Furthermore, the functions β° and β^* are continuous due to the Implicit Function Theorem [For99, pp. 68–71]. Hence, the equivalence must hold for all $\xi \in \mathbb{R}_{\geq 0}$.

It has been proven that $\beta^\circ(0) > \beta^*(0)$ if $\hat{\xi}_m(z) > 0$. Conversely, if $\beta^\circ(0)$ is greater than $\beta^*(0)$, the first statement in Proposition 4.4.3 provides that $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi}(0, \beta^*(0); z)$ is positive. This implies the existence of an $\varepsilon \in \mathbb{R}_{>0}$ with the property

$$\ell^{\mathrm{C}}_{\mathrm{sev}}(\varepsilon, \beta^*(0); z) > \ell^{\mathrm{C}}_{\mathrm{sev}}(0, \beta^*(0); z)\,.$$

Hence, it cannot be $\hat{\xi}_m(z) = 0$ if $\beta^\circ(0) > \beta^*(0)$. □

This corollary supplies a concrete algorithm for the calculation of the maximum likelihood estimator of (ξ, β): first, compute the roots $\beta^\circ(0)$, $\beta^*(0)$ and compare them with each other. If $\beta^\circ(0) \leq \beta^*(0)$, the maximum likelihood estimator is found, $(\hat{\xi}_m, \hat{\beta}_m) = (0, \beta^*(0))$. Otherwise, take a $\xi \in \mathbb{R}_{>0}$ large enough such that $\beta^\circ(\xi) < \beta^*(\xi)$ or, equivalently, $\frac{\partial \ell^{\mathrm{C}}_{\mathrm{sev}}}{\partial\xi}(\xi, \beta^*(\xi); z) < 0$. Then, it must be $\hat{\xi}_m \in (0, \xi)$. Now, a simple bisection method can approx the actual maximum likelihood estimator $\hat{\xi}_m$ up to the desired precision. Section 5.5 works out this procedure in detail.

4.4.3. Confidence Intervals

When observing uncensored, statistically independent and generalized Pareto distributed variates, the maximum likelihood estimators of ξ and β are consistent and asymptotically efficient as long as $\xi \in \mathbb{R}_{>-\frac{1}{2}}$ [Smi84, HW87]. Also here, where only the counts per class can be observed, the numerical results in Section 5.5.2 suggest for $\xi \in \mathbb{R}_{>0}$ that $(\hat{\xi}_m, \hat{\beta}_m)$ is asymptotically efficient, i. e. the maximum likelihood estimator $(\hat{\xi}_m, \hat{\beta}_m)$ is asymptotically jointly normally distributed with mean (ξ, β) and the inverse of the Fisher information matrix as covariance matrix (see Section 2.4.3). Due to Theorem 4.2.2, the inverse of the Fisher information matrix concerning ξ and β is

$$I_{\text{sev}}(\mu, \xi, \beta)^{-1} = \frac{1}{\mu\, \gamma(\xi, \beta)} \sum_{j=1}^{m} \sum_{k=1}^{d-1} \frac{l_j}{b_{jk}(\xi, \beta)} \begin{pmatrix} a_{2jk}(\xi, \beta)^2 & -a_{1jk}(\xi, \beta)\, a_{2jk}(\xi, \beta) \\ -a_{1jk}(\xi, \beta)\, a_{2jk}(\xi, \beta) & a_{1jk}(\xi, \beta)^2 \end{pmatrix},$$

with a_{ijk}, b_{jk} as defined there and

$$\gamma(\xi, \beta) := \prod_{i=1}^{2} \left(\sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{ijk}(\xi, \beta)^2}{b_{jk}(\xi, \beta)} \right) - \left(\sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{1jk}(\xi, \beta)\, a_{2jk}(\xi, \beta)}{b_{jk}(\xi, \beta)} \right)^2 .$$

Thus, asymptotic efficiency means that $\hat{\xi}_m$ is asymptotically $\mathcal{N}(\xi, {\sigma_{\xi,m}}^2)$ distributed and $\hat{\beta}_m$ is asymptotically $\mathcal{N}(\beta, {\sigma_{\beta,m}}^2)$ distributed, where

$$\begin{aligned} {\sigma_{\xi,m}}^2 &:= \frac{1}{\mu\, \gamma(\xi, \beta)} \sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{2jk}(\xi, \beta)^2}{b_{jk}(\xi, \beta)}, \\ {\sigma_{\beta,m}}^2 &:= \frac{1}{\mu\, \gamma(\xi, \beta)} \sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{1jk}(\xi, \beta)^2}{b_{jk}(\xi, \beta)}. \end{aligned}$$

With $\boldsymbol{Z} := (\boldsymbol{Z}_{\mathrm{v}(j)})_{1 \le j \le m}$ this leads to

$$\begin{aligned} \lim_{m \to \infty} \mathbb{P}_\vartheta\left(-q_{1-\alpha/2} \le \tfrac{\hat{\xi}_m(\boldsymbol{Z}) - \xi}{\hat{\sigma}_{\xi,m}(\boldsymbol{Z})} \le q_{1-\alpha/2} \right) &= 1 - \alpha, \\ \lim_{m \to \infty} \mathbb{P}_\vartheta\left(-q_{1-\alpha/2} \le \tfrac{\hat{\beta}_m(\boldsymbol{Z}) - \beta}{\hat{\sigma}_{\beta,m}(\boldsymbol{Z})} \le q_{1-\alpha/2} \right) &= 1 - \alpha, \end{aligned}$$

where $q_{1-\alpha/2}$ is the $\left(1 - \frac{\alpha}{2}\right) 100\,\%$ quantile of the standard normal distribution, and $\hat{\sigma}_{\xi,m}(\boldsymbol{Z})$ equates $\sigma_{\xi,m} = \sqrt{{\sigma_{\xi,m}}^2}$ while $\hat{\sigma}_{\beta,m}(\boldsymbol{Z})$ equates $\sigma_{\beta,m} = \sqrt{{\sigma_{\beta,m}}^2}$ with μ, ξ and β replaced by the estimators $\hat{\mu}_m(\boldsymbol{Z})$, $\hat{\xi}_m(\boldsymbol{Z})$ and $\hat{\beta}_m(\boldsymbol{Z})$ respectively. Eventually, if z is a realization of $\boldsymbol{Z}$, the intervals

$$\begin{aligned} C_\xi(\alpha, z) &:= \left[\hat{\xi}_m(z) - \hat{\sigma}_{\xi,m}(z)\, q_{1-\alpha/2},\ \hat{\xi}_m(z) + \hat{\sigma}_{\xi,m}(z)\, q_{1-\alpha/2} \right], \\ C_\beta(\alpha, z) &:= \left[\hat{\beta}_m(z) - \hat{\sigma}_{\beta,m}(z)\, q_{1-\alpha/2},\ \hat{\beta}_m(z) + \hat{\sigma}_{\beta,m}(z)\, q_{1-\alpha/2} \right] \end{aligned} \tag{4.7}$$

are approximate actual confidence intervals of ξ and β with confidence level $(1-\alpha)$ $(\alpha \in (0,1))$.

If it is $\xi = 0$, the term ${(\hat{\xi}_m(\boldsymbol{Z})-\xi)}/{\sigma_{\xi,m}} = {\hat{\xi}_m(\boldsymbol{Z})}/{\sigma_{\xi,m}}$ cannot be asymptotically standard normally distributed, because $\hat{\xi}_m$ is bounded below by 0. Since the anti-diagonal of the Fisher information matrix is nonzero, the boundedness of $\hat{\xi}_m$ also affects the distribution of $\hat{\beta}_m(\boldsymbol{Z})$. The following heuristic approach shall yield approximate confidence intervals of ξ and β in case of $\xi = 0$.

For this purpose, assume for a while that also negative shape parameters are allowed. Due to the asymptotic efficiency, the maximum likelihood estimator of ξ is asymptotically normally distributed with mean $\xi = 0$. Consequently, the true parameter $\xi = 0$ is underestimated and overestimated each with probability $1/2$ if the sample size m is large. Every realization z of $(\boldsymbol{Z}_{\mathrm{v}(j)})_{1 \le j \le m}$ which causes an underestimation of $\xi = 0$ in the entire generalized Pareto model yields $\hat{\xi}_m(z) = 0$ in the counting model $\mathcal{E}_{\mathrm{C}}$. Hence, for those realizations the cumulative distribution function of $\hat{\xi}_m$ is

$$\Phi_{\xi-}(t) := \mathbb{1}_{\mathbb{R}_{\ge 0}}(t) \qquad \forall t \in \mathbb{R}.$$

Basically, ξ is simply known and β is the only unknown parameter. Thus, for those realizations $\hat{\beta}_m$ is asymptotically normally distributed with mean β and variance ${\tau_{\beta,m}}^2$, where ${\tau_{\beta,m}}^2$ equates the inverse of the Fisher information concerning β,

$${\tau_{\beta,m}}^2 := \frac{1}{I_{\mathrm{sev}}(\mu,\xi,\beta)_{22}} = \frac{1}{\mu \sum_{j=1}^{m} l_j \sum_{k=1}^{d-1} \frac{a_{2jk}(\xi,\beta)^2}{b_{jk}(\xi,\beta)}}$$

(see Theorem 4.2.2). Hence, the cumulative distribution function of $\hat{\beta}_m$ is

$$\Phi_{\beta-}(t) := \int_{-\infty}^{t} \frac{1}{\tau_{\beta,m}\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\beta}{\tau_{\beta,m}}\right)^2} \,\mathrm{d}x \qquad \forall t \in \mathbb{R}.$$

On the other hand, a realization z of $(\boldsymbol{Z}_{\mathrm{v}(j)})_{1 \le j \le m}$ which causes an overestimation of $\xi = 0$ in the entire generalized Pareto model yields $\hat{\xi}_m(z) > 0$ in the counting model $\mathcal{E}_{\mathrm{C}}$. For all these realizations, $\hat{\xi}_m$ is asymptotically truncated normally distributed with lower bound 0,

$$\Phi_{\xi+}(t) := 2\int_{0}^{\max\{t,0\}} \frac{1}{\sigma_{\xi,m}\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\sigma_{\xi,m}}\right)^2} \,\mathrm{d}x \qquad \forall t \in \mathbb{R}.$$

Since the anti-diagonal of the inverse Fisher information matrix $I_{\mathrm{sev}}(\mu,\xi,\beta)^{-1}$ has negative entries, $\hat{\xi}_m$ and $\hat{\beta}_m$ are negative correlated for large sample sizes. As a consequence, $\hat{\beta}_m(z)$ tends to underestimate the true scale parameter β, because $\hat{\xi}_m(z)$ overestimates the true shape $\xi = 0$. Therefore, it is worse trying to assume that $\hat{\beta}_m$ is truncated normally distributed with upper bound β. So, it has the cumulative distribution function

$$\Phi_{\beta+}(t) := 2 \int_{-\infty}^{\min\{t,\beta\}} \frac{1}{\sigma_{\beta,m}\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\beta}{\sigma_{\beta,m}}\right)^2} \,\mathrm{d}x \qquad \forall t \in \mathbb{R}.$$

Near the true scale β the approximation of the distribution of $\hat{\beta}_m$ through $\Phi_{\beta+}$ is expected to be bad, because $\hat{\beta}_m$ is not really bounded from above. However, the results presented in Section 5.5.3 offer that this approach yields an adequate approximation of the extreme quantiles.

Eventually, the approximate cumulative distribution functions Φ_ξ of $\hat{\xi}_m$ and Φ_β of $\hat{\beta}_m$ are

$$\begin{aligned} \Phi_\xi(t) :=& \frac{\Phi_{\xi-}(t) + \Phi_{\xi+}(t)}{2} = \frac{\mathbb{1}_{\mathbb{R}_{\geq 0}}(t)}{\sigma_{\xi,m}\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\sigma_{\xi,m}}\right)^2} \,\mathrm{d}x, \\ \Phi_\beta(t) :=& \frac{\Phi_{\beta-}(t) + \Phi_{\beta+}(t)}{2} \\ =& \int_{-\infty}^{t} \left(\frac{\mathbb{1}_{\mathbb{R}_{\leq\beta}}(x)}{\sigma_{\beta,m}\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\beta}{\sigma_{\beta,m}}\right)^2} + \frac{0.5}{\tau_{\beta,m}\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\beta}{\tau_{\beta,m}}\right)^2} \right) \mathrm{d}x, \end{aligned} \tag{4.8}$$

If Φ_ξ^{-1} denotes the inverse of $\Phi_\xi \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 1/2}$, and Φ_β^{-1} denotes the inverse of Φ_β, then it holds

$$\Phi_\xi^{-1}(1-\alpha) = \sigma_{\xi,m}\, q_{1-\alpha} \quad \text{and} \quad \Phi_\beta^{-1}\left(1 - \tfrac{\alpha}{2}\right) = \beta + \tau_{\beta,m}\, q_{1-\alpha} \qquad \forall \alpha \in (0, 0.5],$$

where $q_{1-\alpha}$ is the $(1-\alpha)\,100\,\%$ quantile of the standard normal distribution. All in all, the intervals

$$\begin{aligned} C_\xi^0(\alpha, z) &:= \left[0,\ \hat{\sigma}_{\xi,m}(z)\, q_{1-\alpha}\right], \\ C_\beta^0(\alpha, z) &:= \left[\hat{\Phi}_\beta^{-1}\left(\tfrac{\alpha}{2}\right),\ \hat{\beta}_m(z) + \hat{\tau}_{\beta,m}(z)\, q_{1-\alpha}\right] \end{aligned} \tag{4.9}$$

can be taken as approximate actual confidence intervals of ξ and β with confidence level $(1-\alpha)$ if $\hat{\xi}_m$ is estimated to be $\hat{\xi}_m = 0$ ($\alpha \in (0, 0.5]$), where $\hat{\Phi}_\beta^{-1}$, $\hat{\sigma}_{\xi,m}(z)$, $\hat{\sigma}_{\beta,m}(z)$ and $\hat{\tau}_{\beta,m}(z)$ equate Φ_β^{-1}, $\sigma_{\xi,m}(z)$, $\sigma_{\beta,m}(z)$ and $\tau_{\beta,m}(z)$, respectively, with μ, ξ and β replaced by the estimators $\hat{\mu}_m(z)$, $\hat{\xi}_m(z)$ and $\hat{\beta}_m(z)$, respectively.

4.4.4. Equidistant Class Limits and Optimal Class Length

The *BMW Group*'s study which provides the data for distribution fitting (see Section 2.1) is based on an experimental design with classes of equal length.

Accordingly, there is a $\Lambda \in \mathbb{R}_{>0}$ such that

$$\Lambda = t_{jk} - t_{j,k-1} \qquad \forall j \in \mathbb{N}_{\leq m},\ \forall k \in \mathbb{N}_{\leq d-1}.$$

Λ is simply called **class length**. In this situation the class limits can be expressed as

$$t_{jk} = u_{\text{sev}} + k\Lambda \qquad \forall j \in \mathbb{N}_{\leq m},\ \forall k \in \mathbb{N}_{\leq d-1}.$$

This choice of class limits reveals an advantage when deciding whether the exponential model ($\xi = 0$) or the Pareto model ($\xi \in \mathbb{R}_{>0}$) should be used. Corollary 4.4.5 provides that the maximum likelihood estimator $\hat{\xi}_m$ is equal to 0 if and only if the root of $\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \xi}(0,\cdot\,;z)$ does not exceed the root of $\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \beta}(0,\cdot\,;z)$. In the counting model with equidistant class limits these roots can be calculated analytically. The root of $\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \beta}(0,\cdot\,;z)$ was also found by Kulldorff [Kul61, p. 28].

4.4.6 Lemma. *Let $z = (z_{jk})_{\substack{1\leq j\leq m\\ 1\leq k\leq d}} \in {\mathbb{N}_0}^{m\times d}$ be a realization of $(\boldsymbol{Z}_{\mathrm{v}(j)})_{1\leq j\leq m}$ such that $\sum_{j=1}^{m}\sum_{k=1}^{d-1} z_{jk} > 0$ and $\sum_{j=1}^{m}\sum_{k=2}^{d} z_{jk} > 0$. If the class lengths are all equal to Λ ($\Lambda \in \mathbb{R}_{>0}$), the points*

$$\beta_0^{\circ} := \frac{\Lambda}{\log\left(1 + \frac{\sum_{j=1}^{m}\sum_{k=1}^{d-1}(2k-1)z_{jk}}{\sum_{j=1}^{m}\sum_{k=2}^{d}(k-1)^2 z_{jk}}\right)}, \qquad \beta_0^{*} := \frac{\Lambda}{\log\left(1 + \frac{\sum_{j=1}^{m}\sum_{k=1}^{d-1} z_{jk}}{\sum_{j=1}^{m}\sum_{k=2}^{d}(k-1) z_{jk}}\right)}$$

are the unique roots of $\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \xi}(0,\cdot\,;z)$ and $\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \beta}(0,\cdot\,;z)$ respectively, i. e.

$$\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \xi}(0,\beta_0^{\circ};z) = 0 \qquad \text{and} \qquad \frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \beta}(0,\beta_0^{*};z) = 0,$$

and both derivatives as function with respect to β change their sign from positive to negative at the particular root.

Proof. If it is chosen $\xi = 0$ and $t_{jk} = u_{\text{sev}} + k\Lambda$ for all $j \in \mathbb{N}_{\leq m}$ and $k \in \mathbb{N}_{\leq d-1}$, then

$$\frac{1 - F_{\text{sev}}(t_{jk})}{1 - F_{\text{sev}}(t_{j,k-1})} = \frac{1 - F_{\text{sev}}(u_{\text{sev}} + k\Lambda)}{1 - F_{\text{sev}}(u_{\text{sev}} + (k-1)\Lambda)} = \mathrm{e}^{-\frac{1}{\beta}\Lambda} \quad \forall j \in \mathbb{N}_{\leq m}, \forall k \in \mathbb{N}_{\leq d-1}.$$

Hence, the partial derivatives of $\ell^{\mathrm{C}}_{\text{sev}}$ given in Lemma 4.4.2 can be rewritten as

$$\begin{aligned}
\frac{\partial \ell^{\mathrm{C}}_{\text{sev}}}{\partial \xi}(0,\beta;z) &= \frac{\Lambda^2}{2\beta^2}\sum_{j=1}^{m}\left(\sum_{k=1}^{d-1} z_{jk}\left(\frac{(k-1)^2}{1-\mathrm{e}^{-\frac{1}{\beta}\Lambda}} - \frac{k^2}{\mathrm{e}^{\frac{1}{\beta}\Lambda}-1}\right) + z_{jd}(d-1)^2\right)\\
&= \frac{\Lambda^2}{2\beta^2(\mathrm{e}^{\frac{1}{\beta}\Lambda}-1)}\left(\sum_{j=1}^{m}\sum_{k=2}^{d}(k-1)^2 z_{jk}\right)\left(\mathrm{e}^{\frac{1}{\beta}\Lambda} - \mathrm{e}^{\frac{1}{\beta_0^{\circ}}\Lambda}\right),
\end{aligned}$$

and

$$\frac{\partial \ell_{\mathrm{sev}}^{\mathrm{C}}}{\partial \beta}(0,\beta;z) = \frac{\Lambda}{\beta^2}\sum_{j=1}^{m}\left(\sum_{k=1}^{d-1} z_{jk}\left(\frac{k-1}{1-\mathrm{e}^{-\frac{1}{\beta}\Lambda}} - \frac{k}{\mathrm{e}^{\frac{1}{\beta}\Lambda}-1}\right) + z_{jd}(d-1)\right)$$
$$= \frac{\Lambda}{\beta^2(\mathrm{e}^{\frac{1}{\beta}\Lambda}-1)}\left(\sum_{j=1}^{m}\sum_{k=2}^{d}(k-1)z_{jk}\right)\left(\mathrm{e}^{\frac{1}{\beta}\Lambda} - \mathrm{e}^{\frac{1}{\beta_0^*}\Lambda}\right).$$

Thus, the derivatives both have each a unique root and they chance their sign from positiv to negative at the particular root.

□

It is not difficult to compare the roots β_0° and β_0^* from Lemma 4.4.6 with each other and to apply Corollary 4.4.5 thereafter. The result of this forms the following corollary.

4.4.7 Corollary. *Suppose it is $d \in \mathbb{N}_{\geq 3}$. Let the situation be as in Lemma 4.4.6 with the roots β_0° and β_0^* from there, then the following statements are equivalent:*

(i) The root of $\frac{\partial \ell_{\mathrm{sev}}^{\mathrm{C}}}{\partial \xi}(0,\cdot\,;z)$ is not greater than the root of $\frac{\partial \ell_{\mathrm{sev}}^{\mathrm{C}}}{\partial \beta}(0,\cdot\,;z)$,

$$\beta_0^\circ \leq \beta_0^*.$$

(ii) It holds

$$\begin{pmatrix} y_1 & \dots & y_{d-1}\end{pmatrix}\begin{pmatrix} c_{11} & \dots & c_{1,d-1} \\ \vdots & \ddots & \vdots \\ c_{d-1,1} & \dots & c_{d-1,d-1}\end{pmatrix}\begin{pmatrix} y_2 \\ \vdots \\ y_d\end{pmatrix} \leq 0,$$

where

$$y_k := \sum_{j=1}^{m} z_{jk} \quad \text{and} \quad c_{kk'} := (1-2k+k')k' \qquad \forall k,k' \in \mathbb{N}_{\leq d-1}.$$

(iii) The maximum likelihood estimator $(\hat{\xi}_m, \hat{\beta}_m)$ of (ξ, β) based on z is given by

$$\big(\hat{\xi}_m(z), \hat{\beta}_m(z)\big) = (0, \beta_0^*).$$

Proof. (**i**) ⇔ (**ii**): From the definitions of β_0° and β_0^* in Lemma 4.4.6 it follows that the condition $\beta_0^\circ \leq \beta_0^*$ is equivalent to the two equivalent relations

$$\frac{\sum_{k=1}^{d-1}(2k-1)y_k}{\sum_{k'=1}^{d-1} k'^2 y_{k'+1}} \geq \frac{\sum_{k=1}^{d-1} y_k}{\sum_{k'=1}^{d-1} k' y_{k'+1}} \quad \Leftrightarrow \quad 0 \geq \sum_{k=1}^{d-1}\sum_{k'=1}^{d-1} k'(k'-2k+1)\, y_k\, y_{k'+1}.$$

The right-hand side of the last inequality is nothing else than the result of the matrix product in the second statement.

(iii) $\Leftrightarrow$ **(i)**: If $\sum_{j=1}^{m}\sum_{k=2}^{d-1} z_{jk} = 0$, it is always

$$\beta_0^{\circ} = \frac{\Lambda}{\log\Big(1+\frac{y_1}{(d-1)^2\,y_d}\Big)} > \frac{\Lambda}{\log\Big(1+\frac{y_1}{(d-1)\,y_d}\Big)} = \beta_0^{*}.$$

On the other hand, due to Theorem 4.4.4, the maximum likelihood estimator does not exist.

Otherwise, the maximum likelihood estimator exists (see Theorem 4.4.4). Everything else follows from Corollary 4.4.5. □

This corollary yields some interesting conclusions. Suppose, for instance, that only the two lowest classes are filled, i. e. $\sum_{j=1}^{m}\sum_{k=3}^{d} z_{jk} = 0$, then the matrix product in the second statement of corollary 4.4.7 is

$$\left(\sum_{j=1}^{m} z_{j1}\right)\left(\sum_{j=1}^{m} z_{j2}\right) c_{11} + \left(\sum_{j=1}^{m} z_{j2}\right)^2 c_{21} = -2\left(\sum_{j=1}^{m} z_{j2}\right)^2 < 0.$$

Consequently, in this situation the maximum likelihood method always prefers the exponential model, $\big(\hat{\xi}_m(z), \hat{\beta}_m(z)\big) = (0, \beta_0^*)$. If only one single class is filled, i. e. $\sum_{j=1}^{m}\sum_{k=1}^{d} z_{jk} = \sum_{j=1}^{m} z_{jk_0}$ for an $k_0 \in \{2,\dots,d-1\}$, the exponential model is preferred, too, because then the matrix product in the second statement of corollary 4.4.7 is

$$\left(\sum_{j=1}^{m} z_{jk_0}\right)^2 c_{k_0,k_0-1} = -k_0(k_0-1)\left(\sum_{j=1}^{m} z_{jk_0}\right)^2 < 0.$$

In both situations there is not enough information about the shape of F_{sev}, and so the shape parameter ξ is estimated to be 0.

Besides the possibility of calculating the roots of the partial derivatives of $\ell_{\text{sev}}^{\text{C}}$ analytically, a second advantage with regard to equidistant class limits is the existence of some kind of optimal class length. In this context, optimality means that the accuracy of estimate can be optimized by choosing a certain class length. Suppose, the chosen class length is extremely high, $\Lambda \to \infty$, then, with probability tending to one, any SOLE will be observed in the lowest class, $(u_{\text{sev}}, u_{\text{sev}} + \Lambda]$. On the other hand, in case of $\Lambda \to 0$, only the highest class, $(u_{\text{sev}} + (d-1)\Lambda, \infty)$, has the chance to get filled. In both cases there is not enough information to estimate (ξ, β) (see Theorem 4.4.4). This fact is also reflected by the Fisher information matrix concerning ξ and β, $I_{\text{sev}}(\mu,\xi,\beta)$ (see Theorem 4.2.2), because with equidistant class limits it holds

$$\lim_{\Lambda\to 0,\infty} I_{\text{sev}}(\mu,\xi,\beta) = \mu \sum_{j=1}^{m} l_j \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

Indeed, each entry of the 2×2 matrix $I_{\text{sev}}(\mu,\xi,\beta)$ as function with respect to Λ grows up until a unique maximum is achieved and decreases to 0 again as Λ approaches ∞. As long as $d\in\mathbb{N}_{\geq 3}$, the determinant of $I_{\text{sev}}(\mu,\xi,\beta)$ as function with respect to Λ,

$$\begin{aligned}
&\det\big(I_{\text{sev}}(\mu,\xi,\beta)\big)\\
&\quad = \left(\mu\sum_{j=1}^{m} l_j\right)^2 \left(\prod_{i=1}^{2}\sum_{k=1}^{d-1}\frac{a_{i1k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)} - \left(\sum_{k=1}^{d-1}\frac{a_{11k}(\xi,\beta)\,a_{21k}(\xi,\beta)}{b_{1k}(\xi,\beta)}\right)^2\right)
\end{aligned}$$

(a_{ijk}, b_{jk} as defined in Theorem 4.2.2), behaves the same way.

This behavior together with the asymptotic efficiency of the maximum likelihood estimator of (ξ,β) brings about a plausible definition of an *optimal* class length. Based on the principle of D-optimality from theory of experimental design [AD92, Puk06], the optimal class length shall be the maximizer of the determinant of the concerning Fisher information matrix. Since the confidence intervals of ξ and β are derived from the inverse of the Fisher information matrix, maximizing the determinant means minimizing the volume of the (asymptotic) confidence region.

4.4.8 Definition. In the counting model with equidistant class limits the **optimal class length** Λ_{opt} is the class length which maximizes the determinant of $I_{\text{sev}}(\mu,\xi,\beta)$ from Theorem 4.2.2.

As an example, let us calculate the optimal class length in the particular case that $\xi = 0$. For this purpose, the following corollary brings the matrix $I_{\text{sev}}(\mu,\xi,\beta)$ into an easily viewable structure.

4.4.9 Corollary. *Define for $n\in\{1,2,3\}$ the polynomials R_n and the terms Q_n by*

$$\begin{aligned}
R_1(x) &:= 1 - x^{d-1},\\
R_2(x) &:= \tfrac{1}{2} + \tfrac{1}{2}x - \left(d-\tfrac{1}{2}\right)x^{d-1} + \left(d-\tfrac{3}{2}\right)x^{d},\\
R_3(x) &:= \tfrac{1}{4} + \tfrac{3}{2}x + \tfrac{1}{4}x^2 - \left(d-\tfrac{1}{2}\right)^2 x^{d-1} + \left(2d^2-4d+\tfrac{1}{2}\right)x^{d} - \left(d-\tfrac{3}{2}\right)^2 x^{d+1},\\
Q_n(x) &:= x\,(-\log(x))^{1+n}\,(1-x)^{-(1+n)}\,R_n(x),
\end{aligned}$$

for all $x \in (0,1)$. If the class lengths are all equal to Λ, the matrix $I_{\text{sev}}(\mu,\xi,\beta)$ from Theorem 4.2.2 reads for $\xi = 0$ as follows:

$$I_{\text{sev}}(\mu,0,\beta) = \mu \sum_{j=1}^{m} l_j \begin{pmatrix} Q_3\Big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big) & \frac{1}{\beta}\,Q_2\Big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big) \\ \frac{1}{\beta}\,Q_2\Big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big) & \frac{1}{\beta^2}\,Q_1\Big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big) \end{pmatrix}.$$

Thus, the determinant of $I_{\text{sev}}(\mu,0,\beta)$ is

$$\det\big(I_{\text{sev}}(\mu,0,\beta)\big) = \left(\frac{\mu\sum_{j=1}^{m} l_j}{\beta}\right)^2 Q_0\Big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big),$$

where, for all $x \in (0,1)$,

$$\begin{aligned} Q_0(x) &:= x^3\,\log(x)^6\,(1-x)^{-6}\,R_0(x), \\ R_0(x) &:= 1-(d-1)^2x^{d-2}+(2d^2-4d)x^{d-1}-(d-1)^2x^d+x^{2d-2}. \end{aligned}$$

Proof. If $s_{jk} = t_{jk} - u_{\text{sev}} = k\Lambda$ for all $j \in \mathbb{N}_{\le m}$ and $k \in \mathbb{N}_{\le d-1}$, the terms a_{ijk}, b_{jk} in Theorem 4.2.2 satisfy

$$a_{1jk}(0,\beta) = \big(k-\tfrac{1}{2}\big)\,\Lambda^2, \quad a_{2jk}(0,\beta) = \Lambda, \quad b_{jk}(0,\beta) = \beta^4\Big(1-\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big)\,\mathrm{e}^{\frac{1}{\beta}k\Lambda},$$

and therefore

$$\sum_{k=1}^{d-1}\frac{a_{ijk}(0,\beta)\,a_{hjk}(0,\beta)}{b_{jk}(0,\beta)} = \frac{\Lambda^{6-i-h}}{\beta^4\Big(1-\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big)}\sum_{k=1}^{d-1}\big(k-\tfrac{1}{2}\big)^{4-i-h}\Big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\Big)^k$$

for all $i,h \in \{1,2\}$. For all $a \in \{1,2,3\}$, the relation

$$\sum_{k=1}^{d-1}\big(k-\tfrac{1}{2}\big)^{a-1}x^k = \frac{x\,R_a(x)}{(1-x)^a} \qquad \forall x \in \mathbb{R}_{>0}$$

can easily be verified via mathematical induction on d ($d \in \mathbb{N}_{\ge 2}$).

The determinant of $I_{\text{sev}}(\mu,0,\beta)$ is

$$\det\big(I_{\text{sev}}(\mu,0,\beta)\big) = \left(\frac{\mu\sum_{j=1}^{m} l_j}{\beta}\right)^2 \big(Q_1Q_3-{Q_2}^2\big)\big(\mathrm{e}^{-\frac{1}{\beta}\Lambda}\big).$$

The fact that $x\,R_0(x) = R_1(x)R_3(x) - R_2(x)^2$ for all $x \in \mathbb{R}$ finishes the proof. □

All the R_n from Corollary 4.4.9 satisfy $R_n(1) = 0$ ($n \in \{0, \dots, 3\}$). Moreover, l'Hôpital's Rule [For04, p. 171] helps to verify

$$\lim_{x \searrow 0} x^{n_1}(-\log(x))^{n_2} = \left(\lim_{x \searrow 0} \frac{\log(x)}{-x^{-\frac{n_1}{n_2}}}\right)^{n_2} = \left(\lim_{x \searrow 0} \frac{n_2}{n_1}\, x^{\frac{n_1}{n_2}}\right)^{n_2} = 0 \quad \forall n_1, n_2 \in \mathbb{N},$$

$$\lim_{x \nearrow 1} \frac{-\log(x)}{1-x} = \lim_{x \nearrow 1} \frac{1}{x} = 1.$$

Therefore, the Q_n from Corollary 4.4.9 ($n \in \{0, \dots, 3\}$) have the characteristics

$$\lim_{x \searrow 0} Q_n(x) = 0, \quad \lim_{x \nearrow 1} Q_n(x) = 0 \qquad \text{and} \qquad Q_n(x) > 0 \quad \forall x \in (0,1),$$

and so they each must have a maximizer. In fact, each Q_n has got a unique maximizer which shall be denoted by $\tilde{\Lambda}_{d,n}$ (the polynomials R_n depend on the number of classes d). The $\tilde{\Lambda}_{d,n}$ can be found by calculating the derivatives of the Q_n,

$$\tfrac{\mathrm{d}Q_0}{\mathrm{d}x}(x) = Q_0(x)\left(\frac{3}{x} + \frac{6}{x\log(x)} + \frac{6}{1-x} + \frac{\frac{\mathrm{d}R_0}{\mathrm{d}x}(x)}{R_0(x)}\right),$$

$$\tfrac{\mathrm{d}Q_n}{\mathrm{d}x}(x) = Q_n(x)\left(\frac{1}{x} + \frac{1+n}{x\log(x)} + \frac{1+n}{1-x} + \frac{\frac{\mathrm{d}R_n}{\mathrm{d}x}(x)}{R_n(x)}\right) \qquad \forall n \in \{1, 2, 3\},$$

because the maximizers $\tilde{\Lambda}_{d,n}$ are the unique roots of these derivatives. The Q_n are positive on $(0, 1)$, and so the maximizers satisfy the equation(s)

$$\left(1 + \frac{1+n+\mathbb{1}_{\{0\}}(n)}{\log(x)} + \frac{1+n+\mathbb{1}_{\{0\}}(n)}{\frac{1}{x}-1} + \frac{x\,\frac{\mathrm{d}R_n}{\mathrm{d}x}(x)}{\left(1+2\cdot\mathbb{1}_{\{0\}}(n)\right)R_n(x)}\right|_{x=\tilde{\Lambda}_{d,n}} = 0.$$

Since $\tilde{\Lambda}_{d,0}$ is the maximizer of Q_0, the optimal class length in the sense of Definition 4.4.8 must satisfy $\mathrm{e}^{-\frac{1}{\beta}\Lambda_{\text{opt}}} = \tilde{\Lambda}_{d,0}$ (see Corollary 4.4.9) or, equivalently,

$$\Lambda_{\text{opt}} = -\log\left(\tilde{\Lambda}_{d,0}\right)\beta.$$

Particularly, Λ_{opt} is a linear function with respect to the parameter β. Incidentally, this is true even if $\xi > 0$. This follows from the fact that $\det(I_{\text{sev}}(\mu, \xi, \beta))$, or rather $\beta^2 \det(I_{\text{sev}}(\mu, \xi, \beta))$, does not really depend on β and Λ but on the quotient Λ/β. Section 5.5.4 expounds this in detail. In addition, it presents some numerical results for the optimal class length. The results endorse some particular results of Kulldorff [Kul61, pp. 34–36] with regard to optimal grouping.

It is not surprising that the optimal class length depends on the values of the parameters ξ and β. The parameters determine the range where most of the

SOLEs will lie, and the class length is optimal if the classes are just large enough to cover this range evenly. Consequently, the design of the experiment can be optimized only if a rough estimate of ξ and β is available in preparation of the data collection. In addition, it is conceivable that the class length is optimized iteratively over the course of the experiment.

4.5. Estimating the Severity of a SOLE in the Counting-Maximum Model

Similar to the situation at the beginning of Section 4.4, also here F_{sev} denotes the cumulative distribution function of a shifted generalized Pareto distribution. But this time the parameters ξ and β shall be estimated based on the counts *and* on the maximum SOLEs as described in Section 4.1 and Section 4.2.2. In big parts the counting-maximum model can be traced back to a related counting model. In this way, the results from Section 4.4 can be used here. In this context, *related* means the following:

Suppose, (z, x) is a realization of $\big(\boldsymbol{Z}_{\text{v}(j)}, M_{\text{v}(j)}\big)_{1\le j\le m}$,

$$z = (z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in \mathbb{N}_0{}^{m\times d} \qquad \text{and} \qquad x = (x_j)_{1\le j\le m} \in (\{0\}\cup \mathbb{R}_{>u_{\text{sev}}})^m,$$

and $k_1, \dots, k_m$ are the classes that contain the maximum SOLEs,

$$k_j \coloneqq \sum_{k=1}^{d} k\,\mathbb{1}_{A_{jk}}(x_j) \qquad \forall j \in \mathbb{N}_{\le m}.$$

Without loss of generality it is $x \in \mathbb{R}_{>u_{\text{sev}}}{}^m$, because the observations without any SOLE do not influence the likelihood function (see Proposition 4.2.1). A glance at the log-likelihood function $\ell_{\text{sev}}^{\text{CM}}$ in Proposition 4.2.1 as well as the discussion in connection with Theorem 3.4.4 induce to define new class limits

$$\bar{t}_{jk} \coloneqq \begin{cases} t_{jk}, & \text{if } k \in \{0, \dots, k_j - 1\},\\ x_j, & \text{if } k = k_j,\\ \infty, & \text{if } k = k_j + 1, \end{cases} \qquad \text{and} \qquad \bar{s}_{jk} \coloneqq \bar{t}_{jk} - u_{\text{sev}},$$

which leads to the new partitioning

$$B_{j1} \coloneqq (u_{\text{sev}}, \bar{t}_{j1}], \quad B_{j2} \coloneqq (\bar{t}_{j1}, \bar{t}_{j2}], \quad \dots, \quad B_{j,k_j+1} \coloneqq (\bar{t}_{j,k_j}, \infty) \qquad \forall j \in \mathbb{N}_{\le m}.$$

Notice that the number of classes has been changed. For a fixed $j \in \mathbb{N}_{\le m}$ the new number of classes is $k_j + 1$. Particularly, if $x_j \in (t_{j,d-1}, \infty)$, the detection range $\mathcal{S}$ is even divided into $k_j+1 = d+1$ classes. However, define for any given $y = (y_{jk})_{\substack{1\le j\le m\\ 1\le k\le d+1}} \in \mathbb{N}_0{}^{m\times d+1}$ the function $\overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}(\,\cdot\,;y)$ by

$$\begin{aligned}\overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}(\xi,\beta;y) &:= \sum_{j=1}^{m}\sum_{k=1}^{k_j+1} y_{jk}\log\bigl(p_{B_{jk}}\bigr)\\ &= \sum_{j=1}^{m}\sum_{k=1}^{k_j+1} y_{jk}\log\bigl(F_{\mathrm{sev}}\bigl(\overline{t}_{jk}\bigr) - F_{\mathrm{sev}}\bigl(\overline{t}_{j,k-1}\bigr)\bigr)\,,\end{aligned} \tag{4.10}$$

and modify the counts z in two different ways by defining $\overline{z} := (\overline{z}_{jk})_{\substack{1\le j\le m\\ 1\le k\le d+1}}$ and $\tilde{z} := (\tilde{z}_{jk})_{\substack{1\le j\le m\\ 1\le k\le d+1}}$ with

$$\overline{z}_{jk} := \begin{cases} z_{jk}, & \text{if } k\in\mathbb{N}_{\le k_j-1},\\ z_{jk}-1, & \text{if } k=k_j,\\ 1, & \text{if } k=k_j+1,\\ 0, & \text{if } k\in\{k_j+2,\dots,d+1\},\end{cases}$$

$$\tilde{z}_{jk} := \begin{cases} z_{jk}, & \text{if } k\in\mathbb{N}_{\le k_j-1},\\ z_{jk}-1, & \text{if } k=k_j,\\ 0, & \text{if } k\in\{k_j+1,\dots,d+1\}.\end{cases}$$

Based on all these modifications the following fact can be noted.

4.5.1 Lemma. *In the previously described situation it holds for all $(\xi,\beta)\in\Theta_{\mathrm{sev}}$*

$$\begin{aligned}\ell^{\mathrm{CM}}_{\mathrm{sev}}(\xi,\beta;z,x) &= \overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}(\xi,\beta;\tilde{z}) + \sum_{j=1}^{m}\log\left(\tfrac{1-F_{\mathrm{sev}}(x_j)}{\beta+\xi\overline{s}_{jk_j}}\right)\\ &= \overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}(\xi,\beta;\overline{z}) - \sum_{j=1}^{m}\log\bigl(\beta+\xi\overline{s}_{jk_j}\bigr)\,.\end{aligned}$$

Proof. The definitions in the run-up of this lemma and the results of Proposition 4.2.1 ensure

$$\begin{aligned}\ell^{\mathrm{CM}}_{\mathrm{sev}}(\xi,\beta;z,x) &= \sum_{j=1}^{m}\sum_{k=1}^{k_j}\overline{z}_{jk}\log\bigl(p_{B_{jk}}\bigr) + \sum_{j=1}^{m}\log\bigl(\tfrac{\mathrm{d}F_{\mathrm{sev}}}{\mathrm{d}t}(x_j)\bigr)\\ &= \overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}(\xi,\beta;\tilde{z}) + \sum_{j=1}^{m}\log\bigl(\tfrac{\mathrm{d}F_{\mathrm{sev}}}{\mathrm{d}t}(x_j)\bigr)\,.\end{aligned}$$

The derivative of F_{sev} is

$$\tfrac{\mathrm{d}F_{\mathrm{sev}}}{\mathrm{d}t}(t) = \left.\begin{cases}\frac{1}{\beta}\left(1+\frac{\xi}{\beta}(t-u)\right)^{-\frac{1+\xi}{\xi}}, & \text{if } \xi>0\\ \frac{1}{\beta}\mathrm{e}^{-\frac{1}{\beta}(t-u)}, & \text{if } \xi=0\end{cases}\right\} = \frac{1-F_{\mathrm{sev}}(t)}{\beta+\xi(t-u)} \quad \forall t\in\mathbb{R}_{>u_{\mathrm{sev}}}.$$

Due to $\overline{s}_{jk_j} = \overline{t}_{jk_j} - u_{\mathrm{sev}} = x_j - u_{\mathrm{sev}}$, this means

$$\begin{aligned}\log\bigl(\tfrac{\mathrm{d}F_{\mathrm{sev}}}{\mathrm{d}t}(x_j)\bigr) &= \log\bigl(1 - F_{\mathrm{sev}}(\overline{t}_{jk_j})\bigr) - \log\bigl(\beta + \xi \overline{s}_{jk_j}\bigr) \\ &= \log\Bigl(p_{B_{j,k_j+1}}\Bigr) - \log\bigl(\beta + \xi \overline{s}_{jk_j}\bigr)\,,\end{aligned}$$

and the lemma holds. □

The functions $\ell^{\mathrm{C}}_{\mathrm{sev}}$ (defined in Proposition 4.2.1) and $\overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}$ are in principle the same. The only difference is that $\ell^{\mathrm{C}}_{\mathrm{sev}}$ is based on observations with a constant number of classes (d), while $\overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}$ is based on observations with different numbers of classes ($k_1+1, \dots, k_m+1$). However, in principle the partial derivatives of $\overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}$ can be looked up in Lemma 4.4.2 (just change the summation $\sum_{k=1}^{d-1}$ to $\sum_{k=1}^{k_j}$).

In the counting-model the maximum likelihood estimator of (ξ, β) does not exist if the lowest or the highest classes are empty (see Theorem 4.4.4). In the counting-maximum model the maximum likelihood estimator of (ξ, β) exists as long as at least one SOLE has been observed. This is the statement of the following theorem. The associated proof verifies the existence of a global maximum. The verification of the uniqueness of this maximum would be too much to be discussed here. Instead, it should be pointed out that the uniqueness of the global maximum is shown for the counting model (see Proposition 4.4.3 and Theorem 4.4.4) and the similarity of $\ell^{\mathrm{C}}_{\mathrm{sev}}$ and $\ell^{\mathrm{CM}}_{\mathrm{sev}}$ (see Lemma 4.5.1) ensures that also in the counting-maximum model the global maximum is unique.

4.5.2 Theorem. *Let the situation be as defined in the beginning of this section, then the maximum likelihood estimator $(\hat{\xi}_m, \hat{\beta}_m)$ of (ξ, β) based on (z, x) exists. Moreover, it holds*

$$\hat{\beta}_m(z, x) \le \max_{1\le j\le m} x_j - u_{\mathrm{sev}}.$$

Proof. Lemma 4.5.1 provides that

$$\ell^{\mathrm{CM}}_{\mathrm{sev}}(\xi, \beta; z, x) \le \sum_{j=1}^{m} \log\left(\frac{1-F_{\mathrm{sev}}(x_j)}{\beta+\xi\overline{s}_{jk_j}}\right) \le -\sum_{j=1}^{m} \log\bigl(\beta + \xi\overline{s}_{jk_j}\bigr)\,,$$

because $\overline{\ell^{\mathrm{C}}_{\mathrm{sev}}}$ is not positive. On the one hand, l'Hôpital's Rule [For04, p. 171] helps to verify

$$\begin{aligned}&\lim_{\beta\to 0} \frac{1 - F_{\mathrm{sev}}(t)}{\beta + \xi(t - u_{\mathrm{sev}})} \\ &= \left\{\begin{array}{ll} \lim\limits_{\beta\to 0} \beta^{\frac{1}{\xi}}\,\bigl(\beta + \xi(t - u_{\mathrm{sev}})\bigr)^{-\frac{1+\xi}{\xi}}\,, & \text{if } \xi > 0, \\ \lim\limits_{\beta^{-1}\to\infty} \frac{\beta^{-1}}{\mathrm{e}^{(t-u_{\mathrm{sev}})\beta^{-1}}} = \lim\limits_{\beta^{-1}\to\infty} \frac{1}{(t-u_{\mathrm{sev}})\,\mathrm{e}^{(t-u_{\mathrm{sev}})\beta^{-1}}}, & \text{if } \xi = 0, \end{array}\right\} = 0\end{aligned}$$

for all $t \in \mathbb{R}_{>u_{\text{sev}}}$. On the other hand, $\log\bigl(\beta + \xi \overline{s}_{jk_j}\bigr)$ tends to ∞ if β or ξ tend to ∞. Hence, for all $(\xi_0, \beta_0) \in \bigl(\mathbb{R}_{\geq 0} \times \{0, \infty\}\bigr) \cup \bigl(\{\infty\} \times \mathbb{R}_{\geq 0}\bigr) \cup \bigl(\{\infty\} \times \{\infty\}\bigr)$ it follows

$$\ell_{\text{sev}}^{\text{CM}}(\xi, \beta; z, x) \xrightarrow{(\xi,\beta)\to(\xi_0,\beta_0)} -\infty.$$

This means that $\ell_{\text{sev}}^{\text{CM}}(\,\cdot\,; z, x)$ approaches $-\infty$ on the boundary of its domain. Consequently, $\ell_{\text{sev}}^{\text{CM}}(\,\cdot\,; z, x)$ has got a global maximum. As mentioned in the run-up of this theorem, the uniqueness of this maximum can be derived from the uniqueness of the maximum in the counting model.

To verify the upper boundary for $\hat{\beta}_m$, have a look at the partial derivative of $\ell_{\text{sev}}^{\text{CM}}$,

$$\begin{aligned}\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(\xi, \beta; z, x) &= \frac{\partial \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \beta}(\xi, \beta; \tilde{z}) + \sum_{j=1}^{m} \frac{\partial}{\partial \beta} \log\left(\frac{1 - F_{\text{sev}}(x_j)}{\beta + \xi \overline{s}_{jk_j}}\right) \\ &= \frac{\partial \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \beta}(\xi, \beta; \tilde{z}) + \sum_{j=1}^{m} \frac{\overline{s}_{jk_j} - \beta}{\beta^2 + \xi \beta \overline{s}_{jk_j}}.\end{aligned}$$

With $\beta_0 := \max\{\overline{s}_{jk_j} \mid 1 \leq j \leq m\} = \max\{x_j - u_{\text{sev}} \mid 1 \leq j \leq m\}$ it follows

$$\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(\xi, \beta_0; z, x) \leq \frac{\partial \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \beta}(\xi, \beta_0; \tilde{z}) \qquad \forall \xi \in \mathbb{R}_{\geq 0}.$$

Due to the similarity of $\ell_{\text{sev}}^{\text{C}}$ and $\overline{\ell_{\text{sev}}^{\text{C}}}$, Proposition 4.4.3 also holds for $\overline{\ell_{\text{sev}}^{\text{C}}}$ if the varying numbers of classes $(k_j + 1)_{1 \leq j \leq m}$ supersede the constant number d. Since it is $\tilde{z}_{j,k_j+1} = 0$ for all $j \in \mathbb{N}_{\leq m}$ by definition, the second statement of Proposition 4.4.3 provides that $\frac{\partial \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \beta}(\xi, \beta_0; \tilde{z})$ is negative if there is a $j \in \mathbb{N}_{\leq m}$ such that $\tilde{z}_{jk} > 0$ for a $k \in \{2, \ldots, k_j\}$. On the other hand, if only the lowest class is filled, $\sum_{j=1}^{m} \sum_{k=2}^{k_j+1} \tilde{z}_{jk} = 0$, Lemma 4.4.2 yields that $\frac{\partial \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \beta}(\xi, \beta_0; \tilde{z})$ is not positive. All in all, it must be

$$\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(\xi, \beta_0; z, x) \leq 0 \qquad \forall \xi \in \mathbb{R}_{\geq 0}.$$

In the proof of Proposition 4.4.3 it is verified that $\beta \frac{\partial \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \beta}(\xi, \beta; \tilde{z})$ is decreasing (as function with respect to β). Due to

$$\frac{\partial}{\partial \beta}\left(\frac{\overline{s}_{jk_j} - \beta}{\beta + \xi \overline{s}_{jk_j}}\right) = -\frac{\overline{s}_{jk_j}(\xi + 1)}{\bigl(\beta + \xi \overline{s}_{jk_j}\bigr)^2} < 0,$$

the terms $(\overline{s}_{jk_j} - \beta)/(\beta + \xi \overline{s}_{jk_j})$ are strictly decreasing, too. Consequently, it must hold

$$\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(\xi, \beta; z, x) < 0 \qquad \forall \xi \in \mathbb{R}_{\geq 0},\ \forall \beta \in \mathbb{R}_{>\beta_0}.$$

Since $\hat{\beta}_m(z, x)$ is a root of $\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(\hat{\xi}_m(z, x), \cdot\,; z, x)$, it cannot be greater than β_0. □

It should be noted that searching for the maximum likelihood estimators $\hat{\xi}_m(z,x)$ and $\hat{\beta}_m(z,x)$ works in the same way as in the counting model. Corollary 4.4.5 can be formulated for the counting-maximum model as well. Thus, compute the roots $\beta^\circ(0)$ of $\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \xi}(0,\cdot\,;z,x)$ and $\beta^*(0)$ of $\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(0,\cdot\,;z,x)$, first. If $\beta^\circ(0) \le \beta^*(0)$, the maximum likelihood estimator is $(\hat{\xi}_m, \hat{\beta}_m) = (0, \beta^*(0))$. Otherwise, take a $\xi \in \mathbb{R}_{>0}$ large enough such that the root of $\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \beta}(\xi,\cdot\,;z,x)$ exceeds the root of $\frac{\partial \ell_{\text{sev}}^{\text{CM}}}{\partial \xi}(\xi,\cdot\,;z,x)$. Then, it must be $\hat{\xi}_m \in (0,\xi)$, and a simple bisection method can approx the real maximum likelihood estimator $\hat{\xi}_m$ up to the desired precision. Section 5.6 works out this procedure in detail.

In the counting model the estimation of confidence intervals of ξ and β uses the Fisher information from Theorem 4.2.2 due to the asymptotic efficiency of the maximum likelihood estimators (see intervals (4.7) on page 101). Since the calculation of the Fisher information of the counting-maximum model is still pending, the **observed Fisher information** shall be used instead. The entries of this matrix are minus the second partial derivatives of the log-likelihood function ℓ_{CM}. Since the parameters concerning the number of SOLEs, $\nu \in \Theta_{\text{num}}$, and the parameters $\varsigma = (\xi,\beta) \in \Theta_{\text{sev}}$ can be estimated separately, it is sufficient to look at the observed Fisher information matrix concerning ξ and β defined by

$$\mathcal{I}_{\text{sev}}(z,x) := - \left.\begin{pmatrix} \frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \xi^2}(\xi,\beta;z,x) & \frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \xi \partial \beta}(\xi,\beta;z,x) \\ \frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \beta \partial \xi}(\xi,\beta;z,x) & \frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \beta^2}(\xi,\beta;z,x) \end{pmatrix}\right|_{\substack{\xi=\hat{\xi}_m(z,x)\\ \beta=\hat{\beta}_m(z,x)}} .$$

Under some regularity conditions the expectation of the observed Fisher information is equal to the Fisher information matrix [LC98, p. 116]. Efron and Hinkley [EH78] even found that sometimes the observed Fisher information is more suitable for estimating the variance of estimators.

With standard rules from linear algebra the observed Fisher information matrix can be inverted,

$$\mathcal{I}_{\text{sev}}(z,x)^{-1} = \frac{1}{\det(\mathcal{I}_{\text{sev}}(z,x))} \left.\begin{pmatrix} -\frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \beta^2}(\xi,\beta;z,x) & \frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \xi \partial \beta}(\xi,\beta;z,x) \\ \frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \beta \partial \xi}(\xi,\beta;z,x) & -\frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \xi^2}(\xi,\beta;z,x) \end{pmatrix}\right|_{\substack{\xi=\hat{\xi}_m(z,x)\\ \beta=\hat{\beta}_m(z,x)}} .$$

Now, the calculation of actual (approximate) confidence intervals works in the same way as in the counting model (see calculation of the intervals (4.7) on page 101 and the intervals (4.9) on page 103) with the observed Fisher information matrix $\mathcal{I}_{\text{sev}}$ instead of the Fisher information matrix I_{sev}. Having said that, if $\hat{\xi}_m$ is estimated to be $\hat{\xi}_m \in \mathbb{R}_{>0}$, the actual (approximate) confidence intervals of ξ and β with confidence level $1-\alpha$ ($\alpha \in (0,1)$) are

$$\begin{aligned} C_\xi(\alpha,z,x) &:= \left[\hat{\xi}_m(z,x) - \hat{\sigma}_{\xi,m}(z,x)\, q_{1-\alpha/2},\ \hat{\xi}_m(z,x) + \hat{\sigma}_{\xi,m}(z,x)\, q_{1-\alpha/2}\right], \\ C_\beta(\alpha,z,x) &:= \left[\hat{\beta}_m(z,x) - \hat{\sigma}_{\beta,m}(z,x)\, q_{1-\alpha/2},\ \hat{\beta}_m(z,x) + \hat{\sigma}_{\beta,m}(z,x)\, q_{1-\alpha/2}\right], \end{aligned} \tag{4.11}$$

respectively, where, this time,

$$\begin{aligned} \hat{\sigma}_{\xi,m}(z,x) &:= -\frac{\frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \beta^2}\left(\hat{\xi}_m(z,x), \hat{\beta}_m(z,x); z, x\right)}{\det(\mathcal{I}_{\text{sev}}(z,x))}, \\ \hat{\sigma}_{\beta,m}(z,x) &:= -\frac{\frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \xi^2}\left(\hat{\xi}_m(z,x), \hat{\beta}_m(z,x); z, x\right)}{\det(\mathcal{I}_{\text{sev}}(z,x))}, \end{aligned}$$

and, again, $q_{1-\alpha/2}$ is the $\left(1-\frac{\alpha}{2}\right)100\,\%$ quantile of the standard normal distribution. If $\hat{\xi}_m = 0$, the actual (approximate) confidence intervals of ξ and β with confidence level $1-\alpha$ ($\alpha \in (0, 0.5]$) are

$$\begin{aligned} C_\xi^0(\alpha,z,x) &:= \left[0,\ \hat{\sigma}_{\xi,m}(z,x)\, q_{1-\alpha}\right], \\ C_\beta^0(\alpha,z,x) &:= \left[\hat{\Phi}_\beta^{-1}\left(\tfrac{\alpha}{2}\right),\ \hat{\beta}_m(z,x) + \hat{\tau}_{\beta,m}(z,x)\, q_{1-\alpha}\right], \end{aligned} \tag{4.12}$$

respectively, where

$$\hat{\tau}_{\beta,m}(z,x) := -\frac{1}{\frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \beta^2}\left(\hat{\xi}_m(z,x), \hat{\beta}_m(z,x); z, x\right)},$$

and $\hat{\Phi}_\beta^{-1}$ is the inverse of Φ_β as defined in Equation (4.8) on page 103 with the terms $\hat{\sigma}_{\beta,m}(z,x)$ and $\hat{\tau}_{\beta,m}(z,x)$ from this section here (instead of $\sigma_{\beta,m}$ and $\tau_{\beta,m}$ respectively).

The variance terms $\hat{\sigma}_{\xi,m}(z,x)$, $\hat{\sigma}_{\beta,m}(z,x)$ and $\hat{\tau}_{\beta,m}(z,x)$ can be calculated explicitly. In order to use a compact notation, define θ_1 and θ_2 as $\theta_1 := \xi$ and $\theta_2 := \beta$. Then, with the notation of Lemma 4.5.1 one gets

$$\frac{\partial^2 \ell_{\text{sev}}^{\text{CM}}}{\partial \theta_h \partial \theta_i}(\xi,\beta;z,x) = \frac{\partial^2 \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \theta_h \partial \theta_i}(\xi,\beta;\overline{z}) + \sum_{j=1}^{m} \frac{\overline{s}_{jk_j}{}^{4-i-h}}{(\beta + \xi \overline{s}_{jk_j})^2} \qquad \forall i,h \in \{1,2\}.$$

Due to the definition of $\overline{\ell_{\text{sev}}^{\text{C}}}$ in Equation (4.10) on page 111, the second partial derivatives of $\overline{\ell_{\text{sev}}^{\text{C}}}$ are

$$\begin{aligned} \frac{\partial^2 \overline{\ell_{\text{sev}}^{\text{C}}}}{\partial \theta_h \partial \theta_i}(\xi,\beta;\overline{z}) &= \frac{\partial}{\partial \theta_h} \sum_{j=1}^{m} \sum_{k=1}^{k_j+1} \overline{z}_{jk} \frac{\frac{\partial}{\partial \theta_i} p_{B_{jk}}}{p_{B_{jk}}} \\ &= \sum_{j=1}^{m} \sum_{k=1}^{k_j+1} \overline{z}_{jk} \left(\frac{\frac{\partial^2}{\partial \theta_h \partial \theta_i} p_{B_{jk}}}{p_{B_{jk}}} - \frac{\frac{\partial}{\partial \theta_h} p_{B_{jk}}}{p_{B_{jk}}} \frac{\frac{\partial}{\partial \theta_i} p_{B_{jk}}}{p_{B_{jk}}} \right), \end{aligned}$$

and therefore

$$\frac{\partial^2 \ell_{\mathrm{sev}}^{\mathrm{CM}}}{\partial\theta_h\partial\theta_i}(\xi,\beta;z,x) = \sum_{j=1}^{m}\sum_{k=1}^{k_j+1} \overline{z}_{jk}\left(\frac{\frac{\partial^2}{\partial\theta_h\partial\theta_i}p_{B_{jk}}}{p_{B_{jk}}} - \frac{\frac{\partial}{\partial\theta_h}p_{B_{jk}}}{p_{B_{jk}}}\frac{\frac{\partial}{\partial\theta_i}p_{B_{jk}}}{p_{B_{jk}}}\right) + \sum_{j=1}^{m}\frac{\overline{s}_{jk_j}{}^{4-i-h}}{(\beta+\xi\overline{s}_{jk_j})^2}.$$

Finally, Lemma A.1 in the appendix provides the partial derivatives of the probability terms $p_{B_{jk}} = \big(1+\frac{\xi}{\beta}\overline{s}_{j,k-1}\big)^{-\frac{1}{\xi}} - \big(1+\frac{\xi}{\beta}\overline{s}_{jk}\big)^{-\frac{1}{\xi}}$ for $k\in\mathbb{N}_{\le k_j}$,

$$\frac{\frac{\partial}{\partial\theta_i}p_{B_{jk}}}{p_{B_{jk}}} = \frac{\varphi_i\left(\frac{\xi}{\beta},\overline{s}_{j,k-1}\right)\frac{\overline{s}_{j,k-1}}{\beta^2+\xi\beta\overline{s}_{j,k-1}}}{1-\frac{F^*_{\mathrm{sev}}(\overline{t}_{jk})}{F^*_{\mathrm{sev}}(\overline{t}_{j,k-1})}} - \frac{\varphi_i\left(\frac{\xi}{\beta},\overline{s}_{jk}\right)\frac{\overline{s}_{jk}}{\beta^2+\xi\beta\overline{s}_{jk}}}{\frac{F^*_{\mathrm{sev}}(\overline{t}_{j,k-1})}{F^*_{\mathrm{sev}}(\overline{t}_{jk})}-1},$$

$$\frac{\frac{\partial^2}{\partial\theta_h\partial\theta_i}p_{B_{jk}}}{p_{B_{jk}}} = \frac{\phi_{hi}(\xi,\beta,\overline{s}_{j,k-1})\left(\frac{\overline{s}_{j,k-1}}{\beta^2+\xi\beta\overline{s}_{j,k-1}}\right)^2}{1-\frac{F^*_{\mathrm{sev}}(\overline{t}_{jk})}{F^*_{\mathrm{sev}}(\overline{t}_{j,k-1})}} - \frac{\phi_{hi}(\xi,\beta,\overline{s}_{jk})\left(\frac{\overline{s}_{jk}}{\beta^2+\xi\beta\overline{s}_{jk}}\right)^2}{\frac{F^*_{\mathrm{sev}}(\overline{t}_{j,k-1})}{F^*_{\mathrm{sev}}(\overline{t}_{jk})}-1},$$

and for $k = k_j+1$,

$$\frac{\frac{\partial}{\partial\theta_i}p_{B_{j,k_j+1}}}{p_{B_{j,k_j+1}}} = \varphi_i\left(\frac{\xi}{\beta},\overline{s}_{jk_j}\right)\frac{\overline{s}_{jk_j}}{\beta^2+\xi\beta\overline{s}_{jk_j}},$$

$$\frac{\frac{\partial^2}{\partial\theta_h\partial\theta_i}p_{B_{j,k_j+1}}}{p_{B_{j,k_j+1}}} = \phi_{hi}\big(\xi,\beta,\overline{s}_{jk_j}\big)\left(\frac{\overline{s}_{jk_j}}{\beta^2+\xi\beta\overline{s}_{jk_j}}\right)^2,$$

where $F^*_{\mathrm{sev}} := 1-F_{\mathrm{sev}}$ and φ_i, ϕ_{hi} as defined in Lemma A.1 ($i,h\in\{1,2\}$).

5. Implementation and Results

This chapter presents some numerical results concerning the theoretical model from Chapter 3 and Chapter 4. All the defined quantities from the previous chapters are presumed to be known. Section 5.1 describes how to generate random samples drawn from the distributions of number of SOLEs and maximum SOLEs. Section 5.2 explores the hypothesis test for Poisson approach from Section 3.5.3. The approximation of the distribution of the test statistic under the null hypothesis and the power of the test are analyzed. The subsequent three sections introduce algorithms for the calculation of the maximum likelihood estimators. Moreover, the maximum likelihood estimators of exponent ϱ of the number of SOLEs per kilometer (Section 5.3), mean μ of the number of SOLEs per kilometer (Section 5.4), and shape ξ and scale β of the severity of SOLEs in the counting model (Section 5.5) and in the counting-maximum model (Section 5.6) are analyzed by means of Monte Carlo simulations. Finally, a dataset from the BMW study is evaluated based on the results from this thesis.

5.1. Generating Random Samples

If the distributions of the number of events per kilometer and the severity of a single event, F_{num} and F_{sev}, are known, it is possible to generate random samples drawn from the random variables $Z_{l,A}$ and M_{sev}^{*l} (see Definition 3.1.4 and Definition 3.4.1), and from the random vectors $(Z_{l,A_1}, \dots, Z_{l,A_d})$ and $(M_{\mathrm{sev}}^{*l}, Z_{l,A_1}, \dots, Z_{l,A_d})$, where $l \in \mathbb{N}$, $A \in \mathfrak{S}$, $d \in \mathbb{N}_{\geq 2}$, $A_1, \dots, A_d \in \mathfrak{S}$ disjoint. The direct approach for this is in line with the experimental design described in Section 2.2. At first, l random numbers $n_1, \dots, n_l$ drawn from F_{num} must be generated. Then, $n := \sum_{i=1}^{l} n_i$ is the randomly generated total number of SOLEs during l kilometers. Next, n random severities $x_1, \dots, x_n$ are drawn from F_{sev}. Now, just count all x_i that lay within the set A, and a random realization of $Z_{l,A}$ is found. Furthermore, the maximum $x := \max\{x_1, \dots, x_n\}$ is a random realization of M_{sev}^{*l}.

This procedure is quite simple, but it is very costly, too. In order to get one realization of $(M_{\mathrm{sev}}^{*l}, Z_{l,A_1}, \dots, Z_{l,A_d})$ on the average $l\,\mathbb{E}[N_{\mathrm{num}}]$ random numbers must be generated. If l is large, this can be very time-consuming.

However, there are much faster algorithms for generating random samples. The methods described in the following subsections make use of the immediate distributions of $Z_{l,A}$ and M_{sev}^{*l}.

5.1.1. Samples from $Z_{l,A}$

According to Section 3.5, let F_{num} be the cumulative distribution function of either the binomial, the Poisson or the negative binomial distribution. If N_{num} is binomially, Poisson or negative binomially distributed, $Z_{l,A}$ is in the same distribution family (see Example 3.2.2). Many common mathematical tools include functions for generating (pseudo-)random samples drawn from these three distributions (e. g. MATLAB, R, SPSS STATISTICS). Some of the basic methods which are used by these tools can be found in [Dev86, AD82a, KS88]. For example, when generating negative binomial random samples, it can be taken advantage of the fact that the negative binomial distribution is a gamma-Poisson mixture distribution (see Section 3.5.4) — provided that one is able to generate Poisson and gamma samples (with help of Ahrens and Dieter [AD82a, AD82b], for example).

5.1.2. Samples from $(Z_{l,A_1}, \ldots, Z_{l,A_d})$

Theorem 3.3.2 provides that the variates $Z_{l,A_1}, \ldots, Z_{l,A_d}$ are statistically independent if N_{num} is Poisson distributed. In this case, just realize the Z_{l,A_k} separately by dint of the methods cited above in Section 5.1.1. If, for all $k \in \mathbb{N}_{\leq d}$, z_k is the drawn realization of Z_{l,A_k}, then $(z_1, \ldots, z_d)$ is a realization of $(Z_{l,A_1}, \ldots, Z_{l,A_d})$.

If N_{num} is negative binomially distributed, $N_{\text{num}} \sim \text{NBin}(\varrho, \mu)$, first, take a gamma distributed random variable W, $W \sim \Gamma\big(\varrho, \frac{\mu}{\varrho}\big)$, and generate a realization ω from it (with help of Ahrens and Dieter [AD82b], for example). The gamma-Poisson mixture property of the negative binomial distribution (see Section 3.5.4) ensures that the variates Z_{l,A_k} given $W = \omega$ are Poisson distributed,

$$Z_{l,A_k}|W{=}\omega \;\sim\; \text{Poi}(\mu l p_{A_k}) \qquad \forall k \in \mathbb{N}_{\leq d}.$$

From here on, the situation is exactly the same as in case of a Poisson distributed N_{num}, which is described above.

Finally, let N_{num} be binomially distributed, $N_{\text{num}} \sim \text{Bin}(r, q)$. Divide the unit interval $[0, 1)$ up into $d + 1$ subintervals,

$$[0, 1) = \bigcup_{k=1}^{d+1} [\delta_{k-1}, \delta_k)$$

with $\delta_0 := 0$, $\delta_{d+1} := 1$ and $\delta_k := q\sum_{i=1}^{k} p_{A_i}$ for all $k \in \mathbb{N}_{\leq d}$. Generate rl realizations $y_1, \ldots, y_{rl}$ of a uniform distribution on $[0,1)$. If z_k means the number of realizations which lay within the interval $[\delta_{k-1}, \delta_k)$,

$$z_k := \sum_{n=1}^{rl} \mathbb{1}_{[\delta_{k-1},\delta_k)}(y_n) \qquad \forall k \in \mathbb{N}_{\leq d},$$

then $(z_1, \ldots, z_d)$ is a realization of $(Z_{l,A_1}, \ldots, Z_{l,A_d})$.

5.1.3. Samples from M_{sev}^{*l}

The **inverse transform sampling** [Fis96, p. 149] is a basic technique for generating (pseudo-) random numbers drawn from a distribution with known cumulative distribution function F. The method is based upon the property that $F^{-1}(U)$ is distributed according to F if U is a uniformly distributed random variable on $(0,1)$ with cumulative distribution function $x\,\mathbb{1}_{(0,1)}(x)$ and F^{-1} is the quantile function of F,

$$F^{-1}\colon (0,1) \to \mathbb{R}\colon\ x \longmapsto \inf\left\{t \in \mathbb{R} \,|\, F(t) \geq x\right\}.$$

The cumulative probability function of the maximum SOLE during l kilometers ($l \in \mathbb{N}$) is needed to specify the quantile function of M_{sev}^{*l},

$$F^{-1}_{M_{\text{sev}}^{*l}}\colon (0,1) \to \mathbb{R}\colon\ x \longmapsto \inf\left\{t \in \mathbb{R} \,\middle|\, \mathbb{P}\Big(M_{\text{sev}}^{*l} \leq t\Big) \geq x\right\}.$$

The selfsame function is given in Proposition 3.4.2,

$$\mathbb{P}\Big(M_{\text{sev}}^{*l} \leq t\Big) = G_{\text{num}}(F_{\text{sev}}(t))^l\ \mathbb{1}_{\mathbb{R}_{\geq 0}}(t) \qquad \forall t \in \mathbb{R}.$$

Since G_{num} denotes a probability-generating function, it is strictly increasing and continuous on $[0,1]$. Thus, it is possible to define the inverse function of G_{num},

$$G_{\text{num}}^{-1}\colon [G_{\text{num}}(0)\,, 1] \to [0,1]\colon\ G_{\text{num}}(t) \longmapsto t.$$

Furthermore, F_{sev}^{-1} shall be the quantile function of F_{sev},

$$F_{\text{sev}}^{-1}\colon (0,1) \to \mathcal{S}\colon\ x \longmapsto \inf\{t \in \mathbb{R} \,|\, F_{\text{sev}}(t) \geq x\}.$$

This gives the quantile function of M_{sev}^{*l},

$$F^{-1}_{M_{\text{sev}}^{*l}}(x) = \begin{cases} 0, & \text{if } x \in \big(0, G_{\text{num}}(0)^l\,\big], \\ F_{\text{sev}}^{-1}\big(G_{\text{num}}^{-1}(\sqrt[l]{x})\big)\,, & \text{if } x \in \big(G_{\text{num}}(0)^l\,, 1\big). \end{cases}$$

According to Section 3.6, F_{sev} is the cumulative distribution function of a shifted generalized Pareto distribution with threshold u_{sev}. Thus, the quantile function F_{sev}^{-1} simply is the inverse function of F_{sev},

$$F_{\mathrm{sev}}^{-1}(x) = \begin{cases} u_{\mathrm{sev}} + \frac{\beta}{\xi}\left((1-x)^{-\xi} - 1\right), & \text{if } \xi \neq 0, \\ u_{\mathrm{sev}} - \beta \log(1-x), & \text{if } \xi = 0, \end{cases} \qquad \forall x \in (0,1).$$

The probability-generating functions of the binomial, Poisson and negative binomial distributions are listed in Definition 2.4.2. The inverse functions are:

- $N_{\mathrm{num}} \sim \mathrm{Bin}(r,q) \quad \Rightarrow \quad G_{\mathrm{num}}^{-1}(x) = 1 + \frac{\sqrt[r]{x}-1}{q} \quad \forall x \in \left[(1-q)^r, 1\right],$
- $N_{\mathrm{num}} \sim \mathrm{Poi}(\lambda) \quad \Rightarrow \quad G_{\mathrm{num}}^{-1}(x) = 1 + \frac{\log(x)}{\lambda} \quad \forall x \in \left[\mathrm{e}^{-\lambda}, 1\right],$
- $N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu) \quad \Rightarrow \quad G_{\mathrm{num}}^{-1}(x) = 1 + \frac{\varrho}{\mu}\left(1 - \sqrt[\varrho]{\frac{1}{x}}\right) \ \forall x \in \left[\left(\frac{\varrho}{\varrho+\mu}\right)^{\varrho}, 1\right].$

Together, in case of $N_{\mathrm{num}} \sim \mathrm{Bin}(r,q)$, the quantile function of M_{sev}^{*l} is

$$F_{M_{\mathrm{sev}}^{*l}}^{-1}(x) = \mathbb{1}_{((1-q)^{rl},1)}(x) \cdot \begin{cases} u_{\mathrm{sev}} + \frac{\beta}{\xi}\left(\left(\frac{q}{1-\sqrt[rl]{x}}\right)^{\xi} - 1\right), & \text{if } \xi \neq 0, \\ u_{\mathrm{sev}} + \beta \log\left(\frac{q}{1-\sqrt[rl]{x}}\right), & \text{if } \xi = 0, \end{cases} \qquad \forall x \in (0,1),$$

in case of $N_{\mathrm{num}} \sim \mathrm{Poi}(\lambda)$, it is

$$F_{M_{\mathrm{sev}}^{*l}}^{-1}(x) = \mathbb{1}_{(\mathrm{e}^{-\lambda l},1)}(x) \cdot \begin{cases} u_{\mathrm{sev}} + \frac{\beta}{\xi}\left(\left(-\frac{\lambda l}{\log(x)}\right)^{\xi} - 1\right), & \text{if } \xi \neq 0, \\ u_{\mathrm{sev}} + \beta \log\left(-\frac{\lambda l}{\log(x)}\right), & \text{if } \xi = 0, \end{cases} \qquad \forall x \in (0,1),$$

and finally, if $N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu)$, the quantile function is

$$F_{M_{\mathrm{sev}}^{*l}}^{-1}(x) = \mathbb{1}_{((\varrho/\varrho+\mu)^{\varrho l},1)}(x) \cdot \begin{cases} u_{\mathrm{sev}} + \frac{\beta}{\xi}\left(\left(\frac{\varrho}{\mu}\left(\sqrt[\varrho l]{\frac{1}{x}} - 1\right)\right)^{-\xi} - 1\right), & \text{if } \xi \neq 0, \\ u_{\mathrm{sev}} + \beta \log\left(\frac{\mu}{\varrho}\right) - \beta \log\left(\sqrt[\varrho l]{\frac{1}{x}} - 1\right), & \text{if } \xi = 0 \end{cases}$$

for all $x \in (0,1)$.

Due to the inverse transform sampling method, take realizations $y_1, \ldots, y_n$ of a uniform distribution on $(0,1)$ ($n \in \mathbb{N}$), and $F_{M_{\mathrm{sev}}^{*l}}^{-1}(y_1), \ldots, F_{M_{\mathrm{sev}}^{*l}}^{-1}(y_n)$ is a random sample of size n drawn from M_{sev}^{*l}.

5.1.4. Samples from $(M^{*l}_{\text{sev}}, Z_{l,A_1}, \ldots, Z_{l,A_d})$

Without loss of generality, suppose it is $\mathbb{P}\Big(S_{\text{sev}} \in \bigcup_{k=1}^{d} A_k\Big) = 1$. Otherwise, define $A_{d+1} := \mathcal{S} \setminus \bigcup_{k=1}^{d} A_k$, generate samples from $\big(M^{*l}_{\text{sev}}, Z_{l,A_1}, \ldots, Z_{l,A_d+1}\big)$, and forget the realizations of Z_{l,A_d+1}.

Also without loss of generality, suppose that $A_1, \ldots, A_d$ are intervals,

$$A_1 = (t_0, t_1], \quad \ldots, \quad A_d = (t_{d-1}, t_d),$$

with interval limits $0 < u_{\text{sev}} = t_0 < t_1 < \ldots < t_{d-1} < t_d = \infty$. Otherwise, use the fact that every Borel set is almost surely equal to a union of intervals,

$$A_k = (t_{k0}, t_{k1}] \cup \ldots \cup (t_{k,c_k-1}, t_{kc_k}] \qquad \mathbb{P}\text{-a.s.}$$

with $u_{\text{sev}} \leq t_{k0} < t_{k1} < \ldots t_{k,c_k-1} < t_{kc_k} \leq \infty$ ($c_k \in \mathbb{N}$ for all $k \in \mathbb{N}_{\leq d}$). Define

$$B_{ki_k} := (t_{k,i-1}, t_{ki}] \qquad \forall i_k \in \mathbb{N}_{\leq c_k},\ \forall k \in \mathbb{N}_{\leq d},$$

generate samples from $\big(M^{*l}_{\text{sev}}, Z_{l,B_{11}}, \ldots, Z_{l,B_{1c_1}}, \ldots, Z_{l,B_{d1}}, \ldots, Z_{l,B_{dc_d}}\big)$, and because of

$$Z_{l,A_k} = \sum_{i=1}^{c_k} Z_{l,B_{ki}} \qquad \mathbb{P}\text{-a.s.},$$

the sought-after realization is found.

With that, it manages to generate a sample from $(M^{*l}_{\text{sev}}, Z_{l,A_1}, \ldots, Z_{l,A_d})$ in two steps. First, generate a realization $(z_1, \ldots, z_d)$ of $(Z_{l,A_1}, \ldots, Z_{l,A_d})$ as described in Section 5.1.2. In order to prepare for the second step, define

$$E_Z := \big\{Z_{l,A_1} = z_1, \ldots, Z_{l,A_d} = z_d\big\} \qquad \text{and} \qquad k_0 := \max\big\{k \in \mathbb{N}_{\leq d} \,\big|\, z_k > 0\big\}.$$

Secondly, if $(a, b]$ denotes the largest nonempty class and z is the generated number of SOLEs within this class, $(a, b] := (t_{k_0-1}, t_{k_0}]$ and $z := z_{k_0}$, the proof of Theorem 3.4.4 accompanies the equation

$$\mathbb{P}\Big(M^{*l}_{\text{sev}} \leq t \,\Big|\, E_Z\Big) = \left(\frac{p_{(a,t]}}{p_{(a,b]}}\right)^z = \left(\frac{F_{\text{sev}}(t) - F_{\text{sev}}(a)}{F_{\text{sev}}(b) - F_{\text{sev}}(a)}\right)^z.$$

Hence, the quantile function of M^{*l}_{sev} given E_Z,

$$F^{-1}_{M^{*l}_{\text{sev}}|E_Z} : (0,1) \to (a, b] \colon \ x \longmapsto \inf\Big\{t \in \mathbb{R} \,\Big|\, \mathbb{P}\Big(M^{*l}_{\text{sev}} \leq t \,\big|E_Z\Big) \geq x\Big\},$$

is in general specified through

$$F^{-1}_{M^{*l}_{\text{sev}}|E_Z}(x) = F^{-1}_{\text{sev}}\Big(F_{\text{sev}}(a) + \sqrt[z]{x}\big(F_{\text{sev}}(b) - F_{\text{sev}}(a)\big)\Big).$$

With the generalized Pareto approach for F_{sev} one gets

$$F^{-1}_{M^{*l}_{\text{sev}}|E_Z}(x) = \begin{cases} u_{\text{sev}} - \frac{\beta}{\xi} + \dfrac{\frac{\beta}{\xi}+a-u_{\text{sev}}}{\left(1-\sqrt[z]{x}+\sqrt[z]{x}\left(\frac{\beta+\xi(a-u_{\text{sev}})}{\beta+\xi(b-u_{\text{sev}})}\right)^{\frac{1}{\xi}}\right)^{\xi}}, & \text{if } \xi \neq 0, \\ a - \beta \log\left(1 - \sqrt[z]{x} + \sqrt[z]{x}\, \mathrm{e}^{-\frac{1}{\beta}(b-a)}\right), & \text{if } \xi = 0, \end{cases} \quad \forall x \in (0,1).$$

At last, take a realization y of a uniform distribution on $(0,1)$, and, according to the inverse transform sampling method described in Section 5.1.3, the vector $\left(F^{-1}_{M^{*l}_{\text{sev}}|E_Z}(y),\ z_1,\ \ldots,\ z_d\right)$ is a realization drawn from the random vector $(M^{*l}_{\text{sev}}, Z_{l,A_1}, \ldots, Z_{l,A_d})$.

5.2. Accuracy and Power of Index-of-Dispersion Hypothesis Test

In Section 3.5.3 a hypothesis test is developed to check whether the index of dispersion of the number of SOLEs per kilometer, $\mathbb{D}[N_{\text{num}}]$, differs from 1 significantly. In this connection, the mileage covered by a vehicle is treated as a random variable called L. If the data consists of m observations ($m \in \mathbb{N}$), the underlying mileages $l_1, \ldots, l_m$ are realizations of the statistically independent variates $L_1, \ldots, L_m$ which are all distributed according to L. The observed total number of SOLEs $n_1, \ldots, n_m$ are realizations of the statistically independent variates $\mathcal{N}_1, \ldots, \mathcal{N}_m$ defined by

$$\mathcal{N}_j := \sum_{i=1}^{L_j} N_{ij} \qquad \forall j \in \mathbb{N}_{\leq m},$$

where the N_{ij} are statistically independent random variables distributed according to the number of SOLEs per kilometer N_{num}. Theorem 3.5.2 and Corollary 3.5.3 provide that the term $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is approximately standard normally distributed if N_{num} is Poisson distributed, where $\hat{D}_2$ denotes the estimator of $\mathbb{D}[N_{\text{num}}]$ defined in Equation (3.6) on page 40,

$$\hat{D}_2 = \hat{D}_2\big((\mathcal{N}_j, L_j)_{1 \leq j \leq m}\big) = \frac{\sum_{j=1}^{m} \frac{\mathcal{N}_j^{\,2}}{L_j} - \frac{\left(\sum_{j=1}^{m} \mathcal{N}_j\right)^2}{\sum_{j=1}^{m} L_j}}{\sum_{j=1}^{m} \frac{\mathcal{N}_j}{L_j}}.$$

Based on this, the hypothesis test in Section 3.5.3 suggests to reject the hypothesis that $\mathbb{D}[N_{\text{num}}]$ is equal to 1 if and only if $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is outside the interval $\big[-q_{1-\alpha/2},\, q_{1-\alpha/2}\big]$, where $q_{1-\alpha/2}$ denotes the $\big(1 - \frac{\alpha}{2}\big)\, 100\,\%$ quantile of the standard normal distribution ($\alpha \in (0,1)$).

Using MATLAB [MAT12], computer simulations were run to consider the power of the mentioned hypothesis test and the actual distribution of $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ if N_{num} is Poisson distributed.

5.2.1. Accuracy

The significance test is based on the assumption that $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is approximately standard normally distributed. It is necessary to get an appraisal of the quality of this approximation. For this purpose, the simulation generates samples of sizes $m = 10$, 20, 50, 100, 500 and 1000. N_{num} is set to be Poisson distributed, $N_{\text{num}} \sim \text{Poi}(\mu)$, with means $\mu = 10^{-4}$, 10^{-3} and 10^{-2}. This corresponds to average waiting times of 10 000, 1000 and 100 kilometers for a SOLE. The simulation of the mileage is implemented in two different ways: firstly, L is (discrete) uniformly distributed on $\{1000, 1001, \ldots, 50\,999, 51\,000\}$, secondly, L is distributed according to $1000 + \lfloor \tilde{L} + 1/2 \rfloor$, where $\tilde{L}$ is exponentially distributed with mean 25 000 (corresponds to a generalized Pareto distribution with shape $\xi = 0$ and scale $\beta = 25\,000$). The term $\lfloor x \rfloor$ means the largest integer not greater than x. In both cases, L is not smaller than 1000 and the average mileage is 26 000 kilometers.

For each combination of distribution of L and values of m and μ, 10^6 random samples from $(N_1, L_1), \ldots, (N_m, L_m)$ were generated, and for each sample the quantity $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ was calculated. The k-statistics $\hat{k}_1, \ldots, \hat{k}_4$, the unbiased and consistent estimators of the first four cumulants (see Equation (3.7) on page 43), were determined from the resultant values $\delta_1, \ldots, \delta_{10^6}$. Since the first cumulant equates to the expectation, the second cumulant equates to the variance, and all other cumulants vanish in case of a normal distribution [JKB94, p. 89], $\hat{k}_1(\delta_1, \ldots, \delta_{10^6})$, $\hat{k}_3(\delta_1, \ldots, \delta_{10^6})$ and $\hat{k}_4(\delta_1, \ldots, \delta_{10^6})$ are expected to be approximately equal to 0, while the statistic $\hat{k}_2(\delta_1, \ldots, \delta_{10^6})$ should be approximately equal to 1. The actual values can be found in Table B.1 in the appendix. In this table, the column "km" refers to the mileage L which is either uniformly distributed (U) or, more or less, exponentially distributed (Exp).

Table B.1 shows that for all settings the mean of $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is slightly negative and approaches 0 if the sample size m increases. The variance is a little too small, too. However, basically, the approximation of $\hat{k}_2(\delta_1, \ldots, \delta_{10^6})$ through 1 is very accurate. For sample sizes $m \in \{500, 1000\}$, the disparity is only visible in the fourth digit after the decimal point. The third- and fourth-order cumulants are positive, but they approach 0 if the sample size m increases. Consequently, due to Cramér [Cra62, pp. 183/187], for a small amount of samples the frequency curve of $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is right-tailed (or right-skewed) and more tall and slim than the normal curve in the neighborhood of the mode. The latter characteristic is also called **leptokurtic** [UC11]. The approximation of the third- and fourth-order cumulants through 0 becomes better if the average number of SOLEs μ increases, whereas the first- and second-order cumulants are essentially not influenced by μ.

Some of these characteristics become visible in Table B.2, too. This table collects the deviation of some population quantiles of $\delta_1, \dots, \delta_{10^6}$ from the particular quantiles of the standard normal distribution. If $q_{1-\alpha}$ denotes the $(1-\alpha)100\,\%$ quantile of the standard normal distribution, then MATLAB [MAT12] determines for $\alpha = 0.99$, 0.95, 0.9, 0.75, 0.5, 0.25, 0.1, 0.05 and 0.01 the rounded values

$$\begin{aligned} q_{0.01} &\approx -2.326, & q_{0.25} &\approx -0.674, & q_{0.9} &\approx 1.282, \\ q_{0.05} &\approx -1.645, & q_{0.5} &= 0, & q_{0.95} &\approx 1.645 \\ q_{0.1} &\approx -1.282, & q_{0.75} &\approx 0.674, & q_{0.99} &\approx 2.326. \end{aligned}$$

The method for determining the population quantiles of $\delta_1, \dots, \delta_{10^6}$ is based on the `quantile` command of MATLAB [MAT12] which is denoted as $\hat{Q}_5(p)$ by Hyndman and Fan [HF96]. According to this, the ith smallest value $\delta_{(i)}$ is taken as the $(i - \frac{1}{2})10^{-6}$ quantile, and linear interpolation is used to compute quantiles of probabilities between these values. Because the quantity $\tilde{\alpha} := (1-\alpha)10^6$ is an integer for all values of α from above, the $(1-\alpha)100\,\%$ quantile of $\delta_1, \dots, \delta_{10^6}$ simply is $\frac{1}{2}(\delta_{(\tilde{\alpha})} + \delta_{(\tilde{\alpha}+1)})$. Thus, the quantity

$$q_\Delta(1-\alpha) := \frac{\delta_{(\tilde{\alpha})} + \delta_{(\tilde{\alpha}+1)}}{2} - q_{1-\alpha}$$

specifies the desired deviation of the $(1-\alpha)100\,\%$ quantile of $\delta_1, \dots, \delta_{10^6}$ from the $(1-\alpha)100\,\%$ quantile of the standard normal distribution.

Table B.2 in the appendix shows that the actual distribution of the term $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is slightly right-skewed and leptokurtic if the sample size m is small. The deviation regarding extreme quantiles becomes smaller if μ increases. However, the quartiles are not influenced by the average number of SOLEs.

The level of skewness is illustrated in Figure C.1 and Figure C.2 in the appendix. Figure C.1 shows the standard normal distribution (*red*) next to the frequency distribution of the simulated values $\delta_1, \dots, \delta_{10^6}$ based on different sample sizes $m = 10$, 50, 1000, constant mean $\mu = 10^{-3}$, and uniformly distributed mileages. On the other hand, Figure C.2 shows the standard normal distribution (*red*) next to the frequency distribution of the simulated values $\delta_1, \dots, \delta_{10^6}$ based on different means $\mu = 10^{-4}$, 10^{-3} and 10^{-2}, constant sample size $m = 20$, and uniformly distributed mileages.

5.2.2. Power

The power of a hypothesis test is the probability of accepting the alternative when it is in fact true [UC11]. In this special test, the null hypothesis H_0 says that the index of dispersion $\mathbb{D}[N_{\text{num}}]$ is equal to 1. The alternative hypothesis '$\mathbb{D}[N_{\text{num}}] \neq 1$' can be partitioned into H_{1-} and H_{1+},

$$H_{1-}\colon \mathbb{D}[N_{\text{num}}] \in (0,1), \qquad H_0\colon \mathbb{D}[N_{\text{num}}] = 1, \qquad H_{1+}\colon \mathbb{D}[N_{\text{num}}] \in \mathbb{R}_{>1}.$$

To evaluate the power of the test, a computer simulation generates samples of sizes $m = 10$, 20, 50, 100, 500 and 1000. The index of dispersion is set to be $\mathbb{D}[N_{\text{num}}] = 0.25$, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 3 and 5. The simulations of N_{num} and L work as follows:

- $\mathbb{D}[N_{\text{num}}] > 1$: $N_{\text{num}} \sim \text{NBin}\big(\frac{\mu}{\mathbb{D}[N_{\text{num}}]-1}, \mu\big)$ (negative binomial distribution) with means $\mu = 10^{-4}$, 10^{-3}, 10^{-2}. L is as simulated as in Section 5.2.1: uniformly distributed on $\{1000, 1001, \ldots, 50\,999, 51\,000\}$ and distributed according to $1000 + \lfloor \tilde{L} + 1/2 \rfloor$, where $\tilde{L}$ is exponentially distributed with mean 25 000.
- $\mathbb{D}[N_{\text{num}}] = 1$: $N_{\text{num}} \sim \text{Poi}(\mu)$ (Poisson distribution) with means $\mu = 10^{-4}$, 10^{-3} and 10^{-2}. L is as simulated as above.
- $\mathbb{D}[N_{\text{num}}] < 1$: $N_{\text{num}} \sim \text{Bin}\Big(\frac{r}{1-\mathbb{D}[N_{\text{num}}]}, 1 - \mathbb{D}[N_{\text{num}}]\Big)$ (binomial distribution) with trials $r = 1$, 10, 100. L is either uniformly distributed on $\{1,2,3,4,5\}$ or distributed according to $1 + \overline{L}$ with $\overline{L} \sim \text{NBin}(1,2)$. In both cases, L is positive with expectation value 3.

The different simulation method of the mileage L in case of underdispersion ensures that in all cases similar numbers of events are generated:

$$\begin{aligned} \mathbb{D}[N_{\text{num}}] \geq 1: \qquad & \mathbb{E}\Big[N_{\text{num}}^{*L}\Big] = \mathbb{E}[L]\,\mathbb{E}[N_{\text{num}}] = 26\,000\mu \in \{2.6,\ 26,\ 260\}, \\ \mathbb{D}[N_{\text{num}}] < 1: \qquad & \mathbb{E}\Big[N_{\text{num}}^{*L}\Big] = \mathbb{E}[L]\,\mathbb{E}[N_{\text{num}}] = 3r \in \{3,\ 30,\ 300\}. \end{aligned}$$

Note that the NBin(1, 2) distribution is identical to a geometric distribution, which is the discrete analogue of the exponential distribution.

For each combination of distribution of L and values of m, $\mathbb{D}[N_{\text{num}}]$, μ and r, 10^5 random samples from $(N_1, L_1), \ldots, (N_m, L_m)$ were generated, and for each sample the quantity $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ was calculated. It was counted how often the null hypothesis H_0 is rejected in favour of H_{1+} due to $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big) > q_{1-\alpha/2}$, how often the null hypothesis H_0 is rejected in favour of H_{1-} due to the relation $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big) < -q_{1-\alpha/2}$, and how often the null hypothesis cannot be rejected. The significance level is set to be $\alpha = 0.05$, so that $q_{1-\alpha/2} \approx 1.9600$. Table B.3, Table B.4 and Table B.5 in the appendix collect the results of these counts. The column "km" refers to the mileage L which is either uniformly distributed (U), negative binomially distributed (NBin) or, more or less, exponentially distributed (Exp). The abbreviation "IOD" stands for index of dispersion.

Since the significance level is set to be $\alpha = 0.05$, the acceptance rate of the null hypothesis is expected to be about 95 % if $\mathbb{D}[N_{\text{num}}] = 1$. Table B.3 indicates that the test holds the nominal level quite well when sample size is not to small, i. e. $m \geq 100$. The column "IOD=1" also confirms the results of Section 5.2.1: the distribution of $\sqrt{\frac{m}{2}}(\hat{D}_2 - 1)$ is right-skewed and leptokurtic under the Poisson hypothesis.

Expectedly, the power increases if the index of dispersion increases. For large sample sizes, i. e. $m \geq 500$, the power increases very fast and already approaches 100 % for $\mathbb{D}[N_{\text{num}}] = 1.5$. For medium sample sizes, i. e. $m \in \{50, 100\}$, the power approaches 100 % for $\mathbb{D}[N_{\text{num}}] = 3$. If the sample size is smaller than 50, the power increases slowly. For $m \in \{10, 20\}$ it even happens, although extremely rarely, that underdispersion is suggested even though $\mathbb{D}[N_{\text{num}}] = 2.5$. The power is mostly better for uniformly distributed mileages than for exponentially distributed ones. This is because ${\tau_{\text{iod}}}^2$, the asymptotic variance of $\sqrt{m}\hat{D}_2$, increases linearly with $\mathbb{E}\left[\frac{1}{L}\right]$ (see Theorem 3.5.2), and $\mathbb{E}\left[\frac{1}{L}\right]$ is smaller for the uniformly distributed L. Here, a smaller variance means less mass in the nonrejection range, and, accordingly, more power. Finally, the power increases if the mean μ increases. This also has to do with the approximate variance ${\tau_{\text{iod}}}^2$. Theorem 3.5.2 and Equation (4.4) on page 84 indicate that ${\tau_{\text{iod}}}^2$ increases linearly with μ^{-1} if N_{num} is negative binomially distributed and the index of dispersion is kept constant. The course of power described is also illustrated in Figure C.3.

Table B.5 collects the results concerning indices of dispersion less than 1, i e. $\mathbb{D}[N_{\text{num}}] = 0.25$, 0.5, 0.75 and 0.9. For the first three settings, the sample size $m = 500$ is large enough to ensure a power of minimum 99.5 %. However, for small sample sizes the power increases only very slowly if $\mathbb{D}[N_{\text{num}}]$ decreases. To understand the influence of mileage distribution and average number of SOLEs on the power, let us have a look at ${\tau_{\text{iod}}}^2$, the asymptotic variance of $\sqrt{m}\hat{D}_2$ from Theorem 3.5.2. If N_{num} is binomially distributed, $N_{\text{num}} \sim \text{Bin}(r, q)$, the cumulants are given by the recurrence relation

$$\kappa_{n+1}[N_{\text{num}}] = q(1-q)\, \tfrac{\partial}{\partial q} \kappa_n[N_{\text{num}}] \qquad \forall n \in \mathbb{N}.$$

[JKK05, p. 111]. Since the first cumulant is equal to the mean $\mathbb{E}[N_{\text{num}}] = rq$ (see Section 2.4.6 and Definition 2.4.2), it follows

$$\kappa_1[N_{\text{num}}] = rq, \qquad \kappa_2[N_{\text{num}}] = rq\,(1-q), \qquad \kappa_3[N_{\text{num}}] = rq\,(1-q)\,(1-2q),$$
$$\kappa_4[N_{\text{num}}] = rq\,(1-q)\,(1-6q+6q^2),$$

and so

$$\begin{aligned} {\tau_{\text{iod}}}^2 &= \mathbb{E}\left[\tfrac{1}{L}\right] \frac{-2+5q-3q^2}{r} + 2(1-q)^2 \\ &= \mathbb{E}\left[\tfrac{1}{L}\right] \frac{\mathbb{D}[N_{\text{num}}] - 3\,\mathbb{D}[N_{\text{num}}]^2}{r} + 2\,\mathbb{D}[N_{\text{num}}]^2\,. \end{aligned}$$

Notice that the term $\left(\mathbb{D}[N_{\text{num}}] - 3\,\mathbb{D}[N_{\text{num}}]^2\right)$ is positive if and only if it holds $\mathbb{D}[N_{\text{num}}] \in \left(0, \frac{1}{3}\right)$. In this case, ${\tau_{\text{iod}}}^2$ decreases if either $\mathbb{E}\left[\frac{1}{L}\right]$ or r^{-1} increases. The reverse applies for $\mathbb{D}[N_{\text{num}}] \in \left(\frac{1}{3}, 1\right)$. Depending on whether a decrease of ${\tau_{\text{iod}}}^2$ means more or less mass in the nonrejection range, power decreases or increases. This variable behavior is also illustrated in Figure C.3. It can also be seen there that the sample size is the most influential quantity with regard to the power.

5.3. Calculation and Accuracy of Maximum Likelihood estimator of ϱ

5.3.1. Calculation

Let be $m \in \mathbb{N}$, $(l_j)_{1\leq j\leq m} \in \mathbb{N}^m$ and $(n_j)_{1\leq j\leq m} \in {\mathbb{N}_0}^m$ satisfying

$$\sum_{j=1}^{m} \frac{{n_j}^2}{l_j} - \frac{n^2}{l} > \sum_{j=1}^{m} \frac{n_j}{l_j},$$

where $l := \sum_{j=1}^{m} l_j$ and $n := \sum_{j=1}^{m} n_j$. Given these requirements, define the function Φ by

$$\begin{aligned}\Phi\colon \mathbb{R}_{>0} \to \mathbb{R}\colon\ \varrho \longmapsto &\sum_{j=1}^{m} \sum_{x=0}^{n_j-1} \frac{1}{\varrho + \frac{x}{l_j}} - l \log\Big(1 + \tfrac{n}{\varrho l}\Big) \\ &= \sum_{j=1}^{m} l_j \left(\psi(\varrho l_j + n_j) - \psi(\varrho l_j)\right) - l \log\Big(1 + \tfrac{n}{\varrho l}\Big)\end{aligned}$$

with digamma function ψ, $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log(\Gamma(x))$, [AS65, p. 258]. Theorem 4.3.7 verifies that the maximum likelihood estimator of the actual exponent ϱ is the unique root of Φ. When searching for it, an adequate initial approximation of the root is the point

$$\varrho^* := \frac{n}{l} \left(\frac{\sum_{j=1}^{m} \frac{{n_j}^2}{l_j} - \frac{n^2}{l}}{\sum_{j=1}^{m} \frac{n_j}{l_j}} - 1 \right)^{-1},$$

because ϱ^* is a consistent estimator of the exponent ϱ: n/l is the consistent maximum likelihood estimator of the mean parameter μ (see Theorem 4.3.2 and

Theorem 4.3.3) and the quotient of $\sum_{j=1}^{m} \frac{{n_j}^2}{l_j} - \frac{n^2}{l}$ and $\sum_{j=1}^{m} \frac{n_j}{l_j}$ is a consistent estimator of the index of dispersion of N_{num} (see Theorem 3.5.2), which means

$$\varrho^* \xrightarrow{P} \frac{\mathbb{E}[N_{\text{num}}]}{\mathbb{D}[N_{\text{num}}] - 1} = \frac{\mu}{1 + \frac{\mu}{\varrho} - 1} = \varrho \qquad \text{for } m \to \infty.$$

The most immediate way to find the root of Φ is to use a simple bisection method. First, choose a point $\varrho_1 \in \mathbb{R}_{>0}$ large enough such that $\Phi(\varrho_1) < 0$. Since the root of Φ is a maximizer, the sign of Φ changes from positive to negative there. Thus, the actual root must lie within the interval $(0, \varrho_1)$. Next, check for the center point $\varrho_2 := \frac{1}{2}\varrho_1$ whether $\Phi(\varrho_2)$ is positive or negative. In the first case, the actual root must be in (ϱ_2, ϱ_1). On the other hand, if $\Phi(\varrho_2)$ is negative, the root lies within the interval $(0, \varrho_2)$. Select the proper interval and, again, check for the center point whether Φ is positive or negative there, and so on. Repeat these iterations until a sufficient small interval is obtained. The center point of this last interval is taken as an accurate approximation of the actual root of Φ. The pseudocode in Algorithm 1 below describes the procedure in detail.

Algorithm 1 Bisection method for calculating the maximum likelihood estimator of ϱ

$\varrho_0 \leftarrow 0$ ▷ lower bound for exponent parameter
$\varrho_1 \leftarrow \varrho^*$ ▷ initial guess for upper bound for the exponent parameter
while $\Phi(\varrho_1) > 0$ **do** ▷ ensure that ϱ_1 is upper bound for the exponent parameter
    $\varrho_0 \leftarrow \varrho_1$
    $\varrho_1 \leftarrow 2\varrho_1$
end while
while $\varrho_1 - \varrho_0 > 10^{-7}$ **do** ▷ do bisection iterations up to sufficient accuracy
    $\varrho \leftarrow \varrho_0 + \frac{1}{2}(\varrho_1 - \varrho_0)$
    if $\Phi(\varrho) > 0$ **then**
        $\varrho_0 \leftarrow \varrho$
    else if $\Phi(\varrho) < 0$ **then**
        $\varrho_1 \leftarrow \varrho$
    else
        $\varrho_0 \leftarrow \varrho$
        $\varrho_1 \leftarrow \varrho$
    end if
end while
return ϱ ▷ the approximation of the exponent parameter is ϱ

The presented bisection method is simple, precise and reliable, but it also is relatively slow [BF93, p. 40 *et seqq.*]. A faster algorithm is based on the Newton-Raphson method [BF93, p. 56 *et seqq.*]. Two disadvantages of this method are:

1. the derivative of Φ is needed, and 2. in general there is no guarantee that the method converges. However, the derivative can be calculated easily,

$$\begin{aligned}\frac{\mathrm{d}\Phi}{\mathrm{d}\varrho}(\varrho) &= -\sum_{j=1}^{m}\sum_{x=0}^{n_j-1}\left(\varrho+\tfrac{x}{l_j}\right)^{-2}+\frac{n}{\varrho\left(\varrho+\frac{n}{l}\right)}\\ &= \sum_{j=1}^{m} {l_j}^2\big(\psi_1(\varrho l_j+n_j)-\psi_1(\varrho l_j)\big)+\frac{n}{\varrho\left(\varrho+\frac{n}{l}\right)}\end{aligned}$$

with trigamma function ψ_1, $\psi_1(x)=\frac{\mathrm{d}^2}{\mathrm{d}^2x}\log(\Gamma(x))$, [AS65, p. 260]. On the other hand, Φ is strictly decreasing and convex between 0 and its root. Therefore, the geometric interpretation of the Newton-Raphson method [BF93, p. 57] illustrates that the method converges to the actual root of Φ from below as long as the initial approximation ϱ_0 is smaller than the root. Algorithm 2 below shows the corresponding pseudocode.

Algorithm 2 Newton-Raphson method for calculating the maximum likelihood estimator of ϱ

$\varrho_0 \leftarrow \varrho^*$ ▷ initial approximation of exponent parameter
while $\Phi(\varrho_0) < 0$ **do** ▷ ensure that ϱ_0 is to the left of the exponent parameter
    $\varrho_0 \leftarrow \frac{1}{2}\varrho_0$
end while
$\varrho \leftarrow \varrho_0 - \Phi(\varrho_0)/\frac{\mathrm{d}\Phi}{\mathrm{d}\varrho}(\varrho_0)$ ▷ first Newton iteration
while $|\varrho-\varrho_0| > 10^{-7}$ **do** ▷ do Newton iterations up to sufficient accuracy
    $\varrho_0 \leftarrow \varrho$
    $\varrho \leftarrow \varrho_0 - \Phi(\varrho_0)/\frac{\mathrm{d}\Phi}{\mathrm{d}\varrho}(\varrho_0)$
end while
return ϱ ▷ the approximation of the exponent parameter is ϱ

5.3.2. Accuracy

Section 4.3.2 suggests that the maximum likelihood estimator of ϱ is asymptotically efficient. Confidence intervals (see Equation (4.5) on page 85) are specified on the basis of this assumption. In order to validate the asymptotic efficiency, a Monte Carlo simulation was run for sample sizes $m = 10$, 20, 50, 100, 500, 1000, means $\mu = 10^{-4}$, 10^{-3}, 10^{-2} and exponents $\varrho = 10^{-5}$, 10^{-4}, 10^{-3}, 10^{-2} using MATLAB [MAT12]. For each combination of m, μ and ϱ, at first a sample of mileages was generated, $l_1, \ldots, l_m$. These mileages are realizations of either a uniform distribution on $\{1000, 1001, \ldots, 50\,999, 51\,000\}$ or the variate $1000+\lfloor \tilde{L}+1/2 \rfloor$, where $\tilde{L}$ is exponentially distributed with mean 25 000. The term $\lfloor x \rfloor$ means the largest integer not greater than x. After that 10^5 samples were generated, $\big(n_j^{(1)}\big)_{1\le j\le m}, \ldots, \big(n_j^{(10^5)}\big)_{1\le j\le m}$, where each $n_j^{(i)}$ is a realization of a negative binomial distribution with exponent ϱl_j and mean μl_j, $\mathrm{NBin}(\varrho l_j, \mu l_j)$. Finally, for each sample $\big(n_j^{(i)}\big)_{1\le j\le m}$ the maximum likelihood estimator ρ_i was calculated, provided that the estimator exists. If the quotient μ/ϱ is too small, than the simulation generates samples that do not satisfy the condition in Theorem 4.3.7 (cf. Section 5.2.2). In these situations, sample means, sample variances, etc. cannot be determined, and the tables which collect the results (Table B.6, Table B.7 and Table B.8 in the appendix) must remain empty there.

Piegorsch [Pie90] uses the reparametrization $1/\varrho$. His simulation study shows that the maximum likelihood estimator of $1/\varrho$ has a negative bias. This suggests that $\hat{\varrho}_m$ is biased with positive bias. Table B.6 in the appendix confirms this assumption. The mean $\overline{\rho} := 10^{-5}\sum_{i=1}^{10^5}\rho_i$ is always larger than the true parameter value ϱ, provided that the mean exists. The bias slightly decreases if the mean parameter μ increases. If ϱ decreases, $\overline{\rho}$ is closer to ϱ on a percentage basis. The standard deviation $\sqrt{\frac{1}{10^5-1}\sum_{i=1}^{10^5}(\rho_i-\overline{\rho})^2}$ behaves in the same way. Apart from the actual standard deviation of $\hat{\varrho}_m$, Table B.6 lists the term $1/\varrho\sqrt{I_\varrho}$, where I_ϱ denotes the Fisher information concerning ϱ from Theorem 4.2.2 under the assumption of Remark 4.2.3,

$$\begin{aligned} I_\varrho :=& I_{\mathrm{num}}(\varrho,\mu)_{11} \\ =& \sum_{j=1}^{m} l_j{}^2 \left(\mathbb{E}_\vartheta\left[\left(\sum_{n=0}^{\overset{*}{N}{}^{l_j}_{\mathrm{num}}-1}\frac{1}{\varrho l_j+n}\right)^2\right] - \log\left(1+\frac{\mu}{\varrho}\right)^2\right) - \frac{\mu}{\varrho(\varrho+\mu)}\sum_{j=1}^{m} l_j. \end{aligned}$$

Since $\hat{\varrho}_m$ is asymptotically efficient and the information inequality holds, the term $1/\sqrt{I_\varrho}$ is a lower bound for the standard deviation, and they approach each other if the sample size m increases. The square root of the inverse Fisher information, $1/\sqrt{I_\varrho}$, is also illustrated in Figure C.10. The plot verifies that an increase of the term μ/ϱ improves the feasible accuracy of estimate of ϱ.

Table B.7 collects the k-statistics $\hat{k}_1, \ldots, \hat{k}_4$ of the sample $\big(\sqrt{I_\varrho}(\rho_i-\varrho)\big)_{1\le i\le 10^5}$, which are unbiased and consistent estimators of the first four cumulants of $\sqrt{I_\varrho}(\hat{\varrho}_m-\varrho)$ (see Equation (3.7) on page 43). The column "km" refers to the mileage which is drawn from either a uniform distribution (U) or a rounded exponential distribution (Exp). Due to the asymptotic efficiency of ϱ, the statistics $\hat{k}_1$, $\hat{k}_3$ and $\hat{k}_4$ are expected to be approximately equal to 0 for large sample sizes, whereas $\hat{k}_2$ is approximately equal to 1, because these values conform the cumulants of a standard normal distribution [JKB94, p. 89]. All of the entries in Table B.7 exceed the expected values. Consequently, $\hat{\varrho}_m$ is biased with positive bias, right-skewed and leptokurtic [UC11, Cra62, pp. 183/187]. The approximation of the distribution of $\sqrt{I_\varrho}(\hat{\varrho}_m - \varrho)$ through the standard normal distribution is suitable as long as the sample size is large enough. Which sample size is sufficient depends on the actual parameters ϱ and μ. In general, the approximation is more accurate if μ is large and ϱ is small. The accuracy of approximation is illustrated in Figure C.4, Figure C.5 and Figure C.6, too. The plots show samples with uniformly generated mileages.

Table B.8 shows how the discrepancies between the k-statistics from above and the cumulants of the standard normal distribution affect the quantiles of the distribution of $\hat{\varrho}_m$. The term $q_\Delta(1-\alpha)$ denotes the difference between the $(1-\alpha)100\,\%$ population quantile of the distribution of $\hat{\varrho}_m$ and the actual $(1-\alpha)100\,\%$ quantile of the standard normal distribution. The calculation of $q_\Delta(1-\alpha)$ works exactly as described in Section 5.2.1.

5.4. Accuracy of Maximum Likelihood Estimator of μ

The maximum likelihood estimator of μ is just the total number of observed SOLEs divided by the absolute mileage (see Theorem 4.3.2). Theorem 4.3.3 provides that $\hat{\mu}_m$ is efficient, i. e. $\hat{\mu}_m$ is unbiased and the variance is equal to the inverse Fisher information. The Monte Carlo study to investigate the accuracy of $\hat{\varrho}_m$ described in Section 5.3.2 also yields an evaluation of the distribution of $\hat{\mu}_m$. In addition, for the same sample sizes $m = 10, 20, 50, 100, 500, 1000$, means $\mu = 10^{-4}, 10^{-3}, 10^{-2}$ and mileages $(l_1, \ldots, l_m)$ (from either a uniform or rounded exponential distribution), realizations of Poisson variates were drawn, $\mathrm{Poi}(\mu l_1), \ldots, \mathrm{Poi}(\mu l_m)$, using MATLAB [MAT12].

The resultant values in Table B.9 in the appendix affirm that $\hat{\mu}_m$ is efficient. The row "$\varrho = \infty$" refers to the Poisson sample, the column "km" refers to the mileage which is drawn from either a uniform distribution (U) or a rounded exponential distribution (Exp). The term I_μ denotes the Fisher information concerning μ from Theorem 4.2.2,

$$I_\mu := \begin{cases} I_{\mathrm{num}}(\mu) = \frac{1}{\mu}\sum_{j=1}^m l_j, & \text{if } N_{\mathrm{num}} \sim \mathrm{Poi}(\mu), \\ I_{\mathrm{num}}(\varrho,\mu)_{22} = \frac{1}{\mu\left(1+\frac{\mu}{\varrho}\right)}\sum_{j=1}^m l_j, & \text{if } N_{\mathrm{num}} \sim \mathrm{NBin}(\varrho,\mu). \end{cases}$$

Since $\hat{\mu}_m$ is efficient, the square root of the inverse Fisher information

$$\frac{1}{\sqrt{I_\mu}} = \frac{\mu}{\sqrt{\sum_{j=1}^m l_j}}\sqrt{\frac{1}{\mu}+\frac{1}{\varrho}}$$

equates to the actual standard deviation of $\hat{\mu}_m$. In contrast to the estimate of ϱ, an increase of both ϱ and μ improves a decrease of the relative standard deviation $1/\mu\sqrt{I_\mu}$. This is also illustrated in Figure C.11. A large μ means that many SOLEs occur, and so the (relative) estimate of μ is more accurate. On the other hand, the greater the exponent ϱ, the smaller is the variance of N_{num}, and the better μ can be estimated.

Table B.10 collects the first four cumulants of $\sqrt{I_\mu}(\hat{\mu}_m - \mu)$ estimated via k-statistics as described in Section 5.3.2. Again, the efficiency of $\hat{\mu}_m$ is expressed through the fact that the first- and second-order cumulants, which correspond to expectation and variance, are practically equal to 0 and 1 respectively. For small sample sizes, the third- and fourth-order cumulants are slightly positive. If, additionally, ϱ is very small, $\hat{k}_3$ and $\hat{k}_4$ are even larger. Thus, $\sqrt{I_\mu}(\hat{\mu}_m - \mu)$ is slightly right-skewed and leptokurtic [UC11, Cra62, pp. 183/187]. However, both $\hat{k}_3$ and $\hat{k}_4$ converge to 0 very fast if the sample size increases. The accuracy of approximation through the standard normal distribution is illustrated in Figure C.7, Figure C.8 and Figure C.9, too.

5.5. Calculation and Accuracy of the Maximum Likelihood Estimator of (ξ, β) in the Counting Model

5.5.1. Calculation

Let be $(z_{jk})_{\substack{1\le j\le m\\ 1\le k\le d}} \in \mathbb{N}_0{}^{m\times d}$ and $(s_{jk})_{\substack{1\le j\le m\\ 0\le k\le d}} \in (\mathbb{R}_{\ge 0} \cup \{\infty\})^{m\times d+1}$ with $m \in \mathbb{N}$ and $d \in \mathbb{N}_{\ge 3}$ satisfying

$$\sum_{j=1}^{m}\sum_{k=2}^{d-1} z_{jk} > 0 \quad \text{and} \quad 0 = s_{j0} < s_{j1} < \ldots < s_{j,d-1} < s_{jd} = \infty \quad \forall j \in \mathbb{N}_{\le m}.$$

Define the two functions Φ_1 and Φ_2 by

$\Phi_i \colon \mathbb{R}_{\ge 0} \times \mathbb{R}_{>0} \to \mathbb{R}$:

$$(\xi,\beta) \longmapsto \sum_{j=1}^{m}\sum_{k=1}^{d-1} \frac{s_{jk}\,\varphi_i\left(\frac{\xi}{\beta}, s_{jk}\right)}{\beta^2 + \xi\beta s_{jk}} \left(\frac{z_{j,k+1}}{1 - \frac{F_{j,k+1}(\xi,\beta)}{F_{jk}(\xi,\beta)}} - \frac{z_{jk}}{\frac{F_{j,k-1}(\xi,\beta)}{F_{jk}(\xi,\beta)} - 1} \right)$$

where

$$F_{jk}(\xi,\beta) := \begin{cases} \left(1 + \frac{\xi}{\beta} s_{jk}\right)^{-\frac{1}{\xi}}, & \text{if } \xi > 0, \\ \mathrm{e}^{-\frac{1}{\beta} s_{jk}}, & \text{if } \xi = 0, \end{cases}$$

$$\varphi_i(x,a) := \mathbb{1}_{\{2\}}(i) + \mathbb{1}_{\{1\}}(i) \cdot \begin{cases} \frac{1}{x}\left(\log(1+xa)\left(1 + \frac{1}{xa}\right) - 1\right), & \text{if } xa > 0, \\ \frac{a}{2}, & \text{if } xa = 0, \end{cases}$$

($j \in \mathbb{N}_{\le m}$, $k \in \{0, \dots, d\}$, $i \in \{1,2\}$). Given these requirements, Proposition 4.4.3, Theorem 4.4.4 and Corollary 4.4.5 provide the following procedure for finding the maximum likelihood estimator of (ξ, β): at first, calculate the unique root of $\Phi_2(0,\cdot)$. If β_0 denotes this root, check whether $\Phi_1(0,\beta_0)$ is positive or not. $\Phi_1(0,\beta_0) \le 0$ means that $(0,\beta_0)$ is the maximum likelihood estimator. Otherwise, the joint root of Φ_1 and Φ_2 is the maximum likelihood estimator of (ξ, β). This root can be found by means of a bisection method. Therefore, take a ξ_1 large enough such that for the root β_1 of $\Phi_2(\xi_1,\cdot)$ it holds $\Phi_1(\xi_1,\beta_1) < 0$. Thus, the actual shape parameter ξ must lie within the interval $(0,\xi_1)$. Next, calculate for the center point $\xi_2 := \frac{1}{2}\xi_1$ the root β_2 of $\Phi_2(\xi_2,\cdot)$. If $\Phi_1(\xi_2,\beta_2)$ is positive, the actual shape parameter ξ must be in (ξ_2,ξ_1), and if $\Phi_1(\xi_2,\beta_2)$ is negative, the actual shape parameter ξ lies within $(0,\xi_2)$. Select the proper interval, take the center point of this interval, and so on. Repeat these iterations until a sufficient small interval is obtained. If ξ_n is the center point of this last interval and β_n is the root of $\Phi_2(\xi_n,\cdot)$, then (ξ_n,β_n) is an accurate approximation of the actual joint root of Φ_1 and Φ_2. The pseudocode in Algorithm 3 below describes this procedure in detail.

The only outstanding point is the calculation of the roots of $\Phi_2(\xi,\cdot)$ with ξ given. Similar to the function Φ in Section 5.3.1, also the function $\beta \mapsto \Phi_2(\xi,\beta)$ is strictly decreasing and convex between 0 and its unique root. Thus, the Newton-Raphson method [BF93, p. 56 *et seqq.*] is an appropriate tool for finding the root of $\Phi_2(\xi,\cdot)$ as long as the initial approximation is to the left of the actual root. A good guess for this initial approximation is the minimum of the relative class limits $s_{11}, \dots, s_{m1}$, because the second statement of Proposition 4.4.3 ensures that in some situations $\min\{s_{j1} \mid 1 \le j \le m\}$ is indeed smaller than the actual root. Finally, the Newton-Raphson method needs the partial derivative of Φ_2 with respect to the second dimension,

$$\begin{aligned}\tfrac{\partial\Phi_2}{\partial\beta}(\xi,\beta) = &-\sum_{j=1}^{m}\sum_{k=1}^{d-1}\frac{2\beta s_{jk}+\xi {s_{jk}}^2}{(\beta^2+\xi\beta s_{jk})^2}\left(\frac{z_{j,k+1}}{1-\frac{F_{j,k+1}(\xi,\beta)}{F_{jk}(\xi,\beta)}} - \frac{z_{jk}}{\frac{F_{j,k-1}(\xi,\beta)}{F_{jk}(\xi,\beta)}-1}\right)\\ &+\sum_{j=1}^{m}\sum_{k=1}^{d-1}\frac{\beta s_{jk}}{(\beta^2+\xi\beta s_{jk})^2}\Bigg(\underbrace{\frac{z_{j,k+1}\,\frac{s_{j,k+1}-s_{jk}}{\beta+\xi s_{j,k+1}}}{\frac{F_{jk}(\xi,\beta)}{F_{j,k+1}(\xi,\beta)}+\frac{F_{j,k+1}(\xi,\beta)}{F_{jk}(\xi,\beta)}-2}}_{\text{without this part if } k=d-1}\\ &\qquad - \frac{z_{jk}\,\frac{s_{jk}-s_{j,k-1}}{\beta+\xi s_{j,k-1}}}{\frac{F_{jk}(\xi,\beta)}{F_{j,k-1}(\xi,\beta)}+\frac{F_{j,k-1}(\xi,\beta)}{F_{jk}(\xi,\beta)}-2}\Bigg).\end{aligned}$$

Algorithm 4 below repeats the Newton-Raphson method for calculating the root of $\Phi_2(\xi,\cdot)$.

Algorithm 3 Bisection method for calculating the maximum likelihood estimator of (ξ, β)

$\beta \leftarrow$ root of $\Phi_2(0, \cdot)$ ▷ initial approximation of scale parameter; see Algorithm 4
if $\Phi_1(0, \beta) \leq 0$ **then** ▷ check initial approximation of scale parameter
    $\xi \leftarrow 0$
else
    $\xi_0 \leftarrow 0$ ▷ lower bound for shape parameter
    $\xi_1 \leftarrow 1$ ▷ initial guess for upper bound for shape parameter
    $\beta \leftarrow$ root of $\Phi_2(\xi_1, \cdot)$ ▷ see Algorithm 4
    while $\Phi_1(\xi_1, \beta) > 0$ **do** ▷ ensure that ξ_1 is an upper bound for shape parameter
        $\xi_0 \leftarrow \xi_1$
        $\xi_1 \leftarrow 2\xi_1$
        $\beta \leftarrow$ root of $\Phi_2(\xi_1, \cdot)$ ▷ see Algorithm 4
    end while
    while $\xi_1 - \xi_0 > 10^{-7}$ **do** ▷ do bisection iterations up to sufficient accuracy
        $\xi \leftarrow \xi_0 + \frac{1}{2}(\xi_1 - \xi_0)$
        $\beta \leftarrow$ root of $\Phi_2(\xi, \cdot)$ ▷ see Algorithm 4
        if $\Phi_1(\xi, \beta) > 0$ **then**
            $\xi_0 \leftarrow \xi$
        else if $\Phi_1(\xi, \beta) < 0$ **then**
            $\xi_1 \leftarrow \xi$
        else
            $\xi_0 \leftarrow \xi$
            $\xi_1 \leftarrow \xi$
        end if
    end while
end if
return ξ, β ▷ the approximation of shape and scale is (ξ, β)

Algorithm 4 Newton-Raphson method for calculating the root of $\Phi_2(\xi, \cdot)$

$\beta_0 \leftarrow \min\{s_{j1} | 1 \leq j \leq m\}$ ▷ initial approximation of the root
while $\Phi_2(\xi, \beta_0) < 0$ **do** ▷ ensure that β_0 is to the left of the root
    $\beta_0 \leftarrow \frac{1}{2}\beta_0$
end while
$\beta \leftarrow \beta_0 - {}^{\Phi_2(\xi,\beta_0)}/_{\frac{\partial \Phi_2}{\partial \beta}(\xi,\beta_0)}$ ▷ first Newton iteration
while $|\beta - \beta_0| > 10^{-7}$ **do** ▷ do Newton iterations up to sufficient accuracy
    $\beta_0 \leftarrow \beta$
    $\beta \leftarrow \beta_0 - {}^{\Phi_2(\beta_0)}/_{\frac{\partial \Phi_2}{\partial \beta}(\xi,\beta_0)}$
end while
return β ▷ the approximation of the actual root is β

5.5.2. Accuracy in Case of $\xi \in \mathbb{R}_{>0}$

In order to analyze the accuracy of the maximum likelihood estimator $(\hat{\xi}_m, \hat{\beta}_m)$, a Monte Carlo simulation was run using MATLAB [MAT12]. For each combination of sample sizes $m = 20$, 50, 100, shape parameters $\xi = 0.5$, 1, scale parameters $\beta = 1$, 3, 5, average numbers of SOLEs $\mu = 10^{-4}$, 10^{-3} and numbers of classes $d = 4$, 6 with partitions $(s_{j0}, \dots, s_{j4}) = (0, 4, 8, 12, \infty)$ and $(s_{j0}, \dots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$, at first a sample of mileages were generated, $l_1, \dots, l_m$. These mileages are realizations of either a uniform distribution on $\{1000, 1001, \dots, 50\,999, 51\,000\}$ or the variate $1000 + \lfloor \tilde{L} + 1/2 \rfloor$, where $\tilde{L}$ is exponentially distributed with mean 25 000. The term $\lfloor x \rfloor$ means the largest integer not greater than x. Hereafter, 10^5 samples were generated, $\big(z_{jk}^{(1)}\big)_{\substack{1\le j\le m\\ 1\le k\le d}}, \dots, \big(z_{jk}^{(10^5)}\big)_{\substack{1\le j\le m\\ 1\le k\le d}}$, where each $z_{jk}^{(i)}$ is a realization of a Poisson distribution with mean

$$\mu l_j \left(\left(1 + \tfrac{\xi}{\beta}\, s_{j,k-1}\right)^{-\frac{1}{\xi}} - \left(1 + \tfrac{\xi}{\beta}\, s_{jk}\right)^{-\frac{1}{\xi}} \right).$$

Finally, for each sample $\big(z_{jk}^{(i)}\big)_{\substack{1\le j\le m\\ 1\le k\le d}}$ the maximum likelihood estimator $(\hat{\xi}_m, \hat{\beta}_m)$ was calculated by means of the method described above in Section 5.5.1.

Table B.11 and Table B.12 in the appendix list sample means and sample standard deviations of the resultant values. Both $\hat{\xi}_m$ and $\hat{\beta}_m$ slightly overestimate the true parameter values. The higher μ or m, the more SOLEs can be observed, and the more accurate the estimates are. If the sample size m is small, it may occur that all medium classes remain empty and the maximum likelihood estimator does not exist (see Theorem 4.4.4). Then, sample mean and sample variance cannot be calculated. An increase of number of classes causes a decrease of the variances of $\hat{\xi}_m$ and $\hat{\beta}_m$, but not for all settings the bias decreases, too. The corresponding entries of the inverse Fisher information matrix,

$$\begin{aligned} J_\xi &:= \left[I_{\mathrm{sev}}(\mu, \xi, \beta)^{-1} \right]_{11} \\ &= \frac{\sum_{k=1}^{d-1} \frac{a_{21k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)}}{\mu \sum_{j=1}^{m} l_j \left(\prod_{i=1}^{2} \sum_{k=1}^{d-1} \frac{a_{i1k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)} - \left(\sum_{k=1}^{d-1} \frac{a_{11k}(\xi,\beta)\, a_{21k}(\xi,\beta)}{b_{1k}(\xi,\beta)} \right)^2 \right)}, \end{aligned}$$

and

$$\begin{aligned} J_\beta &:= \left[I_{\mathrm{sev}}(\mu, \xi, \beta)^{-1} \right]_{22} \\ &= \frac{\sum_{k=1}^{d-1} \frac{a_{11k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)}}{\mu \sum_{j=1}^{m} l_j \left(\prod_{i=1}^{2} \sum_{k=1}^{d-1} \frac{a_{i1k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)} - \left(\sum_{k=1}^{d-1} \frac{a_{11k}(\xi,\beta)\, a_{21k}(\xi,\beta)}{b_{1k}(\xi,\beta)} \right)^2 \right)} \end{aligned}$$

(see Theorem 4.2.2, a_{ijk}, b_{jk} as defined there), approximate the actual variances of $\hat{\xi}_m$ and $\hat{\beta}_m$ respectively very well. This fact justifies the approximate confidence intervals in Section 4.4.3 (see Equation (4.7) on page 101).

The confidence intervals of ξ and β in Section 4.4.3 are based on the assumption that both ${(\hat{\xi}_m-\xi)}/{\sqrt{J_\xi}}$ and ${(\hat{\beta}_m-\beta)}/{\sqrt{J_\beta}}$ are asymptotically standard normally distributed. Table B.13, Table B.14, Table B.15 and Table B.16 in the appendix confirm this assumption. The tables collect the first four cumulants of ${(\hat{\xi}_m-\xi)}/{\sqrt{J_\xi}}$ and ${(\hat{\beta}_m-\beta)}/{\sqrt{J_\beta}}$ estimated via k-statistics as described in Section 5.3.2. Similar to the maximum likelihood estimators of μ and ϱ (see Section 5.4 and Section 5.3.2) both $\hat{\xi}_m$ and $\hat{\beta}_m$ are most often slightly right-skewed and leptokurtic, since the third- and fourth-order cumulants are positive. Only for $\beta = 1$ and $(s_{j0}, \dots s_{j4}) = (0, 4, 8, 12, \infty)$ the fourth-order cumulant of ${(\hat{\beta}_m-\beta)}/{\sqrt{J_\beta}}$ is negative if the sample size m is small. The standard normal distribution, which has a first-, third- and fourth-order cumulant of value 0 and a second-order cumulant of value 1 [JKB94, p. 89], approximates the terms ${(\hat{\xi}_m-\xi)}/{\sqrt{J_\xi}}$ and ${(\hat{\beta}_m-\beta)}/{\sqrt{J_\beta}}$ very well if either μ or m is not too small. This is also illustrated by Figure C.12, Figure C.13, Figure C.14 and Figure C.15 (concerning ξ) as well as in Figure C.16, Figure C.17, Figure C.18 and Figure C.19 (concerning β).

5.5.3. Accuracy in Case of $\xi = 0$

The Monte Carlo simulation described in Section 5.5.2 was also run for $\xi = 0$. In order to analyze the heuristic confidence intervals from Section 4.4.3 (see Equation (4.9) on page 103), this time the simulation was run for sample sizes $m = 20, 50, 100, 500, 1000$ and $\mu = 10^{-4}, 10^{-3}, 10^{-2}$. Table B.17 in the appendix collects the results concerning $\hat{\xi}_m$. Besides the sample mean and sample standard deviation of the 10^5 resultant values, the table lists the number of realizations of $\hat{\xi}_m$ which are equal to 0. By considerations of Section 4.4.3, the probability of the event $\{\hat{\xi}_m = 0\}$ is expected to be 50 %. For large β and, of course, for large μ and m, this assessment is true. Since β represents the average severity of an arbitrary SOLE (see Definition 2.4.2), a small β means that only the lower classes get filled. But for an observation $z \in \mathbb{N}_0{}^d$ with $\sum_{k=3}^{d} z_k = 0$ it always holds $\hat{\xi}_m(z) = 0$ or the maximum likelihood estimator does not exist (see Section 4.4.4). If, in addition, the classes are too big or only less SOLEs can be observed, the frequency of realizations with $\hat{\xi}_m = 0$ can greatly exceed 50 %.

Because the shape is bounded below by 0, $\hat{\xi}_m$ must overestimate the true parameter value $\xi = 0$. The discussion in Section 4.4.3 leads up to expect that for less observations $\hat{\beta}_m$ is biased, too, and the bias is negative. Table B.18 verifies this fact.

Section 4.4.3 provides approximate confidence intervals of ξ $(= 0)$ and β based on the functions Φ_ξ and Φ_β (see Equation (4.8) on page 103). Φ_ξ is taken as an approximation of the cumulative distribution function of $\hat{\xi}_m$, and Φ_β estimates the cumulative distribution function of $\hat{\beta}_m$ focused on the extreme quantiles. According to this, $\Phi_\xi^{-1}(p)$ is an approximation of the $100p\,\%$ quantile of $\hat{\xi}_m$ ($p \in [0.5, 1)$) and $\Phi_\xi^{-1}(p)$ is an approximation of the $100p\,\%$ quantile of $\hat{\beta}_m$ ($p \in (0,1)$) when Φ_ξ^{-1} and Φ_β^{-1} denote the inverse of $\Phi_\xi \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 1/2}$ and Φ_β respectively. To show the accuracy of this approximation, the values of $\Phi_\xi^{-1}(p)$ and $\Phi_\beta^{-1}(p)$ can be compared to the sample quantiles of the 10^5 realizations of $\hat{\xi}_m$ and $\hat{\beta}_m$. The calculation of these sample quantiles works as described in Section 5.2.1.

Table B.19 and Table B.20 collect this comparison with regard to $\hat{\xi}_m$. The columns "p-Q" contain the $100p\,\%$ sample quantiles of the 10^5 realizations of $\hat{\xi}_m$. For the sake of clarity, the table only lists the results with uniformly drawn mileages. The samples with exponentially distributed mileages has very similar quantiles. The table shows that Φ_ξ^{-1} approximates the quantiles of $\hat{\xi}_m$ very well. For $\beta = 1$ the extreme quantiles of $\hat{\xi}_m$, i. e. $p \geq 0.9$, are slightly overestimated if only less observations are available. On the other hand, $\Phi_\xi^{-1}(p)$ is sometimes a little bit to small if $\beta \in \{3, 5\}$.

Table B.21, Table B.22, Table B.23 and Table B.24 in the appendix show the comparison of the quantiles with regard to $\hat{\beta}_m$. Also here, the actual quantiles of $\hat{\beta}_m$ and the values $\Phi_\beta^{-1}(p)$ are quite compatible with each other, especially for the extreme quantiles, i. e. $p \leq 0.1$ and $p \geq 0.9$. Generally, Φ_β^{-1} slightly underestimates the true quantiles of $\hat{\beta}_m$.

As an example, the distribution of $\hat{\xi}_m$ and $\hat{\beta}_m$ from the Monte-Carlo simulation is plotted in case of $(s_{j0}, \ldots, s_{jd}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$, $\xi = 0$, $\beta = 1$, $\mu = 10^{-2}$ and $m = 100$. Figure C.20 in the appendix shows the frequency distribution and the empirical distribution function of $\hat{\xi}_m$. The plots verify that Φ_ξ approximates the distribution of ξ very well. Figure C.21 shows the same plot with regard to $\hat{\beta}_m$. Remember that the approximation Φ_β is put together of $\Phi_{\beta-}$ and $\Phi_{\beta+}$ (see Section 4.4.3): $\Phi_{\beta-}$ and $\Phi_{\beta+}$ are assumed to be the approximate cumulative distribution functions of $\hat{\beta}_m$ given $\hat{\xi}_m = 0$ and of $\hat{\beta}_m$ given $\hat{\xi}_m > 0$, respectively, so that $\Phi_\beta = {}^{(\Phi_{\beta-} + \Phi_{\beta+})}/_2$ is an approximation of the cumulative distribution function of $\hat{\beta}_m$. Figure C.22 verifies that $\Phi_{\beta-}$ is an adequate approximation. Figure C.23 shows that at least the lower quantiles of $\hat{\beta}_m | \hat{\xi}_m = 0$ are well approximated by $\Phi_{\beta+}$. Together, as can be seen in Figure C.21, Φ_β is practical for the calculation of confidence intervals of $\hat{\beta}_m$ as described in Section 4.4.3.

5.5.4. Optimal Class Limits

Section 4.4.4 deals with equidistant class limits and provides the existence of an optimal class length. If there is a $\Lambda \in \mathbb{R}_{>0}$ such that $s_{jk} = t_{jk} - u_{\text{sev}} = k\Lambda$ for all $j \in \mathbb{N}_{\leq m}$ and $k \in \{0, \ldots, d-1\}$, the optimal class length Λ_{opt} is due to Definition 4.4.8 the unique maximizer of the determinant of the Fisher information matrix concerning ξ and β as function with respect to Λ,

$$\begin{aligned}&\det\bigl(I_{\text{sev}}(\mu,\xi,\beta)\bigr)\\&\quad=\left(\mu\sum_{j=1}^{m} l_j\right)^2\left(\prod_{i=1}^{2}\sum_{k=1}^{d-1}\frac{a_{i1k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)}-\left(\sum_{k=1}^{d-1}\frac{a_{11k}(\xi,\beta)\,a_{21k}(\xi,\beta)}{b_{1k}(\xi,\beta)}\right)^2\right)\end{aligned}$$

(see Theorem 4.2.2; for ξ and β constant, $a_{ijk}(\xi,\beta)$ and $b_{jk}(\xi,\beta)$ are functions with respect to the relative class limits $s_{11}, \ldots, s_{1,d-1}$). The actual value of Λ_{opt} only depends on the parameter values ξ and β. The terms $b_{jk}(\xi,\beta)/\beta^4$, $a_{1jk}(\xi,\beta)/\beta^2$ and $a_{2jk}(\xi,\beta)/\beta$ do not really depend on β and Λ but on the quotient Λ/β. Therefore, if Λ is replaced by $\beta\Lambda$, then

$$\begin{aligned}&\Lambda \to \beta\Lambda\\&\rightsquigarrow\quad \frac{b_{jk}(\xi,\beta)}{\beta^4}\to b_{jk}(\xi,1),\quad \frac{a_{1jk}(\xi,\beta)}{\beta^2}\to a_{1jk}(\xi,1),\quad \frac{a_{2jk}(\xi,\beta)}{\beta}\to a_{2jk}(\xi,1)\\&\rightsquigarrow\quad \det\bigl(I_{\text{sev}}(\mu,\xi,\beta)\bigr)\,\beta^2 \;\to\; \det\bigl(I_{\text{sev}}(\mu,\xi,1)\bigr).\end{aligned}\tag{5.1}$$

Consequently, Λ_{opt} is the optimal class length for shape ξ and scale 1 if and only if $\beta\Lambda_{\text{opt}}$ is the optimal class length for shape ξ and scale β, i. e. the optimal class length is in linear proportion to the scale parameter β. For this reason, it is sufficient to calculate Λ_{opt} in case of $\beta = 1$.

A Newton-Raphson method [BF93, p. 56 *et seqq.*] was implemented in MATLAB [MAT12] to calculate Λ_{opt} for different numbers of classes $d = 3, 4, 5, 6, 7, 8$ and shape parameters $\xi = 0, 0.1, 0.2, 0.3, 0.4, 0.5$. Table B.30 in the appendix lists the resultant values. As expected, Λ_{opt} increases if either ξ increases or d decreases. The larger ξ, the more extreme SOLEs may occur. Hence, the class length must increase to offer as much of these extreme SOLEs as possible. On the other hand, if d decreases, the class length must increase to offer the same range as before.

Besides the optimal class length Λ_{opt}, Table B.30 lists the quantiles of the resultant (medium) relative class limits,

$$F^{-1}_{\text{GPar}(\xi,\beta)}\bigl(\Lambda_{\text{opt}}\bigr),\quad F^{-1}_{\text{GPar}(\xi,\beta)}\bigl(2\Lambda_{\text{opt}}\bigr),\quad \ldots,\quad F^{-1}_{\text{GPar}(\xi,\beta)}\bigl((d-1)\Lambda_{\text{opt}}\bigr),$$

where $F^{-1}_{\text{GPar}(\xi,\beta)}$ denotes the inverse of the cumulative distribution function of the generalized Pareto distribution with shape ξ and scale β (see Definition 2.4.2). It is striking that all the quantiles $F^{-1}_{\text{GPar}(\xi,\beta)}\bigl(k\Lambda_{\text{opt}}\bigr)$ decrease if ξ increases, except for $k = 1$ if $d \in \{7, 8\}$.

The demand for equidistant class limits is a strict constraint. Without this constraint, it may be possible to find a configuration of class limits which yields a larger value for $\det(I_{\mathrm{sev}}(\mu, \xi, \beta))$ than any configuration of equidistant class limits. The terms $a_{ijk}(\xi, \beta)$ as functions with respect to the class limits are for all $j \in \mathbb{N}_{\leq m}$ the same. Also the terms $b_{jk}(\xi, \beta)$ as functions with respect to the class limits do not differ from each other for distinct j. Thus, without loss of generality, let be $m = 1$. Since the parameter μ and the mileage l_1 do not influence the maximizer of $\det(I_{\mathrm{sev}}(\mu, \xi, \beta))$, without loss of generality let be $\mu = 1$ and $l_1 = 1$, too. Eventually, for the same reason than above, it is sufficient to find a set of relative class limits $s = (s_{11}, \ldots, s_{1,d-1}) \in {\mathbb{R}_{>0}}^{d-1}$ with $0 < s_{11} < \ldots < s_{1,d-1} < \infty$ maximizing the determinant

$$\det\big(I_{\mathrm{sev}}(1, \xi, 1)\big) = \prod_{i=1}^{2} \sum_{k=1}^{d-1} \frac{a_{i1k}(\xi, 1)^2}{b_{1k}(\xi, 1)} - \left(\sum_{k=1}^{d-1} \frac{a_{11k}(\xi, 1)\, a_{21k}(\xi, 1)}{b_{1k}(\xi, 1)} \right)^2 ,$$

because then the set $\beta s = (\beta s_{11}, \ldots, \beta s_{1,d-1})$ is a maximizer of $\det(I_{\mathrm{sev}}(\mu, \xi, \beta))$.

Figure 5.1.: *Determinant of Fisher information concerning ξ and β for every tenth of $2 \cdot 10^6$ randomly generated class limit configurations.*

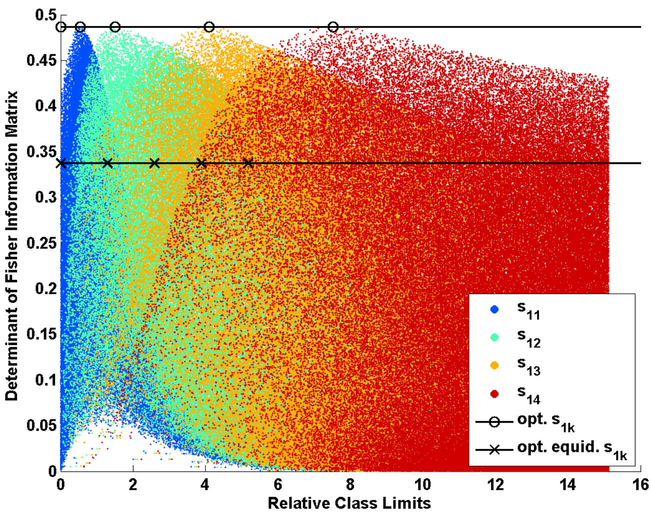
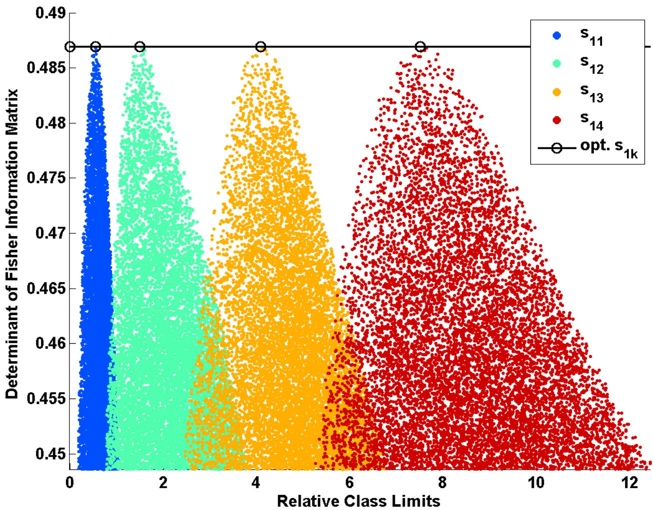


Figure 5.2.: *Determinant of Fisher information concerning ξ and β for top 10^4 of $2 \cdot 10^6$ randomly generated class limit configurations.*



For numbers of classes $d = 3,\ 4,\ 5,\ 6,\ 7,\ 8$ and shapes $\xi = 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5$ Monte Carlo simulations were run using MATLAB [MAT12]. For each combination of d and ξ, $(d-1)$ independent realizations $u_1, \ldots, u_{d-1}$ of U were generated, where U is uniformly distributed on $\big(0,\ F^{-1}_{\mathrm{GPar}(\xi,1)}(0.9999)\big)$. Finally, the determinant $\det(I_{\mathrm{sev}}(1,\xi,1))$ was calculated based on the relative class limits $(s_{11}, \ldots, s_{1,d-1}) = (u_{(1)}, \ldots, u_{(d-1)})$, where $u_{(i)}$ denotes the ith smallest value in $\{u_1, \ldots, u_{d-1}\}$. This procedure was repeated $2 \cdot 10^6$ times.

The two million configurations of relative class limits and the corresponding values of the determinant $\det(I_{\mathrm{sev}}(1,\xi,1))$ can be plotted in the following way: suppose, $(\tilde{s}_{11}, \ldots, \tilde{s}_{1,d-1})$ is one of the two million class limit configurations, then all the values $\tilde{s}_{1k}$ are plotted in a scatter plot against the value $\det(I_{\mathrm{sev}}(1,\xi,1))$. As an example, the resultant scatter plot in case of $\xi = 0.1$ and $d = 5$ is illustrated in Figure 5.1 and Figure 5.2. For the sake of clarity, in Figure 5.1 only every tenth of the two million configurations is plotted, and in Figure 5.2 the top ten thousand configurations are plotted. For all the other combinations of ξ and d, the corresponding scatter plots look very similar to that.

The plots show that the best class limit configurations seem to tend to one optimal configuration. Besides the Monte Carlo class limits, Figure 5.1 shows the optimal equidistant relative class limits as calculated above (marked by crosses), and it shows a configuration which may be the optimal one (marked by circles).

This last configuration was found by a sequential quadratic programming algorithm pre-implemented in MATLAB [MAT12] (command `fmincon` with `sqp` algorithm which works as described in Chapter 18 of Nocedal and Wright [NW06]). The algorithm finds the minimum of a constrained nonlinear multivariable function. Applied to the function $-\det(I_\mathrm{sev}(1,\xi,1))$, for all combinations of d and ξ adequate approximations of the actual optimal class limit configurations could be calculated. Table B.29 in the appendix lists the results of this analysis.

The optimal class limits compose very small lower classes, and the upper classes are growing ever larger. This ensures that not all SOLEs will lie within the lower classes. The high quantiles of the generalized Pareto distribution on the basis are well covered. At least the half of the relative class limits $s_{11}, \ldots, s_{1,d-1}$, namely $s_{1,\lceil \frac{d}{2} \rceil}, \ldots, s_{1,d-1}$, are larger than the 88 % quantile. The optimal class limit configuration is a good compromise: the lower classes are small enough such that not all SOLEs lie within them, and the upper classes are correctly dimensioned such that the probability of an empty class is relatively small and, simultaneously, the high quantiles are well covered.

Changing the class limits during the experiment or adjusting arbitrary class limits instead of equidistant ones can involve considerable technical effort. One may ask if this effort will pay off. The observation period which is necessary to get as small confidence intervals as desired can be used as an criterion to decide whether it is worth the effort or not.

Suppose, the observations of all vehicles are based on the same relative class limit configuration $\bar{s} = (0, \bar{s}_{11}, \ldots, \bar{s}_{1,d-1}, \infty) \in \{0\} \times \mathbb{R}_{>0}{}^{d-1} \times \{\infty\}$. The total mileage $\sum_{j=1}^{m} l_j$ of the vehicles shall be denoted by $\bar{l}$. On the other hand, there is a second relative class limit configuration $\tilde{s} = (0, \tilde{s}_{11}, \ldots, \tilde{s}_{1,d-1}, \infty) \in \{0\} \times \mathbb{R}_{>0}{}^{d-1} \times \{\infty\}$, e. g. the optimal one. The question is how long the vehicles must be observed such that the (approximate) confidence intervals of ξ and β have the same size as in case of the configuration $\bar{s}$. The size of the confidence intervals from Section 4.4.3 (see Equation (4.7) on page 101) is in linear proportion to the terms $\sqrt{J_\xi}$ and $\sqrt{J_\beta}$ respectively from above (denoted by $\sigma_{\xi,m}$ and $\sigma_{\beta,m}$ respectively in Section 4.4.3). Thus, if the class limit configuration is changed, a new mileage $\tilde{l}$ must be chosen such that the term J_ξ remained constant. In

Table 5.1.: ***Ratio between the mileage l_opt in case of optimal class limits and the mileage l_eopt in case of optimal equidistant class limits such that the confidence interval of ξ is equal in both cases.***

d	l_opt with regard to ξ					
	$\xi = 0$	$\xi = 0.1$	$\xi = 0.2$	$\xi = 0.3$	$\xi = 0.4$	$\xi = 0.5$
3	0.50 l_eopt	0.44 l_eopt	0.40 l_eopt	0.37 l_eopt	0.34 l_eopt	0.32 l_eopt
4	0.69 l_eopt	0.62 l_eopt	0.56 l_eopt	0.51 l_eopt	0.47 l_eopt	0.44 l_eopt
5	0.80 l_eopt	0.72 l_eopt	0.66 l_eopt	0.60 l_eopt	0.56 l_eopt	0.52 l_eopt
6	0.84 l_eopt	0.77 l_eopt	0.70 l_eopt	0.65 l_eopt	0.61 l_eopt	0.57 l_eopt
7	0.87 l_eopt	0.81 l_eopt	0.75 l_eopt	0.70 l_eopt	0.65 l_eopt	0.61 l_eopt
8	0.89 l_eopt	0.83 l_eopt	0.78 l_eopt	0.73 l_eopt	0.68 l_eopt	0.64 l_eopt

Table 5.2.: *Ratio between the mileage l_{opt} in case of optimal class limits and the mileage l_{eopt} in case of optimal equidistant class limits such that the confidence interval of β is equal in both cases.*

d	l_{opt} with regard to β					
	$\xi = 0$	$\xi = 0.1$	$\xi = 0.2$	$\xi = 0.3$	$\xi = 0.4$	$\xi = 0.5$
3	$0.50\ l_{\text{eopt}}$	$0.46\ l_{\text{eopt}}$	$0.43\ l_{\text{eopt}}$	$0.41\ l_{\text{eopt}}$	$0.39\ l_{\text{eopt}}$	$0.38\ l_{\text{eopt}}$
4	$0.73\ l_{\text{eopt}}$	$0.68\ l_{\text{eopt}}$	$0.63\ l_{\text{eopt}}$	$0.60\ l_{\text{eopt}}$	$0.58\ l_{\text{eopt}}$	$0.56\ l_{\text{eopt}}$
5	$0.77\ l_{\text{eopt}}$	$0.71\ l_{\text{eopt}}$	$0.67\ l_{\text{eopt}}$	$0.64\ l_{\text{eopt}}$	$0.61\ l_{\text{eopt}}$	$0.59\ l_{\text{eopt}}$
6	$0.83\ l_{\text{eopt}}$	$0.78\ l_{\text{eopt}}$	$0.73\ l_{\text{eopt}}$	$0.70\ l_{\text{eopt}}$	$0.68\ l_{\text{eopt}}$	$0.65\ l_{\text{eopt}}$
7	$0.86\ l_{\text{eopt}}$	$0.81\ l_{\text{eopt}}$	$0.77\ l_{\text{eopt}}$	$0.73\ l_{\text{eopt}}$	$0.71\ l_{\text{eopt}}$	$0.68\ l_{\text{eopt}}$
8	$0.88\ l_{\text{eopt}}$	$0.83\ l_{\text{eopt}}$	$0.79\ l_{\text{eopt}}$	$0.76\ l_{\text{eopt}}$	$0.74\ l_{\text{eopt}}$	$0.71\ l_{\text{eopt}}$

other words:

$$1 = \frac{J_\xi \Big|_{\substack{\text{class limits } \tilde{s} \\ \text{tot. mileage } \tilde{l}}}}{J_\xi \Big|_{\substack{\text{class limits } \bar{s} \\ \text{tot. mileage } \bar{l}}}}$$

$$\Leftrightarrow \qquad \frac{\tilde{l}}{\bar{l}} = \frac{\left.\frac{\sum_{k=1}^{d-1} \frac{a_{21k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)}}{\left(\prod_{i=1}^{2} \sum_{k=1}^{d-1} \frac{a_{i1k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)} - \left(\sum_{k=1}^{d-1} \frac{a_{11k}(\xi,\beta)\, a_{21k}(\xi,\beta)}{b_{1k}(\xi,\beta)}\right)^2\right)}\right|_{\text{class limits } \tilde{s}}}{\left.\frac{\sum_{k=1}^{d-1} \frac{a_{21k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)}}{\left(\prod_{i=1}^{2} \sum_{k=1}^{d-1} \frac{a_{i1k}(\xi,\beta)^2}{b_{1k}(\xi,\beta)} - \left(\sum_{k=1}^{d-1} \frac{a_{11k}(\xi,\beta)\, a_{21k}(\xi,\beta)}{b_{1k}(\xi,\beta)}\right)^2\right)}\right|_{\text{class limits } \bar{s}}}.$$

Analogously, a sufficient mileage for J_β can be found.

As an example, suppose that $\bar{s}$ is the optimal configuration for equidistant class limits, and the covered distance is $\bar{l} = l_{\text{eopt}}$. If $\tilde{s}$ is chosen to be the absolute optimal class limit configuration, the sufficient mileage $\tilde{l} = l_{\text{opt}}$ for equal-sized confidence intervals is partly considerable smaller than l_{eopt}. Table 5.1 lists the ratios between l_{opt} and l_{eopt} for confidence intervals of the shape parameter ξ. Since the optimal relative class limits are in linear proportion to the scale parameter β, these results hold for all values of β. For $\xi = 0$, l_{opt} is calculated as in case of a positive shape using the confidence interval C_ξ (see Equation (4.7) on page 101). The table shows that it is possible to save up to 68 % of observation time if the optimal class limit configuration is used instead of the optimal equidistant one. The more classes are chosen, the less time can be saved, because in case of many classes the equidistant class limits cover the observation range quite well. Table 5.2 collects the particular mileages for the confidence intervals of β, and it shows similar ratios.

5.5.5. Comparison with Uncensored Generalized Pareto Model

Even if the optimal class limit configuration from Section 5.5.4 is used, the grouping of the SOLEs is accompanied by a loss of information. In order to quantify this loss of information, let us compare the inverse Fisher information terms J_ξ and J_β from Section 5.5.2, which corresponds to the approximate variances of the maximum likelihood estimators $\hat{\xi}_m$ and $\hat{\beta}_m$, with the inverse Fisher information of an uncensored generalized Pareto distribution. According to Smith [Smi84], the Fisher information of a generalized Pareto experiment is

$$I_{\text{GPar}}(\xi, \beta) = \begin{pmatrix} \frac{2}{(1+\xi)(1+2\xi)} & -\frac{1}{\beta(1+\xi)(1+2\xi)} \\ -\frac{1}{\beta(1+\xi)(1+2\xi)} & \frac{1}{\beta^2(1+2\xi)} \end{pmatrix}.$$

Consequently, the inverse Fisher information is

$$I_{\text{GPar}}(\xi, \beta)^{-1} = \begin{pmatrix} (1+\xi)^2 & \beta(1+\xi) \\ \beta(1+\xi) & 2\beta^2(1+\xi) \end{pmatrix}.$$

Compare the left upper entry of $I_{\text{GPar}}(\xi, \beta)^{-1}$ with J_ξ where also only one SOLE is observed, i. e. $\mu \sum_{j=1}^m l_j = 1$, by determining the quotient

$$H_\xi := \frac{J_\xi \Big|_{\mu \sum_{j=1}^m l_j = 1}}{(1+\xi)^2}.$$

The same can be done with the lower right entry of $I_{\text{GPar}}(\xi, \beta)^{-1}$ and J_β,

$$H_\beta := \frac{J_\beta \Big|_{\mu \sum_{j=1}^m l_j = 1}}{2\beta^2(1+\xi)}.$$

Figure 5.3 shows this quotients H_ξ and H_β where J_ξ and J_β are calculated based on the optimal class limit configurations from Section 5.5.4. Since the

Figure 5.3.: *Quotients H_ξ and H_β of censored and uncensored inverse Fisher information based on the optimal class limit configuration.*

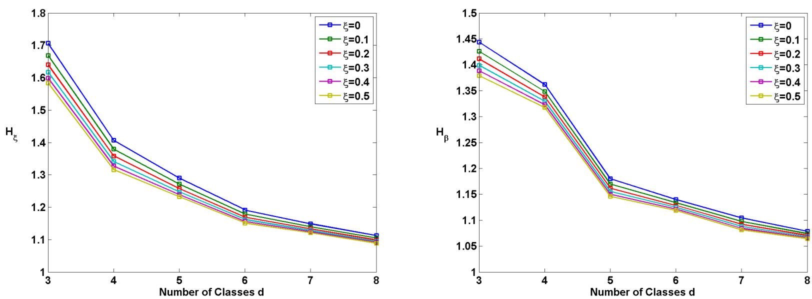
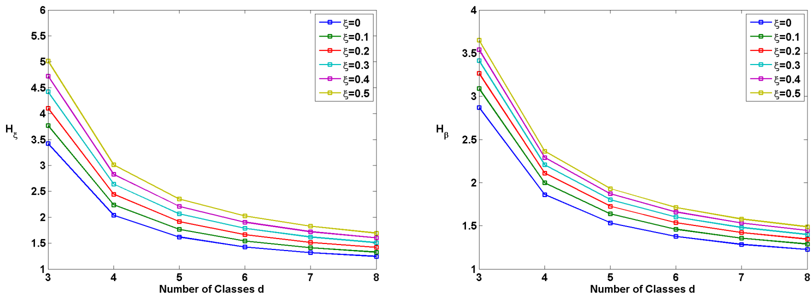


Figure 5.4.: *Quotients H_ξ and H_β of censored and uncensored inverse Fisher information based on the optimal equidistant class limit configuration.*



optimal class limits are in linear proportion to β and the relations (5.1) on page 138 hold, both H_ξ and H_β are constant as function with respect to β as long as the respective optimal class limit configurations are used. The plots show that both H_ξ and H_β approaches 1 if the number of classes decreases. The larger ξ, the closer are H_ξ and H_β to 1.

Figure 5.4 shows the same plots, but this time J_ξ and J_β are calculated based on the optimal equidistant class limit configurations from Section 5.5.4. Again, the values of H_ξ and H_β do not depend on β then. It can be seen that H_ξ and H_β are much larger in the equidistant case. However, they approach 1 for large numbers of classes. This time, both H_ξ and H_β are closer to 1 if ξ is small.

5.6. Calculation and Accuracy of the Maximum Likelihood Estimator of (ξ, β) in the Counting-Maximum Model

5.6.1. Calculation

The calculation of $\hat{\xi}_m$ and $\hat{\beta}_m$ in the counting-maximum model works very similar to the calculation in the counting model (see Section 5.5.1). In the functions Φ_i, the counts $(z_{jk})_{\substack{1\leq j\leq m\\1\leq k\leq d}}$ and the relative class limits $(s_{jk})_{\substack{1\leq j\leq m\\0\leq k\leq d}}$ must be replaced by the transformed counts $(\overline{z}_{jk})_{\substack{1\leq j\leq m\\1\leq k\leq d+1}}$ and the transformed relative class limits $(\overline{s}_{jk})_{\substack{1\leq j\leq m\\0\leq k\leq k_j+1}}$, respectively, as described in Section 4.5. In addition, the particular derivatives with respect to ξ and β of the term $-\sum_{j=1}^{m}\log(\beta+\xi(x_j-u_{\text{sev}}))$

must be added, where $(x_j)_{1 \leq j \leq m}$ are the observed maximum SOLEs per vehicle. In other words:

$$\Phi_i(\xi,\beta)\Big|_{(z_{jk}),(s_{jk})} \longrightarrow \Phi_i(\xi,\beta)\Big|_{(\overline{z}_{jk}),(\overline{s}_{jk})} - \sum_{j=1}^{m} \frac{(x_j - u_{\mathrm{sev}})\mathbb{1}_{\{1\}}(i) + \mathbb{1}_{\{2\}}(i)}{\beta + \xi(x_j - u_{\mathrm{sev}})},$$

$$\frac{\partial \Phi_2}{\partial \beta}(\xi,\beta)\Big|_{(z_{jk}),(s_{jk})} \longrightarrow \frac{\partial \Phi_2}{\partial \beta}(\xi,\beta)\Big|_{(\overline{z}_{jk}),(\overline{s}_{jk})} + \sum_{j=1}^{m} \frac{1}{(\beta + \xi(x_j - u_{\mathrm{sev}}))^2}.$$

Lemma 4.5.1 ensures that with these transformations the calculation of the maximum likelihood estimators works as described in Section 5.5.1.

5.6.2. Accuracy

Section 5.5.2 describes a Monte Carlo simulation whose purpose is to determine the accuracy of $\hat{\xi}_m$ and $\hat{\beta}_m$ in the counting model. The same simulation was also run for the counting-maximum model. The mileages were taken over from the counting model study and the maxima were drawn as described in Section 5.1.4. Since the calculation of $\hat{\xi}_m$ and $\hat{\beta}_m$ is much more time-consuming than in the counting model, only 10^4 realizations of the maximum likelihood estimators were generated. The sample means and sample standard deviations of these 10^4 values are listed in Table B.25, Table B.26 and Table B.27 and Table B.28 in the appendix.

Table B.25 shows that the standard deviation of $\hat{\xi}_m$ is for the setting with four classes by a factor of between 1.2 and 2.5 smaller than in the counting model (cf. Section 5.5.2 and Section 5.5.3). For the setting with six classes (see Table B.26), this factor can still be up to 2. The larger β, the greater is this factor, at least for the applied class limit configurations. In case of $\xi = 0$, the smaller standard deviation causes a smaller bias than in the counting model. For $\xi \in \{0.5, 1\}$, the bias is similar to that in the counting model, but this time the bias is negative. Thus, $\hat{\xi}_m$ slightly underestimates the true shape if only less observations are available. In contrast, in the counting model $\hat{\xi}_m$ overestimates the true shape.

The standard deviation of $\hat{\beta}_m$ is smaller than in the counting model, too. In case of $\xi = 0$, this also results in a smaller bias. For $\xi \in \{0.5, 1\}$ the bias of $\hat{\beta}_m$ is greater than in the counting model. However, the mean squared error of $\hat{\beta}_m$ (variance of $\hat{\beta}_m$ plus the bias of $\hat{\beta}_m$ squared [UC11]) is significantly smaller than in the counting model.

To conclude, the estimate of ξ and β through $\hat{\xi}_m$ and $\hat{\beta}_m$ respectively is significant better if the absolute maximum SOLE is part of the observation. A second advantage is that the maximum likelihood estimators always exist in the counting-maximum model.

5.7. Results of Measurement Study

The study of the *BMW Group* mentioned in Section 2.1 is the motivation for this thesis. A part of the real data from this study shall be analyzed here by means of the presented model. This part of data comprises $m = 8913$ vehicles. All the observations are based on the same class limit configuration: $d = 8$ classes, equidistant class limits with class length $\Lambda \in \mathbb{R}_{>0}$ and severity threshold $u_{\text{sev}} \in \mathbb{R}_{>0}$, so that the class limits are

$$\big(t_{j0}, t_{j1}, \ldots, t_{j7}, t_{j8}\big) = \big(u_{\text{sev}},\ u_{\text{sev}} + \Lambda,\ \ldots,\ u_{\text{sev}} + 7\Lambda,\ \infty\big) \qquad \forall j \in \mathbb{N}_{\leq 8913}.$$

As usual, for vehicle j, l_j denotes the mileage in kilometers, z_{jk} denotes the number of events in the kth class, and x_j is the maximum SOLE ($j \in \mathbb{N}_{\leq 8913}$, $k \in \mathbb{N}_{\leq 8}$).

The mileages of all of the 8913 vehicles add up to

$$l := \sum_{j=1}^{8913} l_j = 97\,385\,008$$

kilometers. A total of

$$n := \sum_{j=1}^{8913} \sum_{k=1}^{8} z_{jk} = 277\,938$$

SOLEs were generated from these vehicles. Hence, due to Theorem 4.3.2, the maximum likelihood estimator of μ takes the value

$$\hat{\mu}_m = \frac{n}{l} = \frac{277\,938}{97\,385\,008} \approx 2.854 \cdot 10^{-3}.$$

The symbol "$\approx$" means that the values are rounded. This symbol is used in the same way throughout this section.

The value of the test statistic $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ (see Equation (3.6) on page 40) is

$$\sqrt{\tfrac{m}{2}}\big(\hat{D}_2 - 1\big) = \sqrt{\tfrac{8913}{2}}\big(39.8128\ldots - 1\big) > 2591.$$

Since the 99.999 % quantile of the standard normal distribution is smaller than 5, the hypothesis that the number of SOLEs per kilometer is Poisson distributed can be rejected with all common significance levels (see hypothesis test in Section 3.5.3). Therefore, according to the decision-making procedure in Section 3.5.7, the number of SOLEs is assumed to be negative binomially distributed. The value of the maximum likelihood estimator of the exponent ϱ is

$$\hat{\varrho}_m \approx 9.538 \cdot 10^{-5}.$$

The 277 938 observed SOLEs are allocated to the eight classes as follows:

$$\left(\sum_{j=1}^{m} z_{j1}, \ldots, \sum_{j=1}^{m} z_{j8}\right) = \big(267\,510,\ 10\,217,\ 206,\ 4,\ 1,\ 0,\ 0,\ 0\big).$$

The matrix product in Corollary 4.4.7,

$$\begin{pmatrix} 267510 & 10217 & 206 & 4 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 & 6 & 12 & 20 & 30 & 42 \\ -2 & -2 & 0 & 4 & 10 & 18 & 28 \\ -4 & -6 & -6 & -4 & 0 & 6 & 14 \\ -6 & -10 & -12 & -12 & -10 & -6 & 0 \\ -8 & -14 & -18 & -20 & -20 & -18 & -14 \\ -10 & -18 & -24 & -28 & -30 & -30 & -28 \\ -12 & -22 & -30 & -36 & -40 & -42 & -42 \end{pmatrix} \begin{pmatrix} 10217 \\ 206 \\ 4 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

results in the value $-102\,115\,826$. Hence, due to this corollary and Lemma 4.4.6, the maximum likelihood estimators of shape ξ and scale β are

$$\hat{\xi}_m = 0 \quad \text{and} \quad \hat{\beta}_m = \frac{\Lambda}{\log\left(1 + \frac{277\,938}{10\,645}\right)} \approx 3.030\,\Lambda \cdot 10^{-1} \qquad \textit{(counting model).}$$

The Fisher information matrix from Theorem 4.2.2 evaluated at the maximum likelihood estimators amounts to

$$I_C\left(\hat{\varrho}_m, \hat{\mu}_m, \hat{\xi}_m, \hat{\beta}_m\right) \approx \begin{pmatrix} 4.478 \cdot 10^{11} & 0 & 0 & 0 \\ 0 & 1.104 \cdot 10^{9} & 0 & 0 \\ 0 & 0 & 4.319 \cdot 10^{5} & 7.055\,\Lambda^{-1} \cdot 10^{5} \\ 0 & 0 & 7.055\,\Lambda^{-1} \cdot 10^{5} & 1.311\,\Lambda^{-2} \cdot 10^{6} \end{pmatrix}.$$

Since the rounded amount of the 97.5 % quantile of the standard normal distribution is 1.960, the 95 % confidence intervals of μ (see Section 4.3.1), ρ (see Section 4.3.2), ξ and β (see Section 4.4.3 for the case $\hat{\xi}_m = 0$) are

$$\begin{aligned} C_\mu\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[2.795,\ 2.913\big] \cdot 10^{-3}, \\ C_\varrho\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[9.245,\ 9.831\big] \cdot 10^{-5}, \\ C_\xi^0\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[0,\ 7.205\big] \cdot 10^{-3}, \\ C_\beta^0\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[2.981\,\Lambda,\ 3.045\,\Lambda\big] \cdot 10^{-1}. \end{aligned}$$

It is striking that the three upper classes are empty. It can be expected that the accuracy of estimate of ξ and β will be better for smaller classes. An optimal class length and, furthermore, an optimal class limit configuration can be found as described in Section 5.5.4. In this example, the optimal (equidistant) class limit configuration is $\big(u_{\text{sev}},\ u_{\text{sev}} + s_{\text{opt},1},\ \ldots,\ u_{\text{sev}} + s_{\text{opt},7},\ \infty\big)$ with

$$\begin{aligned} &\big(s_{\text{opt},1}, \ldots, s_{\text{opt},7}\big) \\ &\approx \begin{cases} \big(0.115\,\Lambda,\ 0.267\,\Lambda,\ 0.500\,\Lambda,\ 0.895\,\Lambda,\ 1.292\,\Lambda,\ 1.741\,\Lambda,\ 2.358\,\Lambda\big) & \textit{(optimal)} \\ \big(0.270\,\Lambda,\ 0.540\,\Lambda,\ 0.810\,\Lambda,\ 1.080\,\Lambda,\ 1.350\,\Lambda,\ 1.620\,\Lambda,\ 1.890\,\Lambda\big) & \textit{(opt. equid.).} \end{cases} \end{aligned}$$

With the optimal equidistant class limit configuration only 22 721 331 kilometers are sufficient to get the same confidence interval C^0_ξ as above. With the optimal class limit configuration only 20 315 151 kilometers are needed. This is just about 23 % and 21 % respectively of the actual observation period $l = 97\,385\,008$ kilometers. In case of the scale parameter β, the optimal equidistant class limit configuration and an observation period of 18 209 398 kilometers yield the confidence interval $\bigl[2.989\,\Lambda\cdot 10^{-1},\ 3.053\,\Lambda\cdot 10^{-1}\bigr]$, which is as small as C^0_β from above. The optimal class limit configuration and the observation period 16 743 319 kilometers lead to a confidence interval of the same size, namely $\bigl[2.990\,\Lambda\cdot 10^{-1},\ 3.054\,\Lambda\cdot 10^{-1}\bigr]$.

In the counting-maximum model, the numerical calculation (see Section 5.6) yields

$$\hat{\xi}_m \approx 3.086\cdot 10^{-3} \quad\text{and}\quad \hat{\beta}_m \approx 2.940\,\Lambda\cdot 10^{-1} \qquad \textit{(counting-maximum model)}.$$

The observed Fisher information matrix from Section 4.5 adds up to

$$\mathcal{I}_{\text{sev}}\bigl((z_{jk})_{jk},(x_j)_j\bigr) \approx \begin{pmatrix} 4.681\cdot 10^5 & 7.806\,\Lambda^{-1}\cdot 10^5 \\ 7.806\,\Lambda^{-1}\cdot 10^5 & 1.541\,\Lambda^{-2}\cdot 10^6 \end{pmatrix},$$

which leads to the 95 % confidence intervals (see Equation (4.11) on page 115)

$$C_\xi\bigl(0.05,(z_{jk})_{j,k},(x_j)_j\bigr) \approx \bigl[-4.189,\ 10.361\bigr]\cdot 10^{-3},$$
$$C_\beta\bigl(0.05,(z_{jk})_{j,k},(x_j)_j\bigr) \approx \bigl[2.900\,\Lambda,\ 2.980\,\Lambda\bigr]\cdot 10^{-1}.$$

From the Fisher information matrices one can see that the information content in the counting-maximum model is higher than in the counting model as expected. In this example, the confidence intervals in the counting model are smaller only because the shape is estimated to be 0 there, and the intervals C^0_ξ, C^0_β are in general smaller than the intervals C_ξ, C_β (see Section 4.4.3).

Based on the estimated parameter values, the distribution of the maximum SOLE during a reference distance can be calculated. If, for instance, the reference distance is 100 000 kilometers, the cumulative distribution function of $M^{*10^5}_{\text{sev}}$ is in the counting-model

$$\mathbb{P}\Bigl(M^{*10^5}_{\text{sev}} \le t\Bigr) = \left(\frac{\hat{\varrho}_m}{\hat{\varrho}_m + \hat{\mu}_m \exp\Bigl(\frac{1}{\hat{\beta}_m}(t-u_{\text{sev}})\mathbb{1}_{\mathbb{R}_{\ge u_{\text{sev}}}}(t)\Bigr)}\right)^{10^5\,\hat{\varrho}_m} \qquad \forall t\in\mathbb{R}_{\ge 0}$$

(since $\hat{\xi}_m = 0$, see Example 3.4.3), and in the counting-maximum model it is

$$\mathbb{P}\Bigl(M^{*10^5}_{\text{sev}} \le t\Bigr) = \left(\frac{\hat{\varrho}_m}{\hat{\varrho}_m + \hat{\mu}_m \Bigl(1+\frac{\hat{\xi}_m}{\hat{\beta}_m}(t-u_{\text{sev}})\mathbb{1}_{\mathbb{R}_{\ge u_{\text{sev}}}}(t)\Bigr)^{-\frac{1}{\hat{\xi}_m}}}\right)^{10^5\,\hat{\varrho}_m} \qquad \forall t\in\mathbb{R}_{\ge 0}.$$

Figure 5.5.: *Probability density function and cumulative distribution function of* $M_{\text{sev}}^{*10^5}$.

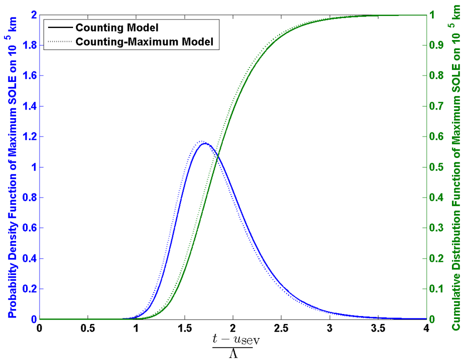


Figure 5.5 plots this distribution functions and the corresponding probability density functions. For example, the values of the 99.9 % quantiles of $M_{\text{sev}}^{*10^5}$ are $u_{\text{sev}}+3.807\,\Lambda$ in the counting model and $u_{\text{sev}}+3.765\,\Lambda$ in the counting maximum-model. Only one vehicle in a thousand will be subjected to higher loads during a distance of 100 000 kilometers.

Since the data are based on various mileages, the adapted model and the data concerning the number of SOLEs cannot be compared in a histogram without further ado. Therefore, for any vehicle the number of observed SOLEs is divided by the mileage, and this number of SOLEs per kilometer is illustrated in a histogram (see Figure 5.6). To compare the resultant values with the adapted model, for each vehicle j a realization of $N_{\text{num}}^{*l_j} \sim \text{NBin}(\hat{\varrho}_m l_j, \hat{\mu}_m l_j)$ is drawn, and, similar to the real data, this realization is divided by the mileage l_j. In this way a total of 10^5 histograms are generated. The averaged histogram is illustrated in Figure 5.6 under the heading “Negative Binomial Fit”. For comparison, a histogram based on the Poisson model is plotted in the same way under the heading “Poisson Fit” in Figure 5.6.

The plot shows that the negative binomial model corresponds quite well to the data. However, within a range close to 0 the distributions vary from each other. One reason for this might be that the sample of the 8913 vehicles is not homogeneous enough. The occurrence rate of SOLEs might be influenced by several factors like country or range of models. These external influences might be found by dint of a factor analysis or analysis of variance. Afterwards, for each subsample an own negative binomial distribution can be fitted.

A second reason for the difference between data and model might be that u_{sev} is set too small. If so, the lowest class is not only filled with SOLEs but with operating load events, too. Since some assumptions about SOLEs are not

Figure 5.6.: *Histogram of the quotient of number of observed SOLEs and mileage for all vehicles (blue) in comparison with the theoretical distribution of this quotient based on the negative binomial model (green) and based on the Poisson model (red).*

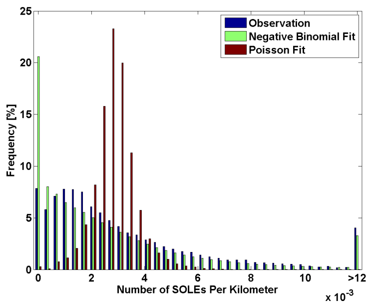
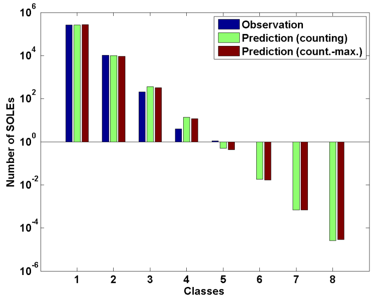


Figure 5.7.: *Observed total number of SOLEs per class (blue) in comparison with the expected number of SOLEs according to the model with parameters estimated in the counting model (green) and in the counting-maximum model (red).*



necessarily right for operating loads, especially Assumption 3.1.3, this might lead to a mixture distribution. If this is the case, disregard the lowest class and use t_{j1} as severity threshold instead (of course, this is only possible if $t_{11} = \ldots = t_{m1}$).

The goodness of fit concerning the severity is visualized in a bar plot (see Figure 5.7). The blue bars represent the total number of observed SOLEs per class. The green and red bars represent the expected number of SOLEs during the observation period $l = 97\,385\,008$ kilometers under the presented model,

$$\mathbb{E}\Big[Z_{l,(t_{j,k-1},t_{jk}]}\Big] = \begin{cases} \hat{\mu}_m l \left(\left(1+\frac{\hat{\xi}_m}{\hat{\beta}_m}(k-1)\Lambda\right)^{-\frac{1}{\hat{\xi}_m}} - \left(1+\frac{\hat{\xi}_m}{\hat{\beta}_m}k\Lambda\right)^{-\frac{1}{\hat{\xi}_m}} \mathbb{1}_{\mathbb{N}_{\le 7}}(k)\right), & \text{if } \hat{\xi}_m > 0, \\ \hat{\mu}_m l \left(\mathrm{e}^{-\frac{1}{\hat{\beta}_m}(k-1)\Lambda} - \mathrm{e}^{-\frac{1}{\hat{\beta}_m}k\Lambda}\mathbb{1}_{\mathbb{N}_{\le 7}}(k)\right), & \text{if } \hat{\xi}_m = 0. \end{cases}$$

(see Proposition 3.2.1). The numerical calculation yields

$$\left(\mathbb{E}\big[Z_{l,(t_{10},t_{11}]}\big], \ldots, \mathbb{E}\Big[Z_{l,(t_{1,d-1},t_{1d}]}\Big]\right) \approx \begin{cases} (267\,685.7,\ 9\,874.2,\ 364.2,\ 13.4,\ 0.50,\ 0.018,\ 0.00067,\ 0.000026) & \textit{(count. model)} \\ (268\,513.1,\ 9\,094.0,\ 318.9,\ 11.6,\ 0.43,\ 0.017,\ 0.00068,\ 0.000029) & \textit{(count.-max. mod.).} \end{cases}$$

On the whole, the model is in line with the data. However, the comparison to the observation per class, which is $(267\,510,\ 10\,217,\ 206,\ 4,\ 1,\ 0,\ 0,\ 0)$, illustrates that in the third and in the fourth class the model predicts too much events. A reason for that might be found in the fact that the ratio of the number of SOLEs in the first class to the number of SOLEs in the third or fourth class is extremely high, e. g. the ratio between first and third class is $^{267510}/_{206} > 1298$. Thus, the lowest classes strongly influence the estimates of ξ and β. Consequently, if the severity threshold u_{sev} is set too small such that the approximation according to the Pickands–Balkema–de Haan Theorem (see Theorem 2.4.4) is not suitable, then the influential first class may distort the estimate of the actual right tail of the distribution. This problem can be solved by disregarding the lowest class and using t_{j1} as severity threshold, provided that $t_{11} = \ldots = t_{m1}$.

In the case of the distributions of both the number of SOLEs and the severity of a SOLE it is proposed to disregard the lowest class $(u_{\text{sev}}, \Lambda]$ in order to achieve a better goodness of fit. When this is done, $t_{11} = u_{\text{sev}} + \Lambda$ is the new severity threshold. This time, the value of the test statistic $\sqrt{\frac{m}{2}}\big(\hat{D}_2 - 1\big)$ is

$$\sqrt{\tfrac{m}{2}}\big(\hat{D}_2 - 1\big) = \sqrt{\tfrac{8913}{2}}\big(23.1219\ldots - 1\big) > 1476.$$

The values of the maximum likelihood estimators are

$$\begin{aligned}
\hat{\mu}_m &\approx 1.071 \cdot 10^{-4}, \\
\hat{\varrho}_m &\approx 3.096 \cdot 10^{-5}, \\
\hat{\xi}_m &\approx 2.761 \cdot 10^{-2} && \textit{(counting model)}, \\
\hat{\beta}_m &\approx 2.427\,\Lambda \cdot 10^{-1} && \textit{(counting model)}, \\
\hat{\xi}_m &\approx 0.642 \cdot 10^{-2} && \textit{(count.-max. model)}, \\
\hat{\beta}_m &\approx 2.563\,\Lambda \cdot 10^{-1} && \textit{(count.-max. model)}.
\end{aligned}$$

and their 95 % confidence intervals are

$$\begin{aligned}
C_\mu\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[1.027,\ 1.114\big] \cdot 10^{-4}, \\
C_\varrho\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[2.916,\ 3.275\big] \cdot 10^{-5}, \\
C_\xi\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[-3.172,\ 8.694\big] \cdot 10^{-2} && \textit{(counting model)}, \\
C_\beta\big(0.05, (z_{jk})_{j,k}\big) &\approx \big[2.116\,\Lambda,\ 2.737\,\Lambda\big] \cdot 10^{-1} && \textit{(counting model)}, \\
C_\xi\big(0.05, (z_{jk})_{j,k}, (x_j)_j\big) &\approx \big[-1.579,\ 2.863\big] \cdot 10^{-2} && \textit{(count.-max. model)}, \\
C_\beta\big(0.05, (z_{jk})_{j,k}, (x_j)_j\big) &\approx \big[2.469\,\Lambda,\ 2.656\,\Lambda\big] \cdot 10^{-1} && \textit{(count.-max. model)}.
\end{aligned}$$

The 99.9 % quantiles of $M_{\text{sev}}^{*10^5}$ are a little bit smaller than before: in the counting model $u_{\text{sev}} + 3.566\,\Lambda$ and in the counting-maximum model $u_{\text{sev}} + 3.450\,\Lambda$. The whole distribution of $M_{\text{sev}}^{*10^5}$ is plotted in Figure 5.8. Take into consideration that the abscissa is shifted by one compared with Figure 5.5.

Figure 5.8.: *Probability density function and cumulative distribution function of $M_{\text{sev}}^{*10^5}$; based on data with class 2 as lowest class.*

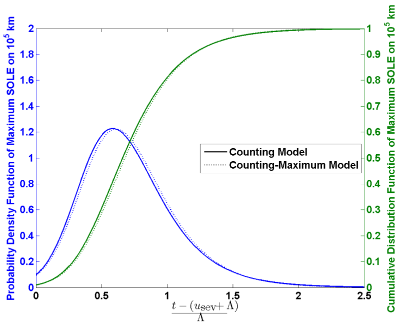
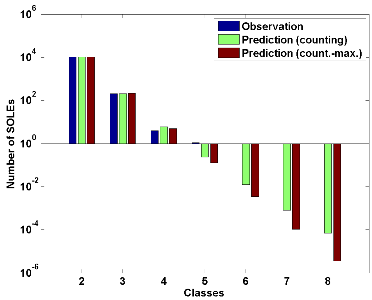


Figure 5.9.: *Observed total number of SOLEs per class (blue) in comparison with the expected number of SOLEs according to the model with parameters estimated in the counting model (green) and in the counting-maximum model (red); based on data with class 2 as lowest class.*



This time, the expected numbers of SOLEs per class are

$$\left(\mathbb{E}\left[Z_{l,(t_{11},t_{12}]}\right], \ldots, \mathbb{E}\left[Z_{l,(t_{1,d-1},t_{1d}]}\right]\right)$$
$$\approx \begin{cases} (10\,217.5,\ 204.3,\ 6.0,\ 0.24,\ 0.012,\ \ 0.00079,\ 0.000068) & \textit{(counting model)} \\ (10\,207.0,\ 215.9,\ 5.0,\ 0.13,\ 0.0035,\ 0.00010,\ 0.0000035) & \textit{(count.-max. mod.)} \end{cases}$$

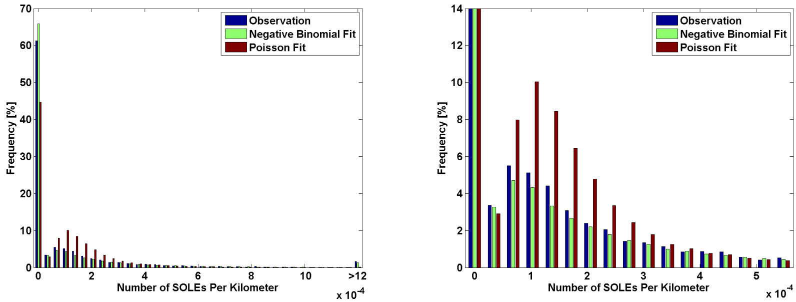
which fits in very well with the data (see Figure 5.9). This can also be verified quantitatively by dint of either Person's chi-squared goodness-of-fit test or the similar likelihood-ratio goodness-of-fit test [UC11, HCBNM02, pp. 14–15]. Pearson's test and the likelihood-ratio test use the test statistics χ^2 and G^2 respectively with

$$\chi^2 = \sum_{k=2}^{8} \frac{\left(\sum_{j=1}^{8913} z_{jk} - \mathbb{E}\left[Z_{l,(t_{1,k-1},t_{1k}]}\right]\right)^2}{\mathbb{E}\left[Z_{l,(t_{1,k-1},t_{1k}]}\right]} \approx 3.121,$$

$$G^2 = 2\sum_{k=2}^{5}\sum_{j=1}^{8913} z_{jk} \log\left(\frac{\sum_{j=1}^{8913} z_{jk}}{\mathbb{E}\left[Z_{l,(t_{1,k-1},t_{1k}]}\right]}\right) \approx 2.122\,.$$

According to theory [HCBNM02, pp. 14–15], under the null hypothesis (i. e. data follow the distribution provided by the model) both χ^2 and G^2 are approximately chi-squared distributed with 4 degrees of freedom (seven classes minus one minus two estimated parameters ξ, β; note that $\hat{\mu}_m l = \sum_{j=1}^{8913}\sum_{k=2}^{8} z_{jk}$ is the total number of events), which corresponds to the distribution of the sum of 4 squared standard normal variates [UC11]. The 95 % quantile of a chi-squared

Figure 5.10.: *Histogram of the quotient of number of observed SOLEs and mileage for all vehicles (blue) in comparison with the theoretical distribution of this ratio based on the negative binomial model (green) and based on the Poisson model (red); based on data with class 2 as lowest class.*



distribution with 4 degrees of freedom approximately is 9.488. Therefore, both tests do not reject the null hypothesis. For the initial estimation above where also the first class $(u_{\text{sev}}, u_{\text{sev}} + \Lambda]$ is taken into account, the test statistics would be $\chi^2 \approx 87.916$ and $G^2 \approx 103.152$. In that situation, both tests would reject the null hypothesis that the data correspond to the model.

The adapted model concerning the number of SOLEs approximates the data very well, too, as can be seen in Figure 5.10. The figure is generated in the same way as Figure 5.6 above. Even though the goodness-of-fit tests from above would reject the null hypothesis that data and model fit together (both test statistics χ^2 and G^2 take values larger than 100), it is nevertheless not advisable to reject the negative binomial distribution, because the courses of the histograms of observation and negative binomial fit are very similar to each other. On the contrary, the distribution of the number of observed SOLEs may be a mixed distribution consisting of several negative binomial distributions since the sample is not homogeneous enough.

6. Résumé

Section 5.7 in the previous chapter illustrates two things: firstly, it shows that the model presented in Chapter 3 is together with the parameter estimation procedure from Chapter 4 suitable for processing the available data and for answering the questions from Chapter 2. Secondly, it describes a workflow for the analysis of the present data:

1. Calculate the maximum likelihood estimator $\hat{\mu}_m$ of the average number μ of SOLEs during one kilometer (see Theorem 4.3.2) and the corresponding actual confidence interval $C_\mu(\alpha, z)$ (see Equation (4.3) on page 77).

2. Utilize the hypothesis test in Section 3.5.3 in order to check whether the number of SOLEs during one kilometer is statistically dispersed or not. Choose the binomial (Bernoulli), Poisson or the negative binomial distribution by following the rules of Section 3.5.7.

3. If step 2 suggests a negative binomial distribution for the number of SOLEs per kilometer, calculate the maximum likelihood estimator $\hat{\varrho}_m$ of the exponent ϱ (see Theorem 4.3.7, Section 5.3.1) and the corresponding actual confidence interval $C_\varrho(\alpha, z)$ (see Equation (4.5) on page 85).

4. Check whether the maximum likelihood estimator $\hat{\xi}_m$ of the shape ξ of the severity of any SOLE is equal to 0 or positive. If the counting model is chosen (see Section 4.2.1) and all class limits are equidistant with the same class length, use Corollary 4.4.7 for that. Otherwise, follow Algorithm 3 on page 134, lines 1-4, with Φ_1, Φ_2 either from Section 5.5.1 or from Section 5.6.1. If $\hat{\xi}_m = 0$, Corollary 4.4.7 and Algorithm 3 respectively guide how to calculate the maximum likelihood estimator $\hat{\beta}_m$ of the scale β of the severity of a SOLE. In addition, calculate the actual confidence intervals $C^0_\xi(\alpha, z)$ and $C^0_\beta(\alpha, z)$ (see Equation (4.9) on page 103) or $C^0_\xi(\alpha, z, x)$ and $C^0_\beta(\alpha, z, x)$ (see Equation (4.12) on page 115).

5. If step 4 suggests $\hat{\xi}_m \in \mathbb{R}_{>0}$, utilize Algorithm 3 on page 134 with Φ_1, Φ_2 from Section 5.5.1 or from Section 5.6.1 in order to calculate the maximum likelihood estimators $\hat{\xi}_m$ and $\hat{\beta}_m$ of shape ξ and scale β of the severity of a SOLE, respectively. Furthermore, calculate the actual confidence intervals $C_\xi(\alpha, z)$ and $C_\beta(\alpha, z)$ (see Equation (4.7) on page 101) or $C_\xi(\alpha, z, x)$ and $C_\beta(\alpha, z, x)$ (see Equation (4.11) on page 115).

6. Prepare the goodness-of-fit plots described in Section 5.7 (see Figure 5.6, Figure 5.7, Figure 5.9 and Figure 5.10) in order to check whether the severity threshold u_{sev} is set large enough. The goodness of fit can also be verified via Pearson's chi-squared or the likelihood-ratio goodness-of-fit test as described in Section 5.7.

7. If needed, the arrangement of the class limits can be optimized (see Section 4.4.4, Section 5.5.4), and the necessary observation period can be estimated (see Example 4.3.6, Example 4.3.8, Section 5.5.4).

8. Use the maximum likelihood estimators to determine the distribution of the number of SOLEs in any measurable set A during any mileage l, $Z_{l,A}$ (see Proposition 3.2.1), the distribution of the maximum SOLE during l kilometers, M^{*l}_{sev} (see Proposition 3.4.2), and the respective common distributions (see Theorem 3.2.3, Theorem 3.4.4).

The particular sections in Chapter 5 clearly demonstrate by examination of typical examples that the maximum likelihood method yields adequate estimates of the distribution parameters. Furthermore, the Poisson hypothesis test developed in Section 3.5 is very accurate and suitable for testing the index of dispersion of the number of SOLEs during one kilometer.

If the goodness-of-fit plots and tests suggest a bad fit to the data, an initial improvement can be achieved by disregarding the lowest class(es). Section 5.7 argues that some model assumptions cannot be correct if the severity threshold is too low, e. g. the generalized Pareto approximation. The data analysis in Section 5.7 shows that the adjustment of the model to the data becomes much better after disregarding the lowest class.

If this procedure does not help, the data are possibly not homogeneous, i. e. the available sample is a mixed population. Some kind of analysis of variance can help to find the particular subpopulations. For this purpose, define influential factors and divide the whole population into cells where each cell represents one combination of levels of the factors. In each cell, calculate the maximum likelihood estimators of the distribution parameters. The vector of the three or four estimators forms the "observation" of the particular cell. In order to point that the estimators of different cells do not have the same accuracy level since the sample sizes, the numbers of observed SOLEs and the mileages are different, the estimators can be weighted by the Fisher information. An analysis of variance can specify if a factor influences the value of one or more estimators. A realization of this idea is in progress.

A. Lemmata

A.1 Lemma. *Let be* $a \in \mathbb{R}_{>0}$ *and*

$$f\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0} \to (0,1)\colon\ (x_1, x_2) \longmapsto \begin{cases} \left(1 + \frac{x_1}{x_2} a\right)^{-\frac{1}{x_1}}, & \text{if } x_1 > 0, \\ \mathrm{e}^{-\frac{1}{x_2} a}, & \text{if } x_1 = 0. \end{cases}$$

1. *f is continuous.*
2. *f is continuously differentiable*[1] *and the partial derivatives are ($i \in \{1,2\}$)*

$$\tfrac{\partial f}{\partial x_i}\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0} \to \mathbb{R}\colon\ (x_1, x_2) \longmapsto f(x_1, x_2)\ \frac{a}{{x_2}^2 + x_1 x_2 a}\ \varphi_i\!\left(\tfrac{x_1}{x_2}, a\right),$$

   where

$$\varphi_i(x, a) := \mathbb{1}_{\{2\}}(i) + \mathbb{1}_{\{1\}}(i) \cdot \begin{cases} \frac{1}{x}\left(\log(1 + xa)\left(1 + \frac{1}{xa}\right) - 1\right), & \text{if } xa > 0, \\ \frac{a}{2}, & \text{if } xa = 0. \end{cases}$$

3. *$\frac{\partial f}{\partial x_i}$ is continuously differentiable*[1] *and the partial derivatives are*

$$\begin{aligned} \tfrac{\partial^2 f}{\partial x_j \partial x_i}&\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0} \to \mathbb{R} \\ &(x_1, x_2) \longmapsto f(x_1, x_2)\left(\frac{a}{{x_2}^2 + x_1 x_2 a}\right)^2 \phi_{ji}(x_1, x_2, a) \end{aligned}$$

   ($i, j \in \{1,2\}$), where

$$\begin{aligned} \phi_{ji}(x_1, x_2, a) := \varphi_i\!\left(\tfrac{x_1}{x_2}, a\right) \Bigg(&\varphi_j\!\left(\tfrac{x_1}{x_2}, a\right) - x_2\, \mathbb{1}_{\{1\}}(j) - \left(2\,\frac{x_2}{a} + x_1\right) \mathbb{1}_{\{2\}}(j) \\ &- \left(\frac{x_2}{a} + x_1\right) \mathbb{1}_{\{1\}}(i)\ \tilde{\varphi}_j\!\left(\tfrac{x_1}{x_2}, a\right)\Bigg), \end{aligned}$$

   φ_i is defined as in the second statement, and

$$\tilde{\varphi}_j(x, a) := \begin{cases} \frac{\log(1+xa)\left(1+\frac{2}{xa}\right)-2}{\log(1+xa)\left(1+\frac{1}{xa}\right)-1}\left(\frac{\mathbb{1}_{\{1\}}(j)}{x} - \mathbb{1}_{\{2\}}(j)\right), & \text{if } xa > 0, \\ \frac{a}{3}\,\mathbb{1}_{\{1\}}(j), & \text{if } xa = 0. \end{cases}$$

[1] at $x_1 = 0$ this means the right derivative

4. *For any $x \in \mathbb{R}_{\geq 0}$ and $y \in \mathbb{R}_{>0}$ it holds*

$$\lim_{\substack{x_1 \to x \\ x_2 \to 0}} f(x_1, x_2) = 0, \qquad \lim_{\substack{x_1 \to x \\ x_2 \to \infty}} f(x_1, x_2) = 1,$$

$$\lim_{\substack{x_1 \to \infty \\ x_2 \to y}} f(x_1, x_2) = 1, \qquad \lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} f(x_1, x_2) = 1.$$

5. *Suppose, $(x_{1,n}, x_{2,n})_{n \in \mathbb{N}} \subseteq \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ is a sequence with $x_{1,n} \to \infty$ and $x_{2,n} \to 0$ for $n \to \infty$ such that the limit*

$$b := \lim_{n \to \infty} x_{2,n}{}^{\frac{1}{x_{1,n}}} \in [0, 1]$$

exists, then

$$\lim_{n \to \infty} f(x_{1,n}, x_{2,n}) = b.$$

Proof. **1.:** The points $(x_1, x_2) \in \{0\} \times \mathbb{R}_{>0}$ are the only values where the continuity could be destroyed. Suppose, $(x_{1,n}, x_{2,n})_{n \in \mathbb{N}} \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ is a sequence with

$$\lim_{n \to \infty} (x_{1,n}, x_{2,n}) = (0, x_0) \in \{0\} \times \mathbb{R}_{>0}.$$

Since $x_{1,n} > 0$ for all $n \in \mathbb{N}$, it is

$$\lim_{n \to \infty} f(x_{1,n}, x_{2,n}) = \lim_{n \to \infty} \left(1 + \frac{x_{1,n}}{x_{2,n}}\, a\right)^{-\frac{1}{x_{1,n}}} = \exp\left(-\lim_{n \to \infty} \frac{\log\left(1 + \frac{x_{1,n}}{x_0}\, a\right)}{x_{1,n}}\right).$$

According to l'Hôpital's Rule [For04, p. 171] the limit within the exponential function is

$$\lim_{n \to \infty} \frac{\log\left(1 + \frac{x_{1,n}}{x_0}\, a\right)}{x_{1,n}} = \lim_{n \to \infty} \frac{\frac{1}{x_0}\, a}{1 + \frac{x_{1,n}}{x_0}\, a} = \frac{1}{x_0}\, a.$$

Hence,

$$\lim_{n \to \infty} f\left(x_{1,n}, x_{2,n}\right) = \exp\left(-\frac{1}{x_0}\, a\right) = f(0, x_0) = f\left(\lim_{n \to \infty} (x_{1,n}, x_{2,n})\right).$$

2.: One manages to calculate the derivative with respect to x_2 with help of the standard rules from differential calculus. Furthermore, $\frac{\partial f}{\partial x_2}$ is continuous because f and $\frac{a}{x_2{}^2 + x_1 x_2 a}$ are continuous.

The derivative of f with respect to x_1 can be calculated easily if $x_1 \in \mathbb{R}_{>0}$. In case of $x_1 = 0$, first, discover with help of l'Hôspital's Rule [For04, p. 171] that

$$\begin{aligned} \lim_{x \searrow 0} \varphi_1(x, a) = \lim_{x \searrow 0} \frac{\log(1 + xa)\,(xa + 1) - xa}{x^2 a} &= \lim_{x \searrow 0} \frac{a \log(1 + xa)}{2xa} \\ &= \lim_{x \searrow 0} \frac{a}{2(1 + ax)} \\ &= \frac{a}{2}, \end{aligned}$$

and so it follows

$$\lim_{x_1 \searrow 0} \tfrac{\partial f}{\partial x_1}(x_1, x_2) = f(0, x_2) \cdot \frac{1}{2} \left(\frac{a}{x_2}\right)^2 \in \mathbb{R} \qquad \forall x_2 \in \mathbb{R}_{>0}.$$

Since this limit exists, it has to be the partial derivative at $x_1 = 0$ from the right. The continuity of $\frac{\partial f}{\partial x_1}$ follows from the continuity of f and the limiting calculations above.

3.: It is easy to check that

$$\begin{aligned}&\phi_{ji}(x_1, x_2, a)\\ &\quad= \varphi_i\Big(\tfrac{x_1}{x_2}, a\Big) \left(\varphi_j\Big(\tfrac{x_1}{x_2}, a\Big) + \frac{\frac{\partial}{\partial x_j} \frac{a}{x_2{}^2 + x_1 x_2 a}}{\left(\frac{a}{x_2{}^2 + x_1 x_2 a}\right)^2} + \frac{\frac{\partial}{\partial x_j} \varphi_i\Big(\frac{x_1}{x_2}, a\Big)}{\varphi_i\Big(\frac{x_1}{x_2}, a\Big) \frac{a}{x_2{}^2 + x_1 x_2 a}} \right).\end{aligned}$$

With that, for $j = 2$, $i \in \{1, 2\}$ and $(x_1, x_2) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ as well as for $j = 1$, $i \in \{1, 2\}$ and $(x_1, x_2) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ it follows from the second statement of this lemma

$$\begin{aligned}&f(x_1, x_2) \left(\frac{a}{x_2{}^2 + x_1 x_2 a} \right)^2 \phi_{ji}(x_1, x_2, a)\\ &\quad= f(x_1, x_2)\, \frac{a\ \varphi_i\Big(\frac{x_1}{x_2}, a\Big)}{x_2{}^2 + x_1 x_2 a} \left(\frac{\frac{\partial f}{\partial x_j}(x_1, x_2)}{f(x_1, x_2)} + \frac{\frac{\partial}{\partial x_j} \frac{a}{x_2{}^2 + x_1 x_2 a}}{\frac{a}{x_2{}^2 + x_1 x_2 a}} + \frac{\frac{\partial}{\partial x_j} \varphi_i\Big(\frac{x_1}{x_2}, a\Big)}{\varphi_i\Big(\frac{x_1}{x_2}, a\Big)} \right)\\ &\quad= \frac{\partial}{\partial x_j} \left(f(x_1, x_2)\ \frac{a}{x_2{}^2 + x_1 x_2 a}\ \varphi_i\Big(\tfrac{x_1}{x_2}, a\Big) \right)\\ &\quad= \tfrac{\partial^2 f}{\partial x_j \partial x_i}(x_1, x_2).\end{aligned}$$

Furthermore, the function $\tilde{\varphi}_j(\,\cdot\,, a)$ is continuous, because on the one hand it holds due to l'Hôspital's Rule [For04, p. 171]

$$\begin{aligned}\lim_{x \searrow 0} \frac{\log(1 + xa)\,(1 + \frac{2}{xa}) - 2}{x^2} &= \lim_{x \searrow 0} \frac{\log(1 + xa)\,(xa + 2) - 2xa}{x^3 a}\\ &= \lim_{x \searrow 0} \frac{a \log(1 + xa) - \frac{xa^2}{xa+1}}{3x^2 a}\\ &= \lim_{x \searrow 0} \frac{\frac{xa^3}{(xa+1)^2}}{6xa} = \frac{a^2}{6},\end{aligned}$$

and, on the other hand, it holds

$$\lim_{x \searrow 0} \frac{\log(1 + xa)\,(1 + \frac{1}{xa}) - 1}{x} = \frac{a}{2}$$

as shown above. Hence, the limit

$$\lim_{x_1 \searrow 0} \frac{\partial^2 f}{\partial x_j \partial x_i}(x_1, x_2) = f(0, x_2)\, \frac{a^2}{{x_2}^4}\, \phi_{ji}(0, x_2, a)$$

exists and must therefore be the second partial derivative with respect to x_1 from the right at $x_1 = 0$.

4.: The first two limits are

$$\lim_{\substack{x_1 \to x \\ x_2 \to 0}} f(x_1, x_2) = \lim_{x_2 \to 0} f(x, x_2) = 0, \qquad \lim_{\substack{x_1 \to x \\ x_2 \to \infty}} f(x_1, x_2) = \lim_{x_2 \to \infty} f(x, x_2) = 1.$$

The third limit follows with help of l'Hôspital's Rule [For04, p. 171]:

$$\begin{aligned}
\lim_{\substack{x_1 \to \infty \\ x_2 \to y}} f(x_1, x_2) = \lim_{x_1 \to \infty} f(x_1, y) &= \exp\left(- \lim_{x_1 \to \infty} \frac{\log\left(1 + \frac{x_1}{y}\, a\right)}{x_1} \right) \\
&= \exp\left(- \lim_{x_1 \to \infty} \frac{\frac{1}{y}\, a}{1 + \frac{x_1}{y}\, a} \right) \\
&= 1.
\end{aligned}$$

At last, since f is monotonically increasing in x_2, it holds

$$1 \geq \lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} f(x_1, x_2) \geq \lim_{\substack{x_1 \to \infty \\ x_2 \to y}} f(x_1, x_2) = 1.$$

5.: The assumptions of the proposition yield

$$\begin{aligned}
\lim_{n \to \infty} f(x_{1,n}, x_{2,n}) &= \lim_{n \to \infty} x_{2,n}^{\frac{1}{x_{1,n}}} \exp\left(- \frac{\log\left(x_{2,n} + x_{1,n}\, a\right)}{x_{1,n}} \right) \\
&= b \exp\left(- \lim_{n \to \infty} \frac{\log\left(x_{1,n}\, a\right)}{x_{1,n}} \right) \\
&= b.
\end{aligned}$$

□

A.2 Lemma. *For any $a \in \mathbb{R}_{>0}$ define*

$$f_a \colon \mathbb{R}_{>0} \to \mathbb{R} \colon\ x \longmapsto \frac{1}{x}\left(\log(1+ax)\left(1+\frac{1}{ax}\right)-1\right),$$

then:

1. *f_a is positive.*
2. *f_a is strictly decreasing.*
3. *f_a is strictly convex.*
4. *Supremum and infimum of f_a are*

$$\sup_{x\in\mathbb{R}_{>0}} f_a(x) = \lim_{x\to 0} f_a(x) = \frac{a}{2} \qquad \text{and} \qquad \inf_{x\in\mathbb{R}_{>0}} f_a(x) = \lim_{x\to\infty} f_a(x) = 0.$$

Proof. **1.:** $f_a(x)$ is positive if and only if

$$\log(1+ax)\,(ax+1) \overset{!}{>} ax.$$

Both sides are equal to 0 if $x = 0$, and the left-hand side increases faster than the right-hand side,

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(\log(1+ax)\,(ax+1)\big) = a(1+\log(1+ax)) > a = \frac{\mathrm{d}}{\mathrm{d}x}\,ax \qquad \forall x \in \mathbb{R}_{>0}.$$

Thus, the inequality above must hold.

2.: f_a is strictly decreasing if and only if the first derivative of f_a,

$$\tfrac{\mathrm{d}f_a}{\mathrm{d}x}(x) = \frac{1}{x^2}\left(2-\log(1+ax)\left(1+\frac{2}{ax}\right)\right) \qquad \forall x \in \mathbb{R}_{>0}$$

is negative [For04, p. 165], and $\frac{\mathrm{d}f_a}{\mathrm{d}x}(x)$ is negative if and only if

$$\log(1+ax)\,(ax+2) \overset{!}{>} 2ax.$$

Both sides are equal to 0 if $x = 0$, and the left-hand side increases faster than the right-hand side,

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(\log(1+ax)\,(ax+2)\big) = a\left(1+\frac{1}{1+ax}+\log(1+ax)\right) > 2a = \frac{\mathrm{d}}{\mathrm{d}x}\,2ax$$

for all $x \in \mathbb{R}_{>0}$. Thus, the inequality above must hold.

3.: f_a is strictly convex if and only if the second derivative of f_a,

$$\frac{\mathrm{d}^2 f_a}{\mathrm{d}x^2}(x) = \frac{1}{x^3}\left(2\,\log(1+ax)\left(1+\frac{3}{ax}\right) - \frac{1}{1+ax} - 5\right) \qquad \forall x \in \mathbb{R}_{>0}$$

is positive [For04, p. 166], and $\frac{\mathrm{d}^2 f_a}{\mathrm{d}x^2}(x)$ is positive if and only if

$$2\log(1+ax)\,(ax+3) \overset{!}{>} 5ax + \frac{ax}{1+ax}.$$

Both sides are equal to 0 if $x = 0$, and the left-hand side increases faster than the right-hand side,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\big(2\log(1+ax)\,(ax+3)\big) &= a\left(2+\frac{1}{1+ax}+2\log(1+ax)\right)\\ &= a\left(2+\frac{1}{(1+ax)^2}+\frac{3+4ax}{(1+ax)^2}+2\log(1+ax)\right)\\ &> a\left(2+\frac{1}{(1+ax)^2}+3\right)\\ &= \frac{\mathrm{d}}{\mathrm{d}x}\left(5ax+\frac{ax}{1+ax}\right) \qquad \forall x \in \mathbb{R}_{>0}.\end{aligned}$$

Thus, the inequality above must hold.

to 4.: It is well-known that

$$\lim_{x\to\infty} f_a(x) = \lim_{x\to\infty}\left(\frac{\log(1+ax)}{x} + \frac{\log(1+ax)}{ax^2} - \frac{1}{x}\right) = 0.$$

Since f_a is decreasing, this limit is also the infimum of f_a.

According to l'Hôpital's Rule [For04, p. 171] it holds

$$\lim_{x\to 0}\frac{\log(1+ax)}{x} = \lim_{x\to 0}\frac{a}{1+ax} = a$$

and

$$\lim_{x\to 0}\frac{\log(1+ax)-ax}{ax^2} = \lim_{x\to 0}\frac{\frac{a}{1+ax}-a}{2ax} = \lim_{x\to 0}\frac{\frac{-a^2}{(1+ax)^2}}{2a} = -\frac{a}{2}.$$

Thus, $f_a(x)$ tends to $\frac{a}{2}$ if x approaches 0. Since f_a is decreasing, this is also its supremum. □

A.3 Lemma. *For any $a \in \mathbb{R}_{>0}$ define*

$$f_a\colon \mathbb{R}_{>0} \to \mathbb{R}\colon\ x \longmapsto \frac{x\left(1-\mathrm{e}^{-\frac{a}{x+1}}\right)}{\mathrm{e}^{\frac{xa}{x+1}}-1},$$

then:

1. *f_a is strictly decreasing.*
2. *Supremum and infimum of f_a are*

$$\sup_{x\in\mathbb{R}_{>0}} f_a(x) = \frac{1-\mathrm{e}^{-a}}{a} < 1 \qquad \text{and} \qquad \inf_{x\in\mathbb{R}_{>0}} f_a(x) = \frac{a}{\mathrm{e}^{a}-1} > \mathrm{e}^{-a}.$$

Proof. **1.:** f_a is strictly decreasing if and only if its derivative is negative [For04, p. 165]. The derivatives of numerator and denominator of f_a are

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x - x\mathrm{e}^{-\frac{a}{x+1}}\right) = 1-\mathrm{e}^{-\frac{a}{x+1}} - \frac{xa}{(x+1)^2}\,\mathrm{e}^{-\frac{a}{x+1}}$$
$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{\frac{xa}{x+1}}-1\right) = \frac{a}{(x+1)^2}\,\mathrm{e}^{\frac{xa}{x+1}}.$$

Since the quotient rule from differential calculus holds [For04, p. 154], f_a is strictly decreasing if the product of numerator and derivative of denominator is greater than the product of denominator and derivative of numerator for every $x \in \mathbb{R}_{>0}$. So, it has to be shown that

$$\left(x - x\mathrm{e}^{-\frac{a}{x+1}}\right)\left(\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{\frac{xa}{x+1}}-1\right)\right) \overset{!}{>} \left(\mathrm{e}^{\frac{xa}{x+1}}-1\right)\left(\frac{\mathrm{d}}{\mathrm{d}x}\left(x - x\mathrm{e}^{-\frac{a}{x+1}}\right)\right) \forall x \in \mathbb{R}_{>0}$$

which is equivalent to the inequation

$$a\left(\mathrm{e}^{a}-1\right) \overset{!}{>} \frac{(x+1)^2}{x}\left(\mathrm{e}^{a} - \mathrm{e}^{\frac{xa}{x+1}} - \mathrm{e}^{\frac{a}{x+1}} + 1\right) \qquad \forall x \in \mathbb{R}_{>0}.$$

The left-hand side can be expressed by integrals,

$$a\left(\mathrm{e}^{a}-1\right) = \int_0^a \left((b+1)\mathrm{e}^{b}-1\right)\,\mathrm{d}b = \int_0^a\int_0^b (2+c)\mathrm{e}^{c}\,\mathrm{d}c\,\mathrm{d}b$$
$$= \int_0^a\int_0^b \left(2+\int_0^c 1\,\mathrm{d}y\right)\mathrm{e}^{c}\,\mathrm{d}c\,\mathrm{d}b,$$

as well as the right-hand side,

$$\begin{aligned}
\frac{(x+1)^2}{x}\left(\mathrm{e}^a - \mathrm{e}^{\frac{xa}{x+1}} - \mathrm{e}^{\frac{a}{x+1}} + 1\right)&\\
&= \frac{(x+1)^2}{x}\int_0^a \left(\mathrm{e}^b - \frac{x}{x+1}\,\mathrm{e}^{\frac{xb}{x+1}} - \frac{1}{x+1}\,\mathrm{e}^{\frac{b}{x+1}}\right)\mathrm{d}b\\
&= \int_0^a\int_0^b \left(\frac{(x+1)^2}{x}\,\mathrm{e}^c - x\,\mathrm{e}^{\frac{xc}{x+1}} - \frac{1}{x}\,\mathrm{e}^{\frac{c}{x+1}}\right)\mathrm{d}c\ \mathrm{d}b\\
&= \int_0^a\int_0^b \left(2 + \int_0^c \frac{x\,\mathrm{e}^{-\frac{y}{x+1}} + \mathrm{e}^{-\frac{yx}{x+1}}}{x+1}\ \mathrm{d}y\right)\mathrm{e}^c\ \mathrm{d}c\ \mathrm{d}b.
\end{aligned}$$

Since it is

$$\frac{x\mathrm{e}^{-\frac{y}{x+1}} + \mathrm{e}^{-\frac{yx}{x+1}}}{x+1} < 1 \qquad \forall x, y \in \mathbb{R}_{>0},$$

the inequation above really holds.

2.: f_a is strictly decreasing, hence

$$\sup_{x\in\mathbb{R}_{>0}} f_a(x) = \lim_{x\to 0} f_a(x) \qquad \text{and} \qquad \inf_{x\in\mathbb{R}_{>0}} f_a(x) = \lim_{x\to\infty} f_a(x).$$

l'Hôspital's Rule [For04, p. 171] yields

$$\lim_{x\to 0} f_a(x) = \lim_{x\to 0} \frac{1 - \mathrm{e}^{-\frac{a}{x+1}} - \frac{xa}{(x+1)^2}\,\mathrm{e}^{-\frac{a}{x+1}}}{\frac{a}{(x+1)^2}\,\mathrm{e}^{\frac{xa}{x+1}}} = \frac{1-\mathrm{e}^{-a}}{a}$$

and

$$\lim_{x\to\infty} f_a(x) = \frac{1}{\mathrm{e}^a - 1}\lim_{x\to\infty}\frac{1 - \mathrm{e}^{-\frac{a}{x+1}}}{\frac{1}{x}} = \frac{1}{\mathrm{e}^a - 1}\lim_{x\to\infty}\frac{-\frac{a}{(x+1)^2}\,\mathrm{e}^{-\frac{a}{x+1}}}{-\frac{1}{x^2}} = \frac{a}{\mathrm{e}^a - 1}.$$

The fact

$$1 - \mathrm{e}^{-a} = \int_0^a \mathrm{e}^{-b}\ \mathrm{d}b < \int_0^a 1\ \mathrm{d}b = a \qquad \forall a \in \mathbb{R}_{>0}$$

finishes the proof. □

A.4 Lemma. *Let the situation be as in Theorem 3.5.2 with E_1, E_2, V_1 and V_2 as defined there, then it holds*

$$\mathbb{E}[E_1] = \mathbb{E}\big[\tfrac{1}{L}\big]\,\mathbb{E}[N_{\text{num}}]\,, \qquad \mathbb{E}[V_1] = \mathbb{E}\big[\tfrac{1}{L}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}]\,,$$
$$\mathbb{E}[E_2] = \mathbb{E}[N_{\text{num}}]\,, \qquad \mathbb{E}[V_2] = \mathbb{V}\text{ar}[N_{\text{num}}] - \frac{1}{m}\,\mathbb{V}\text{ar}[N_{\text{num}}]\,,$$

and

$$\begin{aligned}
\mathbb{V}\text{ar}[E_1] &= \frac{1}{m}\left(\mathbb{E}\big[\tfrac{1}{L^3}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}] + \mathbb{V}\text{ar}\big[\tfrac{1}{L}\big]\,\mathbb{E}[N_{\text{num}}]^2\right),\\
\mathbb{V}\text{ar}[E_2] &= \frac{1}{m}\,\mathbb{E}\big[\tfrac{1}{L}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}]\,,\\
\mathbb{V}\text{ar}[V_1] &= \frac{1}{m}\left(\mathbb{E}\big[\tfrac{1}{L^3}\big]\,\kappa_4[N_{\text{num}}] + \left(3\,\mathbb{E}\big[\tfrac{1}{L^2}\big] - \mathbb{E}\big[\tfrac{1}{L}\big]^2\right)\mathbb{V}\text{ar}[N_{\text{num}}]^2\right) + \mathcal{O}\big(m^{-2}\big)\,,\\
\mathbb{V}\text{ar}[V_2] &= \frac{1}{m}\left(\mathbb{E}\big[\tfrac{1}{L}\big]\,\kappa_4[N_{\text{num}}] + 2\,\mathbb{V}\text{ar}[N_{\text{num}}]^2\right) + \mathcal{O}\big(m^{-2}\big)\,,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}\text{ov}[V_1, E_1] &= \frac{1}{m}\left(\mathbb{E}\big[\tfrac{1}{L^3}\big]\,\kappa_3[N_{\text{num}}] + \mathbb{V}\text{ar}\big[\tfrac{1}{L}\big]\,\mathbb{V}\text{ar}[N_{\text{num}}]\,\mathbb{E}[N_{\text{num}}]\right) + \mathcal{O}\big(m^{-2}\big)\,,\\
\mathbb{C}\text{ov}[V_2, E_2] &= \frac{1}{m}\,\mathbb{E}\big[\tfrac{1}{L}\big]\,\kappa_3[N_{\text{num}}] + \mathcal{O}\big(m^{-2}\big)\,.
\end{aligned}$$

Proof. Before calculating the expectations, variances and covariances, note that the additivity of the cumulants (see Section 2.4.6) ensure that

$$\kappa_n\Big[N_{\text{num}}^{*L}\Big|\,L\Big] = \sum_{i=1}^{L}\kappa_n[N_{\text{num}}|L] = L\,\kappa_n[N_{\text{num}}] \qquad \forall n \in \mathbb{N},$$

because it is $N_{\text{num}}^{*L} = \sum_{i=1}^{L} N_i$ with statistically independent and identically distributed variates N_i, $N_i \sim N_{\text{num}}$. In addition, some formulas in Section 2.4.6 show that the moments of a random variable are polynomials in cumulants. According to these formulas and due to

$$\mathbb{E}\Big[\tfrac{(N_{\text{num}}^{*L})^n}{L^c}\Big] = \mathbb{E}\Big[\tfrac{1}{L^c}\mathbb{E}\Big[\Big(N_{\text{num}}^{*L}\Big)^n\,\Big|\,L\Big]\Big] \qquad \forall c \in \mathbb{R},\ \forall n \in \mathbb{N},$$

it holds for all $c \in \mathbb{R}$

$$\begin{aligned}
\mathbb{E}\Big[\tfrac{N^{*L}_{\mathrm{num}}}{L^c}\Big] &= \mathbb{E}\big[\tfrac{1}{L^{c-1}}\big]\,\kappa_1[N_{\mathrm{num}}]\,,\\
\mathbb{E}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L^c}\Big] &= \mathbb{E}\big[\tfrac{1}{L^{c-1}}\big]\,\kappa_2[N_{\mathrm{num}}] + \mathbb{E}\big[\tfrac{1}{L^{c-2}}\big]\,\kappa_1[N_{\mathrm{num}}]^2\,,\\
\mathbb{E}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^3}{L^c}\Big] &= \mathbb{E}\big[\tfrac{1}{L^{c-1}}\big]\,\kappa_3[N_{\mathrm{num}}] + 3\,\mathbb{E}\big[\tfrac{1}{L^{c-2}}\big]\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]\\
&\qquad + \mathbb{E}\big[\tfrac{1}{L^{c-3}}\big]\,\kappa_1[N_{\mathrm{num}}]^3\,, \qquad\qquad \text{(A.1)}\\
\mathbb{E}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^4}{L^c}\Big] &= \mathbb{E}\big[\tfrac{1}{L^c}\big]\kappa_4[N_{\mathrm{num}}]\\
&\quad + \mathbb{E}\big[\tfrac{1}{L^{c-2}}\big]\,\big(4\,\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + 3\,\kappa_2[N_{\mathrm{num}}]^2\big)\\
&\quad + 6\,\mathbb{E}\big[\tfrac{1}{L^{c-3}}\big]\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]^2\\
&\quad + \mathbb{E}\big[\tfrac{1}{L^{c-4}}\big]\,\kappa_1[N_{\mathrm{num}}]^4\,.
\end{aligned}$$

- calculation of $\mathbb{E}[E_1]$ and $\mathbb{E}[E_2]$:

The equations in Equation (A.1) above directly yield

$$\mathbb{E}[E_i] = \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\Big[\tfrac{N_j}{{L_j}^{3-i}}\Big] = \mathbb{E}\Big[\tfrac{N^{*L}_{\mathrm{num}}}{L^{3-i}}\Big] = \mathbb{E}\big[\tfrac{1}{L^{2-i}}\big]\;\mathbb{E}[N_{\mathrm{num}}] \qquad \forall i \in \{1,2\}.$$

- calculation of $\mathbb{E}[V_1]$:

Since V_1 is an unbiased estimator of the variance of ${N_1}/{L_1} \sim {N^{*L}_{\mathrm{num}}}/{L}$ [LC98, p 55], it holds

$$\mathbb{E}[V_1] = \mathbb{V}\mathrm{ar}\Big[\tfrac{N^{*L}_{\mathrm{num}}}{L}\Big]\,.$$

Now, the equations in Equation (A.1) above yield

$$\mathbb{V}\mathrm{ar}\Big[\tfrac{N^{*L}_{\mathrm{num}}}{L}\Big] = \mathbb{E}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L^2}\Big] - \mathbb{E}\Big[\tfrac{N^{*L}_{\mathrm{num}}}{L}\Big]^2 = \mathbb{E}\big[\tfrac{1}{L}\big]\,\kappa_2[N_{\mathrm{num}}] = \mathbb{E}\big[\tfrac{1}{L}\big]\,\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]\,.$$

- calculation of $\mathbb{E}[V_2]$:

By definition, the expectation of V_2 is

$$\mathbb{E}[V_2] = \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\Big[\tfrac{{N_j}^2}{L_j}\Big] - \frac{1}{m}\,\mathbb{E}\Big[\tfrac{\left(\sum_{j=1}^m N_j\right)^2}{\sum_{j=1}^m L_j}\Big] = \mathbb{E}\Big[\tfrac{\left(N^{*L}_{\mathrm{num}}\right)^2}{L}\Big] - \frac{1}{m}\,\mathbb{E}\Big[\tfrac{\left(\sum_{j=1}^m N_j\right)^2}{\sum_{j=1}^m L_j}\Big]\,.$$

On the one hand, the equations in Equation (A.1) at the beginning of this proof verify that

$$\mathbb{E}\Big[\tfrac{\left(N^{*L}_{\mathrm{num}}\right)^2}{L}\Big] = \mathbb{V}\mathrm{ar}[N_{\mathrm{num}}] + \mathbb{E}[L]\,\mathbb{E}[N_{\mathrm{num}}]^2\,.$$

On the other hand, the definition of $N_{\text{num}}^{*L} = \sum_{i=1}^{L} N_i$ with statistically independent and identically distributed N_i, $N_i \sim N_{\text{num}}$, ensures

$$\mathbb{E}\left[\frac{(\sum_{j=1}^{m} \mathcal{N}_j)^2}{\sum_{j=1}^{m} L_j} \,\middle|\, (L_j)_{1\le j\le m} \right]$$
$$= \frac{\mathbb{V}\mathrm{ar}\left[\sum_{j=1}^{m} \mathcal{N}_j \,\middle|\, (L_j)_{1\le j\le m} \right] + \mathbb{E}\left[\sum_{j=1}^{m} \mathcal{N}_j \,\middle|\, (L_j)_{1\le j\le m} \right]^2}{\sum_{j=1}^{m} L_j}$$
$$= \frac{\sum_{j=1}^{m} \mathbb{V}\mathrm{ar}\left[N_{\text{num}}^{*L_j} \,\middle|\, L_j \right] + \left(\sum_{j=1}^{m} \mathbb{E}\left[N_{\text{num}}^{*L_j} \,\middle|\, L_j \right] \right)^2}{\sum_{j=1}^{m} L_j}$$
$$= \frac{\left(\sum_{j=1}^{m} L_j \right) \mathbb{V}\mathrm{ar}[N_{\text{num}}] + \left(\sum_{j=1}^{m} L_j \right)^2 \mathbb{E}[N_{\text{num}}]^2}{\sum_{j=1}^{m} L_j},$$

and therefore

$$\frac{1}{m}\,\mathbb{E}\left[\frac{(\sum_{j=1}^{m} \mathcal{N}_j)^2}{\sum_{j=1}^{m} L_j} \right] = \frac{1}{m}\,\mathbb{E}\left[\mathbb{E}\left[\frac{(\sum_{j=1}^{m} \mathcal{N}_j)^2}{\sum_{j=1}^{m} L_j} \,\middle|\, (L_j)_{1\le j\le m} \right] \right]$$
$$= \frac{1}{m} \mathbb{V}\mathrm{ar}[N_{\text{num}}] + \mathbb{E}[L]\,\mathbb{E}[N_{\text{num}}]^2 .$$

- calculation of $\mathbb{V}\mathrm{ar}[E_1]$ and $\mathbb{V}\mathrm{ar}[E_2]$:

By definition, the variance of E_i is

$$\mathbb{V}\mathrm{ar}[E_i] = \frac{1}{m^2} \sum_{j=1}^{m} \mathbb{V}\mathrm{ar}\left[\frac{\mathcal{N}_j}{{L_j}^{3-i}} \right] = \frac{1}{m} \mathbb{V}\mathrm{ar}\left[\frac{N_{\text{num}}^{*L}}{L^{3-i}} \right] = \frac{1}{m} \left(\mathbb{E}\left[\frac{(N_{\text{num}}^{*L})^2}{L^{6-2i}} \right] - \mathbb{E}\left[\frac{N_{\text{num}}^{*L}}{L^{3-i}} \right]^2 \right)$$

for all $i \in \{1, 2\}$. Hence, the equations in Equation (A.1) at the beginning of this proof directly yield

$$\mathbb{V}\mathrm{ar}[E_i] = \frac{1}{m} \left(\mathbb{E}\left[\tfrac{1}{L^{5-2i}} \right] \kappa_2[N_{\text{num}}] + \mathbb{E}\left[\tfrac{1}{L^{4-2i}} \right] \kappa_1[N_{\text{num}}]^2 - \mathbb{E}\left[\tfrac{1}{L^{2-i}} \right]^2 \kappa_1[N_{\text{num}}]^2 \right)$$
$$= \frac{1}{m} \left(\mathbb{E}\left[\tfrac{1}{L^{5-2i}} \right] \mathbb{V}\mathrm{ar}[N_{\text{num}}] + \mathbb{V}\mathrm{ar}\left[\tfrac{1}{L^{2-i}} \right] \mathbb{E}[N_{\text{num}}]^2 \right)$$

for all $i \in \{1, 2\}$.

- calculation of $\mathbb{V}\mathrm{ar}[V_1]$:

Since V_1 is the (unbiased) sample variance of ${\mathcal{N}_1}/{L_1} \sim {N_{\text{num}}^{*L}}/{L}$, it holds due to

Cramér [Cra62, pp. 366–367]

$$\begin{aligned}\mathbb{V}\mathrm{ar}[V_1] &= \frac{m\left(C_4 - {C_2}^2\right)}{(m-1)^2} - \frac{2\left(C_4 - 2{C_2}^2\right)}{(m-1)^2} + \frac{\left(C_4 - 3{C_2}^2\right)}{m(m-1)^2} \\ &= \frac{\left(C_4 - {C_2}^2\right)}{m} + \frac{2{C_2}^2}{m(m-1)} \\ &= \frac{\left(C_4 - {C_2}^2\right)}{m} + \mathcal{O}\left(m^{-2}\right),\end{aligned}$$

where

$$C_4 := \mathbb{E}\left[\left(\tfrac{N^{*L}_{\mathrm{num}}}{L} - \mathbb{E}\left[\tfrac{N^{*L}_{\mathrm{num}}}{L}\right]\right)^4\right] \qquad \text{and} \qquad C_2 := \mathbb{V}\mathrm{ar}\left[\tfrac{N^{*L}_{\mathrm{num}}}{L}\right].$$

The facts from Section 2.4.6 and the relations in Equation (3.8) on page 44 yield

$$\begin{aligned}C_4 - {C_2}^2 &= \kappa_4\left[\tfrac{N^{*L}_{\mathrm{num}}}{L}\right] + 2\kappa_2\left[\tfrac{N^{*L}_{\mathrm{num}}}{L}\right]^2 \\ &= \mathbb{E}\left[\tfrac{1}{L^3}\right]\kappa_4[N_{\mathrm{num}}] + 3\,\mathbb{V}\mathrm{ar}\left[\tfrac{1}{L}\right]\kappa_2[N_{\mathrm{num}}]^2 + 2\,\mathbb{E}\left[\tfrac{1}{L}\right]^2\kappa_2[N_{\mathrm{num}}]^2.\end{aligned}$$

Finally, note that it holds by definition $3\,\mathbb{V}\mathrm{ar}\left[\tfrac{1}{L}\right] + 2\,\mathbb{E}\left[\tfrac{1}{L}\right]^2 = 3\,\mathbb{E}\left[\tfrac{1}{L^2}\right] - \mathbb{E}\left[\tfrac{1}{L}\right]^2$ and $\kappa_2[N_{\mathrm{num}}] = \mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]$.

- calculation of $\mathbb{V}\mathrm{ar}[V_2]$:

By definition, the variance of V_2 satisfies

$$\begin{aligned}&\mathbb{V}\mathrm{ar}[V_2] \\ &= \frac{1}{m^2}\sum_{j=1}^{m}\mathbb{V}\mathrm{ar}\left[\tfrac{{\mathcal{N}_j}^2}{L_j}\right] + \frac{1}{m^2}\mathbb{V}\mathrm{ar}\left[\tfrac{\left(\sum_{j=1}^m \mathcal{N}_j\right)^2}{\sum_{j=1}^m L_j}\right] - \frac{2}{m^2}\sum_{j=1}^{m}\mathbb{C}\mathrm{ov}\left[\tfrac{{\mathcal{N}_j}^2}{L_j}, \tfrac{\left(\sum_{i=1}^m \mathcal{N}_i\right)^2}{\sum_{i=1}^m L_i}\right] \\ &= \frac{1}{m}\mathbb{V}\mathrm{ar}\left[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L}\right] + \frac{1}{m^2}\mathbb{V}\mathrm{ar}\left[\tfrac{\left(\sum_{j=1}^m \mathcal{N}_j\right)^2}{\sum_{j=1}^m L_j}\right] - \frac{2}{m}\,\mathbb{C}\mathrm{ov}\left[\tfrac{{\mathcal{N}_1}^2}{L_1}, \tfrac{\left(\sum_{i=1}^m \mathcal{N}_i\right)^2}{\sum_{i=1}^m L_i}\right].\end{aligned}$$

At first, let us calculate both variance terms. The equations in Equation (A.1) at the beginning of this proof yield

$$\begin{aligned}&\frac{1}{m}\mathbb{V}\mathrm{ar}\left[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L}\right] \\ &\quad= \frac{1}{m}\left(\mathbb{E}\left[\tfrac{(N^{*L}_{\mathrm{num}})^4}{L^2}\right] - \mathbb{E}\left[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L}\right]^2\right) \\ &\quad= \frac{1}{m}\Bigg(\mathbb{E}\left[\tfrac{1}{L}\right]\kappa_4[N_{\mathrm{num}}] + 4\,\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + 2\,\kappa_2[N_{\mathrm{num}}]^2 \\ &\qquad\qquad + 4\,\mathbb{E}[L]\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]^2 + \mathbb{V}\mathrm{ar}[L]\,\kappa_1[N_{\mathrm{num}}]^4\Bigg). \qquad \text{(A.2)}\end{aligned}$$

Since the additivity of the cumulants (see Section 2.4.6) ensure

$$\kappa_n\Big[\sum_{j=1}^m \mathcal{N}_j \Big| (L_j)_{j\in\mathbb{N}}\Big] = \sum_{j=1}^m \kappa_n\Big[N_{\text{num}}^{*L_j}\Big| L_j\Big] = \sum_{j=1}^m \sum_{i=1}^{L_j} \kappa_n[N_{\text{num}}|L_j]$$
$$= \sum_{j=1}^m L_j\,\kappa_n[N_{\text{num}}] \qquad \forall n \in \mathbb{N},$$

Equation (A.1) can also be formulated for $\sum_{j=1}^m \mathcal{N}_j$ and $\sum_{j=1}^m L_j$ instead of N_{num}^{*L} and L respectively, which yields

$$\begin{aligned}
&\frac{1}{m^2}\mathbb{V}\text{ar}\Big[\frac{(\sum_{j=1}^m \mathcal{N}_j)^2}{\sum_{j=1}^m L_j}\Big] \\
&\quad = \frac{1}{m^2}\left(\mathbb{E}\Big[\frac{(\sum_{j=1}^m \mathcal{N}_j)^4}{(\sum_{j=1}^m L_j)^2}\Big] + \mathbb{E}\Big[\frac{(\sum_{j=1}^m \mathcal{N}_j)^2}{\sum_{j=1}^m L_j}\Big]^2\right) \\
&\quad = \frac{1}{m^2}\Bigg(\mathbb{E}\Big[\frac{1}{\sum_{j=1}^m L_j}\Big]\,\kappa_4[N_{\text{num}}] + 4\,\kappa_3[N_{\text{num}}]\,\kappa_1[N_{\text{num}}] + 2\,\kappa_2[N_{\text{num}}]^2 \\
&\qquad\qquad + 4m\,\mathbb{E}[L]\,\kappa_2[N_{\text{num}}]\,\kappa_1[N_{\text{num}}]^2 + m\,\mathbb{V}\text{ar}[L]\,\kappa_1[N_{\text{num}}]^4\Bigg) \\
&\quad = \frac{1}{m}\Big(4\,\mathbb{E}[L]\,\kappa_2[N_{\text{num}}]\,\kappa_1[N_{\text{num}}]^2 + \mathbb{V}\text{ar}[L]\,\kappa_1[N_{\text{num}}]^4\Big) + \mathcal{O}\big(m^{-2}\big)\,. \qquad (A.3)
\end{aligned}$$

With this, the variance terms are calculated. Secondly, the covariance term is needed,

$$\mathbb{C}\text{ov}\Big[\frac{\mathcal{N}_1{}^2}{L_1}, \frac{(\sum_{i=1}^m \mathcal{N}_i)^2}{\sum_{i=1}^m L_i}\Big] = \mathbb{E}\Big[\frac{\mathcal{N}_1{}^2}{L_1}\,\frac{(\sum_{i=1}^m \mathcal{N}_i)^2}{\sum_{i=1}^m L_i}\Big] - \mathbb{E}\Big[\frac{\mathcal{N}_1{}^2}{L_1}\Big]\,\mathbb{E}\Big[\frac{(\sum_{i=1}^m \mathcal{N}_i)^2}{\sum_{i=1}^m L_i}\Big]\,.$$

According to Equation (A.1), the rear expectation terms are

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\mathcal{N}_1{}^2}{L_1}\Big]\,\mathbb{E}\Big[\frac{(\sum_{i=1}^m \mathcal{N}_i)^2}{\sum_{i=1}^m L_i}\Big] \\
&= \kappa_2[N_{\text{num}}]^2 + (m+1)\mathbb{E}[L]\,\kappa_2[N_{\text{num}}]\,\kappa_1[N_{\text{num}}]^2 + m\mathbb{E}[L]^2\,\kappa_1[N_{\text{num}}]^4\,. \qquad (A.4)
\end{aligned}$$

For the calculation of the fore expectation term, define the random variables $K_m \coloneqq \sum_{i=2}^m L_j$, $\boldsymbol{K} \coloneqq (L_j)_{j\in\mathbb{N}}$ and $H_m \coloneqq \sum_{j=2}^m \mathcal{N}_j$, so that

$$\mathbb{E}\Big[\frac{\mathcal{N}_1{}^2}{L_1}\,\frac{(\sum_{i=1}^m \mathcal{N}_i)^2}{\sum_{i=1}^m L_i}\Big|(L_j)_{j\in\mathbb{N}}\Big] = \mathbb{E}\Big[\frac{\mathcal{N}_1{}^2}{L_1}\,\frac{(\mathcal{N}_1+H_m)^2}{L_1+K_m}\Big|\,\boldsymbol{K}\Big]\,.$$

Similar to above, the additivity of the cumulants (see Section 2.4.6) ensure

$$\kappa_n[\mathcal{N}_1|\,\boldsymbol{K}] = L_1\,\kappa_n[N_{\text{num}}] \quad \text{and} \quad \kappa_n[H_m|\,\boldsymbol{K}] = K_m\,\kappa_n[N_{\text{num}}] \qquad \forall n \in \mathbb{N}.$$

Moreover, Section 2.4.6 provides formulas which shows that the moments of a random variable are polynomials in cumulants. These formulas together with the additivity of the cumulants yield

$$\begin{aligned}
&\mathbb{E}\Big[\tfrac{\mathcal{N}_1{}^2}{L_1}\,\tfrac{(\mathcal{N}_1+H_m)^2}{L_1+K_m}\Big|\,\boldsymbol{K}\Big] \\
&\quad= \frac{\mathbb{E}\big[\mathcal{N}_1{}^4\big|\,\boldsymbol{K}\big] + 2\,\mathbb{E}\big[\mathcal{N}_1{}^3\big|\,\boldsymbol{K}\big]\,\mathbb{E}[H_m|\,\boldsymbol{K}] + \mathbb{E}\big[\mathcal{N}_1{}^2\big|\,\boldsymbol{K}\big]\,\mathbb{E}\big[H_m{}^2\big|\,\boldsymbol{K}\big]}{L_1(L_1+K_m)} \\
&\quad= \frac{6L_1{}^2+7L_1K_m+K_m{}^2}{L_1+K_m}\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]^2 + \frac{3L_1+K_m}{L_1+K_m}\,\kappa_2[N_{\mathrm{num}}]^2 \\
&\qquad+ \frac{L_1{}^3+2L_1{}^2K_m+L_1K_m{}^2}{L_1+K_m}\,\kappa_1[N_{\mathrm{num}}]^4 + \frac{4L_1+2K_m}{L_1+K_m}\,\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] \\
&\qquad+ \frac{1}{L_1+K_m}\,\kappa_4[N_{\mathrm{num}}]\,.
\end{aligned}$$

This can be simplified to

$$\begin{aligned}
\mathbb{E}\Big[\tfrac{\mathcal{N}_1{}^2}{L_1}\,\tfrac{(\mathcal{N}_1+H_m)^2}{L_1+K_m}\Big|\,\boldsymbol{K}\Big] &= (6L_1+K_m)\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]^2 + \kappa_2[N_{\mathrm{num}}]^2 \\
&\quad+ \big(L_1{}^2+L_1K_m\big)\,\kappa_1[N_{\mathrm{num}}]^4 + 2\,\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] \\
&\quad+ \frac{2L_1\,\big(\kappa_2[N_{\mathrm{num}}]^2+\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]\big)+\kappa_4[N_{\mathrm{num}}]}{L_1+K_m}.
\end{aligned}$$

The expectation of this term is

$$\begin{aligned}
\mathbb{E}\Big[\tfrac{\mathcal{N}_1{}^2}{L_1}\,\tfrac{(\mathcal{N}_1+H_m)^2}{L_1+K_m}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\tfrac{\mathcal{N}_1{}^2}{L_1}\,\tfrac{(\mathcal{N}_1+H_m)^2}{L_1+K_m}\Big|\,\boldsymbol{K}\Big]\Big] \\
&= (5+m)\,\mathbb{E}[L]\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]^2 + \kappa_2[N_{\mathrm{num}}]^2 \\
&\quad+ \big(\mathbb{E}\big[L^2\big]+(m-1)\mathbb{E}[L]^2\big)\,\kappa_1[N_{\mathrm{num}}]^4 + 2\,\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] \\
&\quad+ \mathcal{O}\big(m^{-1}\big)\,.
\end{aligned}$$

Together with Equation (A.4) from above, this yields

$$\begin{aligned}
&\frac{1}{m}\,\mathbb{C}\mathrm{ov}\Big[\tfrac{\mathcal{N}_1{}^2}{L_1},\tfrac{\left(\sum_{i=1}^m \mathcal{N}_i\right)^2}{\sum_{i=1}^m L_i}\Big] \\
&\quad= \frac{1}{m}\Big(4\,\mathbb{E}[L]\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]^2 + \mathbb{V}\mathrm{ar}[L]\,\kappa_1[N_{\mathrm{num}}]^4 + 2\,\kappa_3[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]\Big) \\
&\qquad+ \mathcal{O}\big(m^{-2}\big)\,.
\end{aligned}$$

This last result leads together with Equation (A.2) and Equation (A.3) from above to

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[V_2] &= \frac{1}{m}\mathbb{V}\mathrm{ar}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L}\Big] + \frac{1}{m^2}\mathbb{V}\mathrm{ar}\Big[\tfrac{\left(\sum_{j=1}^m \mathcal{N}_j\right)^2}{\sum_{j=1}^m L_j}\Big] - \frac{2}{m}\,\mathbb{C}\mathrm{ov}\Big[\tfrac{\mathcal{N}_1{}^2}{L_1},\tfrac{\left(\sum_{i=1}^m \mathcal{N}_i\right)^2}{\sum_{i=1}^m L_i}\Big] \\
&= \frac{1}{m}\,\big(\mathbb{E}\big[\tfrac{1}{L}\big]\,\kappa_4[N_{\mathrm{num}}] + 2\,\mathbb{V}\mathrm{ar}[N_{\mathrm{num}}]^2\big) + \mathcal{O}\big(m^{-2}\big)\,.
\end{aligned}$$

• calculation of $\mathbb{C}\mathrm{ov}[V_1, E_1]$:
The covariance of V_1 and E_1 is

$$\begin{aligned}\mathbb{C}\mathrm{ov}[V_1, E_1] &= \mathbb{C}\mathrm{ov}\Big[\tfrac{1}{m-1}\textstyle\sum_{j=1}^{m}\Big(\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j} - \tfrac{1}{m}\textstyle\sum_{i=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,i}}{L_i}\Big)^2, \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big]\\ &= \mathbb{C}\mathrm{ov}\Big[\tfrac{1}{m-1}\textstyle\sum_{j=1}^{m}\Big(\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big]\\ &\qquad - \mathbb{C}\mathrm{ov}\Big[\tfrac{1}{m(m-1)}\Big(\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big].\end{aligned}$$

The first covariance term can be transformed into

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\Big[\tfrac{1}{m-1}\textstyle\sum_{j=1}^{m}\Big(\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big]\\ &\qquad = \frac{1}{m(m-1)}\sum_{j=1}^{m}\sum_{i=1}^{m}\mathbb{C}\mathrm{ov}\Big[\Big(\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{\boldsymbol{N}_{\!L\,i}}{{L_i}^2}\Big]\\ &\qquad = \frac{1}{m(m-1)}\sum_{j=1}^{m}\mathbb{C}\mathrm{ov}\Big[\Big(\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big]\\ &\qquad = \frac{1}{m-1}\,\mathbb{C}\mathrm{ov}\Big[\Big(\tfrac{N^{*L}_{\mathrm{num}}}{L}\Big)^2, \tfrac{N^{*L}_{\mathrm{num}}}{L^2}\Big]\\ &\qquad = \frac{1}{m-1}\Big(\mathbb{E}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^3}{L^4}\Big] - \mathbb{E}\Big[\tfrac{(N^{*L}_{\mathrm{num}})^2}{L^2}\Big]\,\mathbb{E}\Big[\tfrac{N^{*L}_{\mathrm{num}}}{L^2}\Big]\Big).\end{aligned}$$

Again, due to the equations in Equation (A.1) at the beginning of this proof, the first covariance term is

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\Big[\tfrac{1}{m-1}\textstyle\sum_{j=1}^{m}\Big(\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big]\\ &= \frac{1}{m-1}\Big(\mathbb{E}\big[\tfrac{1}{L^3}\big]\,\kappa_3[N_{\mathrm{num}}] + \Big(3\,\mathbb{E}\big[\tfrac{1}{L^2}\big] - \mathbb{E}\big[\tfrac{1}{L}\big]^2\Big)\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]\Big)\\ &= \frac{1}{m}\Big(\mathbb{E}\big[\tfrac{1}{L^3}\big]\,\kappa_3[N_{\mathrm{num}}] + \Big(3\,\mathbb{E}\big[\tfrac{1}{L^2}\big] - \mathbb{E}\big[\tfrac{1}{L}\big]^2\Big)\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]\Big) + \mathcal{O}\big(m^{-2}\big).\end{aligned}$$

Therefore, the proof is established if it can be verified that the second covariance term satisfies

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\Big[\tfrac{1}{m(m-1)}\Big(\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{L_j}\Big)^2, \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_{\!L\,j}}{{L_j}^2}\Big]\\ &\qquad \overset{!}{=} \frac{2}{m}\,\mathbb{E}\big[\tfrac{1}{L^2}\big]\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + \mathcal{O}\big(m^{-2}\big).\end{aligned}$$

Since the random variables (N_i, L_i) and (N_j, L_j) are statistically independent as long as $i \neq j$, the following transformation holds:

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\tfrac{1}{m(m-1)}\left(\sum_{j=1}^{m}\tfrac{N_j}{L_j}\right)^2, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j}{L_j{}^2}\right] \\
&\quad = \frac{1}{m^2(m-1)} \sum_{1\le h,i,j\le m} \mathbb{C}\mathrm{ov}\left[\tfrac{N_h}{L_h}\tfrac{N_i}{L_i}, \tfrac{N_j}{L_j{}^2}\right] \\
&\quad = \frac{1}{m(m-1)} \sum_{1\le i,j\le m} \mathbb{C}\mathrm{ov}\left[\tfrac{N_1}{L_1}\tfrac{N_i}{L_i}, \tfrac{N_j}{L_j{}^2}\right] \\
&\quad = \frac{\mathbb{C}\mathrm{ov}\left[\frac{N_1{}^2}{L_1{}^2}, \frac{N_1}{L_1{}^2}\right] + \sum_{i=2}^{m}\mathbb{C}\mathrm{ov}\left[\frac{N_1}{L_1}\frac{N_i}{L_i}, \frac{N_1}{L_1{}^2}\right] + \sum_{2\le i,j\le m}\mathbb{C}\mathrm{ov}\left[\frac{N_1}{L_1}\frac{N_i}{L_i}, \frac{N_j}{L_j{}^2}\right]}{m(m-1)} \\
&\quad = \frac{\mathbb{C}\mathrm{ov}\left[\frac{N_1{}^2}{L_1{}^2}, \frac{N_1}{L_1{}^2}\right] + (m-1)\,\mathbb{E}\left[\frac{N_2}{L_2}\right]\mathbb{C}\mathrm{ov}\left[\frac{N_1}{L_1}, \frac{N_1}{L_1{}^2}\right] + \mathbb{E}\left[\frac{N_1}{L_1}\right](m-1)\,\mathbb{C}\mathrm{ov}\left[\frac{N_2}{L_2}, \frac{N_2}{L_2{}^2}\right]}{m(m-1)} \\
&\quad = \frac{2}{m}\,\mathbb{E}\left[\tfrac{N_{\mathrm{num}}^{*L}}{L}\right]\mathbb{C}\mathrm{ov}\left[\tfrac{N_{\mathrm{num}}^{*L}}{L}, \tfrac{N_{\mathrm{num}}^{*L}}{L^2}\right] + \mathcal{O}\big(m^{-2}\big) \\
&\quad = \frac{2}{m}\,\mathbb{E}\left[\tfrac{N_{\mathrm{num}}^{*L}}{L}\right]\left(\mathbb{E}\left[\tfrac{(N_{\mathrm{num}}^{*L})^2}{L^3}\right] - \mathbb{E}\left[\tfrac{N_{\mathrm{num}}^{*L}}{L}\right]\mathbb{E}\left[\tfrac{N_{\mathrm{num}}^{*L}}{L^2}\right]\right) + \mathcal{O}\big(m^{-2}\big)\,.
\end{aligned}$$

Thus, due to the equations in Equation (A.1) at the beginning of this proof, the second covariance term is indeed equal to $\frac{2}{m}\,\mathbb{E}\left[\frac{1}{L^2}\right]\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + \mathcal{O}\big(m^{-2}\big)$.

- calculation of $\mathbb{C}\mathrm{ov}[V_1, E_1]$:

The covariance of V_2 and E_2 is

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[V_2, E_2] \\
&\quad = \mathbb{C}\mathrm{ov}\left[\tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j{}^2}{L_j}, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j}{L_j}\right] - \mathbb{C}\mathrm{ov}\left[\tfrac{1}{m}\tfrac{\left(\sum_{j=1}^{m}N_j\right)^2}{\sum_{j=1}^{m}L_j}, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j}{L_j}\right].
\end{aligned}$$

Since the random variables (N_i, L_i) and (N_j, L_j) are statistically independent as long as $i \neq j$, it holds

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left[\tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j{}^2}{L_j}, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j}{L_j}\right] &= \frac{1}{m^2}\sum_{j=1}^{m}\sum_{i=1}^{m}\mathbb{C}\mathrm{ov}\left[\tfrac{N_j{}^2}{L_j}, \tfrac{N_j}{L_j}\right] \\
&= \frac{1}{m}\,\mathbb{C}\mathrm{ov}\left[\tfrac{N_1{}^2}{L_1}, \tfrac{N_1}{L_1}\right] \\
&= \frac{1}{m}\left(\mathbb{E}\left[\tfrac{(N_{\mathrm{num}}^{*L})^3}{L^2}\right] - \mathbb{E}\left[\tfrac{(N_{\mathrm{num}}^{*L})^2}{L}\right]\mathbb{E}\left[\tfrac{N_{\mathrm{num}}^{*L}}{L}\right]\right).
\end{aligned}$$

Again, the equations in Equation (A.1) at the beginning of this proof yield

$$\mathbb{C}\mathrm{ov}\left[\tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j{}^2}{L_j}, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{N_j}{L_j}\right] = \frac{1}{m}\left(\mathbb{E}\left[\tfrac{1}{L}\right]\kappa_3[N_{\mathrm{num}}] + 2\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}]\right).$$

According to this, the proof is established if it can be verified that the second covariance term satisfies

$$\mathbb{C}\mathrm{ov}\left[\tfrac{1}{m}\tfrac{\left(\sum_{j=1}^{m}\boldsymbol{N}_j\right)^2}{\sum_{j=1}^{m}L_j},\, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_j}{L_j}\right] \overset{!}{=} 2\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + \mathcal{O}\big(m^{-2}\big)\,.$$

To verify this, first note that

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\tfrac{1}{m}\tfrac{\left(\sum_{j=1}^{m}\boldsymbol{N}_j\right)^2}{\sum_{j=1}^{m}L_j},\, \tfrac{1}{m}\sum_{j=1}^{m}\tfrac{\boldsymbol{N}_j}{L_j}\right]\\
&\qquad = \frac{1}{m^2}\sum_{i=1}^{m}\mathbb{C}\mathrm{ov}\left[\tfrac{\left(\sum_{j=1}^{m}\boldsymbol{N}_j\right)^2}{\sum_{j=1}^{m}L_j},\, \tfrac{\boldsymbol{N}_i}{L_i}\right]\\
&\qquad = \frac{1}{m}\,\mathbb{C}\mathrm{ov}\left[\tfrac{\left(\sum_{j=1}^{m}\boldsymbol{N}_j\right)^2}{\sum_{j=1}^{m}L_j},\, \tfrac{\boldsymbol{N}_1}{L_1}\right]\\
&\qquad = \frac{1}{m}\left(\mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\,\tfrac{\left(\sum_{i=1}^{m}\boldsymbol{N}_i\right)^2}{\sum_{i=1}^{m}L_i}\right] - \mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\right]\mathbb{E}\left[\tfrac{\left(\sum_{i=1}^{m}\boldsymbol{N}_i\right)^2}{\sum_{i=1}^{m}L_i}\right]\right).
\end{aligned}$$

According to Equation (A.1) at the beginning of this proof, which can also be formulated for $\sum_{j=1}^{m}\boldsymbol{N}_j$ and $\sum_{j=1}^{m}L_j$ instead of N_{num}^{*L} and L respectively, the rear expectation terms are

$$\mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\right]\mathbb{E}\left[\tfrac{\left(\sum_{i=1}^{m}\boldsymbol{N}_i\right)^2}{\sum_{i=1}^{m}L_i}\right] = \kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + m\,\mathbb{E}[L]\,\kappa_1[N_{\mathrm{num}}]^3\,. \tag{A.5}$$

For the calculation of the fore expectation term, define the random variables $K_m := \sum_{i=2}^{m}L_j$, $\boldsymbol{K} := (L_j)_{j\in\mathbb{N}}$ and $H_m := \sum_{j=2}^{m}\boldsymbol{N}_j$, so that

$$\mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\,\tfrac{\left(\sum_{i=1}^{m}\boldsymbol{N}_i\right)^2}{\sum_{i=1}^{m}L_i}\,\middle|\,(L_j)_{j\in\mathbb{N}}\right] = \mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\,\tfrac{(\boldsymbol{N}_1+H_m)^2}{L_1+K_m}\,\middle|\,\boldsymbol{K}\right].$$

From this point, the calculation is very similar to the calculation of $\mathbb{V}\mathrm{ar}[V_2]$ on page 170. Similar to there, the additivity of the cumulants and the formulas from Section 2.4.6 yield

$$\begin{aligned}
&\mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\,\tfrac{(\boldsymbol{N}_1+H_m)^2}{L_1+K_m}\,\middle|\,\boldsymbol{K}\right]\\
&\quad = \frac{\mathbb{E}\big[\boldsymbol{N}_1{}^3\big|\,\boldsymbol{K}\big] + 2\,\mathbb{E}\big[\boldsymbol{N}_1{}^2\big|\,\boldsymbol{K}\big]\,\mathbb{E}[H_m|\,\boldsymbol{K}] + \mathbb{E}[\boldsymbol{N}_1|\,\boldsymbol{K}]\,\mathbb{E}\big[H_m{}^2\big|\,\boldsymbol{K}\big]}{L_1(L_1+K_m)}\\
&\quad = 3\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + (L_1+K_m)\,\kappa_1[N_{\mathrm{num}}]^3 + \frac{\kappa_3[N_{\mathrm{num}}]}{L_1+K_m}.
\end{aligned}$$

The expectation of this term is

$$\begin{aligned}
\mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\,\tfrac{(\boldsymbol{N}_1+H_m)^2}{L_1+K_m}\right] &= \mathbb{E}\left[\mathbb{E}\left[\tfrac{\boldsymbol{N}_1}{L_1}\,\tfrac{(\boldsymbol{N}_1+H_m)^2}{L_1+K_m}\,\middle|\,\boldsymbol{K}\right]\right]\\
&= 3\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + m\,\mathbb{E}[L]\,\kappa_1[N_{\mathrm{num}}]^3 + \mathcal{O}\big(m^{-1}\big)\,.
\end{aligned}$$

Together with Equation (A.5) from above, this yields

$$\mathbb{C}\mathrm{ov}\left[\frac{1}{m}\,\frac{\left(\sum_{j=1}^{m} N\!\!\!L_j\right)^2}{\sum_{j=1}^{m} L_j}, \frac{1}{m}\sum_{j=1}^{m}\frac{N\!\!\!L_j}{L_j}\right] = 2\,\kappa_2[N_{\mathrm{num}}]\,\kappa_1[N_{\mathrm{num}}] + \mathcal{O}\big(m^{-2}\big)\,,$$

which should be proved. □

B. Tables

Table B.1.: *First four cumulants of* $\sqrt{\frac{m}{2}}\,(\hat{D}_2 - 1)$ *under Poisson hypothesis; based on* 10^6 *replications (cf. Section 5.2.1)*

		$\mu = 10^{-4}$				$\mu = 10^{-3}$				$\mu = 10^{-2}$			
m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
10	U	−0.221	0.863	0.933	1.906	−0.221	0.895	0.816	1.173	−0.224	0.901	0.811	1.106
	Exp	−0.220	0.860	1.027	2.413	−0.222	0.890	0.817	1.209	−0.225	0.899	0.810	1.110
20	U	−0.156	0.925	0.743	1.160	−0.157	0.945	0.613	0.629	−0.157	0.949	0.602	0.582
	Exp	−0.154	0.923	0.800	1.378	−0.159	0.945	0.622	0.655	−0.159	0.951	0.608	0.593
50	U	−0.100	0.966	0.509	0.543	−0.100	0.979	0.405	0.272	−0.101	0.980	0.393	0.238
	Exp	−0.099	0.963	0.552	0.653	−0.100	0.979	0.416	0.295	−0.098	0.980	0.394	0.242
100	U	−0.072	0.980	0.373	0.301	−0.070	0.990	0.288	0.136	−0.070	0.988	0.281	0.124
	Exp	−0.069	0.983	0.416	0.391	−0.069	0.990	0.300	0.161	−0.071	0.989	0.282	0.121
500	U	−0.031	0.998	0.175	0.072	−0.030	0.997	0.132	0.029	−0.032	0.998	0.125	0.029
	Exp	−0.031	0.999	0.192	0.092	−0.032	0.997	0.130	0.027	−0.033	0.999	0.126	0.028
1000	U	−0.021	1.000	0.123	0.029	−0.022	0.999	0.093	0.022	−0.021	1.000	0.091	0.008
	Exp	−0.021	0.999	0.142	0.054	−0.024	0.998	0.093	0.014	−0.023	0.999	0.091	0.010

Table B.2.: *Deviation of the quantiles of* $\sqrt{\frac{m}{2}}\,(\hat{D}_2 - 1)$ *under Poisson hypothesis from the quantiles of the standard normal distribution; based on* 10^6 *replications (cf. Section 5.2.1)*

μ	m	km	$q_\Delta(0.01)$	$q_\Delta(0.05)$	$q_\Delta(0.1)$	$q_\Delta(0.25)$	$q_\Delta(0.5)$	$q_\Delta(0.75)$	$q_\Delta(0.9)$	$q_\Delta(0.95)$	$q_\Delta(0.99)$
10^{-4}	10	U	0.619	0.216	0.035	−0.208	−0.375	−0.407	−0.285	−0.139	0.357
		Exp	0.652	0.245	0.057	−0.199	−0.384	−0.421	−0.299	−0.139	0.409
	20	U	0.452	0.144	0.010	−0.166	−0.277	−0.281	−0.176	−0.052	0.329
		Exp	0.485	0.168	0.026	−0.161	−0.282	−0.288	−0.173	−0.043	0.368
	50	U	0.294	0.084	−0.006	−0.118	−0.182	−0.173	−0.091	−0.003	0.247
		Exp	0.327	0.101	0.005	−0.118	−0.187	−0.178	−0.094	0.000	0.282
	100	U	0.215	0.055	−0.010	−0.088	−0.132	−0.123	−0.059	0.004	0.195
		Exp	0.238	0.068	0.000	−0.090	−0.136	−0.122	−0.052	0.018	0.219
	500	U	0.094	0.023	−0.007	−0.044	−0.060	−0.049	−0.019	0.015	0.104
		Exp	0.107	0.027	−0.008	−0.044	−0.061	−0.052	−0.016	0.020	0.108
	1000	U	0.066	0.014	−0.008	−0.031	−0.042	−0.034	−0.009	0.014	0.073
		Exp	0.077	0.023	−0.003	−0.031	−0.043	−0.037	−0.010	0.016	0.093
10^{-3}	10	U	0.561	0.160	−0.014	−0.235	−0.368	−0.366	−0.241	−0.105	0.296
		Exp	0.563	0.164	−0.011	−0.234	−0.369	−0.370	−0.246	−0.110	0.283
	20	U	0.381	0.091	−0.032	−0.182	−0.263	−0.249	−0.145	−0.043	0.239
		Exp	0.384	0.091	−0.031	−0.183	−0.265	−0.253	−0.151	−0.043	0.249
	50	U	0.228	0.043	−0.033	−0.124	−0.169	−0.150	−0.081	−0.011	0.180
		Exp	0.233	0.046	−0.032	−0.124	−0.169	−0.154	−0.080	−0.007	0.186
	100	U	0.154	0.026	−0.029	−0.091	−0.119	−0.103	−0.048	−0.002	0.128
		Exp	0.161	0.030	−0.025	−0.089	−0.119	−0.103	−0.049	0.002	0.145
	500	U	0.074	0.011	−0.015	−0.040	−0.052	−0.044	−0.018	0.006	0.066
		Exp	0.066	0.008	−0.016	−0.042	−0.054	−0.046	−0.022	0.006	0.060
	1000	U	0.043	0.006	−0.011	−0.030	−0.037	−0.032	−0.014	0.003	0.050
		Exp	0.045	0.006	−0.011	−0.030	−0.041	−0.033	−0.014	0.001	0.041
10^{-2}	10	U	0.558	0.153	−0.023	−0.244	−0.371	−0.364	−0.237	−0.097	0.285
		Exp	0.556	0.151	−0.022	−0.243	−0.372	−0.368	−0.237	−0.100	0.279
	20	U	0.373	0.086	−0.037	−0.184	−0.261	−0.244	−0.143	−0.040	0.237
		Exp	0.370	0.082	−0.039	−0.186	−0.265	−0.246	−0.145	−0.041	0.244
	50	U	0.221	0.035	−0.038	−0.124	−0.167	−0.150	−0.080	−0.010	0.170
		Exp	0.224	0.041	−0.036	−0.123	−0.165	−0.146	−0.075	−0.010	0.170
	100	U	0.154	0.027	−0.027	−0.089	−0.118	−0.102	−0.051	−0.004	0.121
		Exp	0.152	0.022	−0.029	−0.090	−0.118	−0.103	−0.053	0.000	0.126
	500	U	0.059	0.009	−0.016	−0.043	−0.053	−0.045	−0.020	0.001	0.056
		Exp	0.059	0.004	−0.018	−0.044	−0.055	−0.044	−0.022	0.001	0.062
	1000	U	0.048	0.005	−0.012	−0.030	−0.036	−0.029	−0.011	0.005	0.048
		Exp	0.049	0.006	−0.011	−0.031	−0.038	−0.031	−0.012	0.003	0.042

Table B.3.: *Power of the test statistic* $\sqrt{\frac{m}{2}}(\hat{D}_2 - 1)$ *in % under Poisson hypothesis* ($\mathrm{IOD} = 1$) *and several alternative hypotheses* ($\mathrm{IOD} > 1$) *for significance level* $1 - \alpha = 0.95$; *based on* 10^5 *replications (cf. Section 5.2.2)*

			IOD=1			IOD=1.25			IOD=1.5			IOD=1.75		
μ	m	km	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}
10^{-4}	10	U	0.05	97.30	2.65	0.03	91.52	8.45	0.01	83.58	16.41	0.01	74.75	25.24
		Exp	0.04	97.29	2.67	0.01	91.40	8.59	0.01	83.77	16.23	0.01	75.83	24.17
	20	U	0.64	96.48	2.88	0.14	86.71	13.16	0.05	72.08	27.87	0.02	56.88	43.10
		Exp	0.51	96.56	2.93	0.12	87.01	12.87	0.05	72.70	27.25	0.02	58.33	41.66
	50	U	1.34	95.80	2.87	0.11	76.37	23.52	0.01	46.48	53.51	0.00	23.66	76.34
		Exp	1.17	95.86	2.98	0.09	76.79	23.12	0.01	47.59	52.40	0.00	25.61	74.39
	100	U	1.74	95.39	2.88	0.03	61.89	38.08	0	20.48	79.53	0	4.65	95.35
		Exp	1.60	95.48	2.92	0.04	62.06	37.90	0.00	21.85	78.15	0	5.30	94.70
	500	U	2.17	95.03	2.81	0	6.91	93.09	0	0.00	100.00	0	0	100
		Exp	2.16	95.03	2.81	0	7.57	92.43	0	0.01	99.99	0	0	100
	1000	U	2.35	94.95	2.70	0	0.24	99.76	0	0	100	0	0	100
		Exp	2.17	95.10	2.73	0	0.30	99.70	0	0	100	0	0	100
10^{-3}	10	U	0.11	97.22	2.67	0.04	90.88	9.08	0.03	81.68	18.30	0.01	71.13	28.86
		Exp	0.12	97.33	2.55	0.06	90.99	8.95	0.02	81.59	18.40	0.01	71.64	28.34
	20	U	0.92	96.25	2.83	0.20	86.69	13.11	0.04	69.81	30.16	0.02	51.93	48.06
		Exp	0.94	96.24	2.82	0.23	86.53	13.24	0.07	69.87	30.07	0.02	52.44	47.54
	50	U	1.64	95.53	2.83	0.14	76.15	23.72	0.01	42.60	57.39	0	18.43	81.57
		Exp	1.64	95.51	2.85	0.12	76.13	23.75	0.01	43.04	56.95	0.00	18.69	81.31
	100	U	1.89	95.31	2.80	0.03	60.81	39.16	0.00	16.81	83.19	0	2.62	97.39
		Exp	1.87	95.39	2.74	0.03	60.60	39.37	0	16.78	83.22	0	2.61	97.39
	500	U	2.34	95.11	2.56	0	5.61	94.39	0	0.00	100.00	0	0	100
		Exp	2.32	95.02	2.66	0	5.58	94.42	0	0.00	100.00	0	0	100
	1000	U	2.39	94.94	2.67	0	0.14	99.86	0	0	100	0	0	100
		Exp	2.38	94.96	2.66	0	0.17	99.84	0	0	100	0	0	100
10^{-2}	10	U	0.13	97.19	2.68	0.05	90.96	8.99	0.02	81.35	18.63	0.02	70.72	29.27
		Exp	0.13	97.10	2.78	0.03	90.97	8.99	0.01	81.32	18.67	0.01	70.78	29.22
	20	U	0.96	96.26	2.79	0.23	86.44	13.33	0.07	69.66	30.28	0.02	51.19	48.79
		Exp	0.99	96.24	2.77	0.24	86.45	13.31	0.05	69.47	30.48	0.02	51.30	48.68
	50	U	1.71	95.57	2.71	0.13	76.26	23.61	0.01	42.06	57.93	0.00	17.70	82.30
		Exp	1.68	95.57	2.75	0.13	76.10	23.78	0.01	42.13	57.86	0.00	17.98	82.02
	100	U	1.92	95.34	2.74	0.04	60.52	39.4	0.00	16.30	83.70	0	2.29	97.71
		Exp	1.95	95.29	2.76	0.03	60.93	39.04	0	16.29	83.71	0	2.33	97.67
	500	U	2.29	95.06	2.65	0	5.49	94.52	0	0.00	100.00	0	0	100
		Exp	2.24	95.12	2.64	0	5.53	94.47	0	0	100	0	0	100
	1000	U	2.41	95.01	2.58	0	0.13	99.87	0	0	100	0	0	100
		Exp	2.41	94.89	2.70	0	0.14	99.86	0	0	100	0	0	100

Table B.4.: *Power of the test statistic* $\sqrt{\frac{m}{2}}\left(\hat{D}_2 - 1\right)$ *in % under several alternative hypotheses* (IOD > 1) *for significance level* $1 - \alpha = 0.95$; *based on* 10^5 *replications (cf. Section 5.2.2)*

			IOD=2			IOD=2.5			IOD=3			IOD=5		
μ	m	km	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}
10^{-4}	10	U	0.00	66.27	33.73	0.00	52.56	47.44	0	42.21	57.79	0	22.21	77.79
		Exp	0.00	67.72	32.28	0.00	54.77	45.23	0	44.64	55.37	0	25.45	74.55
	20	U	0.00	43.71	56.29	0	26.03	73.97	0.00	15.68	84.32	0	3.60	96.40
		Exp	0.01	45.78	54.21	0.00	28.18	71.82	0	17.90	82.10	0	4.79	95.21
	50	U	0	11.19	88.81	0	2.45	97.55	0	0.58	99.42	0	0.01	99.99
		Exp	0	12.88	87.12	0	3.18	96.82	0	0.89	99.11	0	0.03	99.98
	100	U	0	0.79	99.22	0	0.03	99.97	0	0	100	0	0	100
		Exp	0	1.18	98.82	0	0.05	99.95	0	0.01	99.99	0	0	100
	500	U	0	0	100	0	0	100	0	0	100	0	0	100
		Exp	0	0	100	0	0	100	0	0	100	0	0	100
	1000	U	0	0	100	0	0	100	0	0	100	0	0	100
		Exp	0	0	100	0	0	100	0	0	100	0	0	100
10^{-3}	10	U	0.01	60.73	39.27	0.00	43.39	56.61	0	30.33	69.67	0	8.72	91.29
		Exp	0.01	61.38	38.62	0.00	43.97	56.03	0.00	31.17	68.82	0	9.19	90.81
	20	U	0.01	36.84	63.15	0.00	17.44	82.56	0	8.01	91.99	0	0.50	99.50
		Exp	0.01	37.24	62.76	0.00	17.55	82.45	0	8.31	91.69	0	0.55	99.45
	50	U	0	6.92	93.08	0	0.80	99.20	0	0.08	99.92	0	0.00	100.00
		Exp	0	7.07	92.93	0	0.82	99.18	0	0.10	99.91	0	0.00	100.00
	100	U	0	0.26	99.74	0	0.00	100.00	0	0.00	100.00	0	0	100
		Exp	0	0.31	99.70	0	0.00	100.00	0	0.00	100.00	0	0	100
	500	U	0	0	100	0	0	100	0	0	100	0	0	100
		Exp	0	0	100	0	0	100	0	0	100	0	0	100
	1000	U	0	0	100	0	0	100	0	0	100	0	0	100
		Exp	0	0	100	0	0	100	0	0	100	0	0	100
10^{-2}	10	U	0.01	59.61	40.38	0.00	41.65	58.35	0.00	28.97	71.03	0	7.27	92.73
		Exp	0.01	59.93	40.07	0	41.55	58.45	0	28.76	71.24	0	7.41	92.59
	20	U	0.01	35.54	64.45	0.00	15.90	84.10	0	6.98	93.02	0	0.35	99.65
		Exp	0.01	35.88	64.11	0.00	16.32	83.68	0	7.05	92.95	0	0.35	99.65
	50	U	0.00	6.41	93.59	0	0.66	99.34	0	0.07	99.93	0	0	100
		Exp	0	6.37	93.63	0	0.67	99.33	0	0.07	99.93	0	0	100
	100	U	0	0.24	99.76	0	0.01	100.00	0	0	100	0	0	100
		Exp	0	0.23	99.77	0	0.00	100.00	0	0	100	0	0	100
	500	U	0	0	100	0	0	100	0	0	100	0	0	100
		Exp	0	0	100	0	0	100	0	0	100	0	0	100
	1000	U	0	0	100	0	0	100	0	0	100	0	0	100
		Exp	0	0	100	0	0	100	0	0	100	0	0	100

Table B.5.: *Power of the test statistic* $\sqrt{\frac{m}{2}}\left(\hat{D}_2 - 1\right)$ *in % under several alternative hypotheses (*$\mathrm{IOD} < 1$*) for significance level* $1 - \alpha = 0.95$*; based on* 10^5 *replications (cf. Section 5.2.2)*

			IOD=0.25			IOD=0.5			IOD=0.75			IOD=0.9		
r	m	km	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}	H_{1-}	H_0	H_{1+}
1	10	U	17.54	82.46	0	1.58	98.41	0.00	0.34	99.41	0.25	0.17	98.62	1.21
		NBin	17.15	82.85	0	1.61	98.38	0.01	0.35	99.38	0.27	0.17	98.61	1.22
	20	U	93.02	6.98	0	26.11	73.89	0	4.06	95.83	0.11	1.43	97.53	1.04
		NBin	92.29	7.71	0	25.83	74.17	0	3.83	96.09	0.09	1.32	97.64	1.04
	50	U	100	0	0	88.67	11.33	0	18.08	81.91	0.01	4.04	95.39	0.57
		NBin	100	0	0	88.87	11.13	0	17.99	82.00	0.01	4.10	95.34	0.56
	100	U	100	0	0	99.90	0.10	0	43.69	56.31	0.00	7.58	92.12	0.30
		NBin	100	0	0	99.90	0.10	0	43.51	56.49	0	7.74	91.98	0.28
	500	U	100	0	0	100	0	0	99.63	0.37	0	35.27	64.71	0.01
		NBin	100	0	0	100	0	0	99.69	0.32	0	35.11	64.88	0.01
	1000	U	100	0	0	100	0	0	100.00	0.00	0	63.68	36.32	0.00
		NBin	100	0	0	100	0	0	100	0	0	63.65	36.35	0
10	10	U	16.09	83.91	0	1.79	98.21	0.00	0.39	99.30	0.31	0.21	98.51	1.29
		NBin	15.86	84.14	0	1.77	98.23	0.00	0.42	99.28	0.31	0.20	98.46	1.35
	20	U	95.16	4.84	0	28.89	71.11	0	4.91	94.99	0.10	1.82	97.15	1.03
		NBin	95.01	4.99	0	28.80	71.20	0	4.85	95.02	0.13	1.82	97.14	1.05
	50	U	100	0	0	88.08	11.92	0	19.77	80.21	0.01	4.69	94.70	0.61
		NBin	100.00	0.00	0	88.15	11.85	0	19.89	80.10	0.01	4.72	94.67	0.62
	100	U	100	0	0	99.81	0.19	0	44.14	55.87	0	8.22	91.49	0.29
		NBin	100	0	0	99.80	0.20	0	44.18	55.81	0.00	8.55	91.13	0.33
	500	U	100	0	0	100	0	0	99.52	0.48	0	35.62	64.37	0.01
		NBin	100	0	0	100	0	0	99.52	0.48	0	35.87	64.13	0.01
	1000	U	100	0	0	100	0	0	100	0	0	63.04	36.96	0
		NBin	100	0	0	100	0	0	100	0	0	63.43	36.57	0.00
100	10	U	15.98	84.02	0	1.84	98.16	0.00	0.38	99.34	0.28	0.21	98.46	1.33
		NBin	16.07	83.94	0	1.79	98.21	0.00	0.41	99.31	0.28	0.19	98.51	1.30
	20	U	95.34	4.66	0	29.36	70.64	0	5.00	94.89	0.11	1.88	97.07	1.05
		NBin	95.30	4.71	0	28.94	71.06	0.00	4.98	94.90	0.12	1.86	97.05	1.09
	50	U	100	0	0	87.98	12.02	0	19.74	80.25	0.02	4.85	94.56	0.59
		NBin	100	0	0	88.00	12.00	0	20.12	79.87	0.02	4.91	94.49	0.60
	100	U	100	0	0	99.80	0.20	0	44.52	55.48	0.00	8.38	91.28	0.34
		NBin	100	0	0	99.81	0.19	0	44.29	55.71	0.00	8.58	91.15	0.28
	500	U	100	0	0	100	0	0	99.55	0.45	0	35.55	64.44	0.01
		NBin	100	0	0	100	0	0	99.50	0.50	0	35.80	64.19	0.01
	1000	U	100	0	0	100	0	0	100.00	0.00	0	63.30	36.70	0.00
		NBin	100	0	0	100	0	0	100.00	0.00	0	63.21	36.79	0.00

Table B.6.: *Relative mean, relative standard deviation, and relative square root of inverse Fisher information of $\hat{\varrho}_m$; based on 10^5 replications (cf. Section 5.3.2)*

			$\mu = 10^{-4}$			$\mu = 10^{-3}$			$\mu = 10^{-2}$		
ϱ	m	km	$\frac{\text{mean}(\hat{\varrho}_m)}{\varrho}$	$\frac{\text{std}(\hat{\varrho}_m)}{\varrho}$	$\frac{1}{\varrho\sqrt{I_\varrho}}$	$\frac{\text{mean}(\hat{\varrho}_m)}{\varrho}$	$\frac{\text{std}(\hat{\varrho}_m)}{\varrho}$	$\frac{1}{\varrho\sqrt{I_\varrho}}$	$\frac{\text{mean}(\hat{\varrho}_m)}{\varrho}$	$\frac{\text{std}(\hat{\varrho}_m)}{\varrho}$	$\frac{1}{\varrho\sqrt{I_\varrho}}$
10^{-5}	10	U	—	—	0.625	—	—	0.459	1.328	0.733	0.442
		Exp	—	—	0.840	—	—	0.513	1.297	0.688	0.410
	20	U	—	—	0.462	1.142	0.412	0.327	1.123	0.364	0.295
		Exp	—	—	0.489	1.190	0.498	0.380	1.140	0.403	0.327
	50	U	1.094	0.344	0.292	1.052	0.229	0.211	1.045	0.206	0.191
		Exp	1.114	0.388	0.317	1.062	0.253	0.229	1.047	0.210	0.196
	100	U	1.045	0.227	0.211	1.026	0.161	0.155	1.022	0.140	0.135
		Exp	1.055	0.248	0.226	1.030	0.175	0.167	1.025	0.148	0.141
	500	U	1.009	0.096	0.095	1.005	0.069	0.068	1.004	0.060	0.060
		Exp	1.010	0.102	0.100	1.006	0.074	0.073	1.005	0.064	0.064
	1000	U	1.004	0.067	0.066	1.002	0.049	0.048	1.002	0.043	0.043
		Exp	1.005	0.070	0.069	1.003	0.051	0.051	1.002	0.045	0.045
10^{-4}	10	U	—	—	1.060	—	—	0.511	1.403	0.898	0.426
		Exp	—	—	1.060	—	—	0.506	1.359	0.789	0.430
	20	U	—	—	0.689	1.218	0.558	0.351	1.160	0.402	0.301
		Exp	—	—	0.744	1.219	0.556	0.359	1.157	0.405	0.300
	50	U	—	—	0.473	1.075	0.257	0.222	1.056	0.214	0.193
		Exp	—	—	0.480	1.079	0.270	0.232	1.055	0.211	0.190
	100	U	—	—	0.315	1.035	0.167	0.156	1.028	0.143	0.135
		Exp	—	—	0.336	1.035	0.170	0.158	1.027	0.142	0.134
	500	U	1.024	0.152	0.142	1.007	0.071	0.070	1.005	0.061	0.060
		Exp	1.026	0.160	0.149	1.007	0.072	0.071	1.005	0.061	0.061
	1000	U	1.012	0.103	0.101	1.003	0.049	0.049	1.003	0.043	0.043
		Exp	1.013	0.108	0.105	1.003	0.051	0.050	1.003	0.043	0.043
10^{-3}	10	U	—	—	5.219	—	—	0.901	—	—	0.490
		Exp	—	—	5.093	—	—	0.915	—	—	0.488
	20	U	—	—	3.603	—	—	0.640	1.214	0.546	0.346
		Exp	—	—	3.690	—	—	0.647	1.218	0.541	0.345
	50	U	—	—	2.267	—	—	0.405	1.076	0.255	0.219
		Exp	—	—	2.266	—	—	0.407	1.074	0.252	0.218
	100	U	—	—	1.611	—	—	0.287	1.036	0.167	0.155
		Exp	—	—	1.613	—	—	0.288	1.035	0.166	0.154
	500	U	—	—	0.718	1.021	0.136	0.128	1.007	0.070	0.069
		Exp	—	—	0.725	1.022	0.138	0.129	1.007	0.070	0.069
	1000	U	—	—	0.507	1.010	0.093	0.091	1.003	0.049	0.049
		Exp	—	—	0.512	1.010	0.093	0.091	1.003	0.049	0.049
10^{-2}	10	U	—	—	45.255	—	—	4.933	—	—	0.896
		Exp	—	—	45.382	—	—	4.932	—	—	0.897
	20	U	—	—	32.062	—	—	3.485	—	—	0.633
		Exp	—	—	32.086	—	—	3.488	—	—	0.634
	50	U	—	—	20.264	—	—	2.207	—	—	0.401
		Exp	—	—	20.302	—	—	2.209	—	—	0.401
	100	U	—	—	14.334	—	—	1.561	1.128	0.504	0.283
		Exp	—	—	14.364	—	—	1.564	1.127	0.494	0.283
	500	U	—	—	6.413	—	—	0.698	1.021	0.133	0.127
		Exp	—	—	6.425	—	—	0.699	1.021	0.134	0.127
	1000	U	—	—	4.535	—	—	0.494	1.010	0.092	0.090
		Exp	—	—	4.541	—	—	0.494	1.010	0.092	0.090

Table B.7.: *First four cumulants of* $\sqrt{I_\varrho}\,(\hat{\varrho}_m - \varrho) = \sqrt{I_{\mathrm{num}}(\varrho,\mu)_{11}}\,(\hat{\varrho}_m - \varrho)$; *based on* 10^5 *replications (cf. Section 5.3.2)*

			$\mu = 10^{-4}$				$\mu = 10^{-3}$				$\mu = 10^{-2}$			
ϱ	m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
10^{-5}	10	U	—	—	—	—	—	—	—	—	0.741	2.751	11.405	106.812
		Exp	—	—	—	—	—	—	—	—	0.723	2.813	14.097	202.149
	20	U	—	—	—	—	0.434	1.586	2.658	9.475	0.415	1.519	2.300	7.033
		Exp	—	—	—	—	0.500	1.717	3.521	19.023	0.436	1.522	2.219	6.421
	50	U	0.323	1.388	2.152	14.906	0.248	1.183	0.917	1.442	0.238	1.157	0.780	1.022
		Exp	0.361	1.493	2.408	9.607	0.271	1.214	1.044	1.764	0.239	1.157	0.808	1.093
	100	U	0.216	1.163	0.866	1.324	0.169	1.084	0.518	0.483	0.164	1.079	0.500	0.457
		Exp	0.242	1.202	1.014	1.711	0.182	1.099	0.589	0.593	0.178	1.090	0.535	0.477
	500	U	0.098	1.029	0.297	0.135	0.073	1.017	0.201	0.066	0.072	1.013	0.194	0.057
		Exp	0.104	1.041	0.320	0.164	0.080	1.018	0.233	0.102	0.077	1.010	0.200	0.063
	1000	U	0.066	1.014	0.205	0.094	0.051	1.013	0.144	0.029	0.051	1.018	0.135	0.046
		Exp	0.070	1.102	0.220	0.074	0.055	1.006	0.155	0.050	0.054	1.010	0.154	0.058
10^{-4}	10	U	—	—	—	—	—	—	—	—	0.946	4.435	47.182	1443.991
		Exp	—	—	—	—	—	—	—	—	0.833	3.365	20.116	293.276
	20	U	—	—	—	—	0.621	2.523	17.892	625.802	0.533	1.794	3.851	16.772
		Exp	—	—	—	—	0.608	2.399	10.359	123.958	0.523	1.826	4.343	24.766
	50	U	—	—	—	—	0.339	1.339	1.642	4.213	0.292	1.231	1.148	2.126
		Exp	—	—	—	—	0.340	1.351	1.680	4.420	0.291	1.238	1.146	2.107
	100	U	—	—	—	—	0.223	1.146	0.827	1.213	0.209	1.115	0.680	0.769
		Exp	—	—	—	—	0.223	1.155	0.850	1.248	0.203	1.115	0.671	0.788
	500	U	0.167	1.143	0.893	1.412	0.099	1.027	0.295	0.142	0.084	1.015	0.259	0.147
		Exp	0.173	1.146	0.949	1.643	0.101	1.025	0.280	0.149	0.089	1.015	0.236	0.092
	1000	U	0.121	1.059	0.539	0.543	0.065	1.007	0.192	0.044	0.064	1.003	0.194	0.066
		Exp	0.121	1.159	0.553	0.631	0.068	1.021	0.205	0.098	0.060	1.012	0.181	0.043
10^{-3}	10	U	—	—	—	—	—	—	—	—	—	—	—	—
		Exp	—	—	—	—	—	—	—	—	—	—	—	—
	20	U	—	—	—	—	—	—	—	—	0.617	2.490	26.674	2012.840
		Exp	—	—	—	—	—	—	—	—	0.634	2.461	11.407	150.599
	50	U	—	—	—	—	—	—	—	—	0.348	1.358	1.704	4.218
		Exp	—	—	—	—	—	—	—	—	0.339	1.332	1.630	4.021
	100	U	—	—	—	—	—	—	—	—	0.233	1.161	0.881	1.313
		Exp	—	—	—	—	—	—	—	—	0.230	1.154	0.869	1.325
	500	U	—	—	—	—	0.162	1.124	0.841	1.350	0.107	1.033	0.318	0.180
		Exp	—	—	—	—	0.168	1.139	0.878	1.508	0.101	1.030	0.308	0.224
	1000	U	—	—	—	—	0.116	1.060	0.525	0.532	0.071	1.022	0.224	0.090
		Exp	—	—	—	—	0.109	1.048	0.490	0.455	0.071	1.011	0.224	0.091
10^{-2}	10	U	—	—	—	—	—	—	—	—	—	—	—	—
		Exp	—	—	—	—	—	—	—	—	—	—	—	—
	20	U	—	—	—	—	—	—	—	—	—	—	—	—
		Exp	—	—	—	—	—	—	—	—	—	—	—	—
	50	U	—	—	—	—	—	—	—	—	—	—	—	—
		Exp	—	—	—	—	—	—	—	—	—	—	—	—
	100	U	—	—	—	—	—	—	—	—	0.453	3.172	125.359	15525.493
		Exp	—	—	—	—	—	—	—	—	0.449	3.036	181.658	39801.888
	500	U	—	—	—	—	—	—	—	—	0.164	1.110	0.806	1.268
		Exp	—	—	—	—	—	—	—	—	0.165	1.122	0.809	1.202
	1000	U	—	—	—	—	—	—	—	—	0.110	1.053	0.495	0.445
		Exp	—	—	—	—	—	—	—	—	0.113	1.060	0.506	0.463

Table B.8.: *Deviation of the quantiles of $\sqrt{I_\varrho}\,(\hat{\varrho}_m - \varrho) = \sqrt{I_{\mathrm{num}}(\varrho,\mu)_{11}}\,(\hat{\varrho}_m - \varrho)$ from the quantiles of the standard normal distribution; based on 10^5 replications (cf. Section 5.3.2)*

			$\mu = 10^{-2}$								
ϱ	m	km	$q_\Delta(0.01)$	$q_\Delta(0.05)$	$q_\Delta(0.1)$	$q_\Delta(0.25)$	$q_\Delta(0.5)$	$q_\Delta(0.75)$	$q_\Delta(0.9)$	$q_\Delta(0.95)$	$q_\Delta(0.99)$
10^{-5}	10	U	0.946	0.610	0.473	0.333	0.363	0.689	1.406	2.122	4.323
		Exp	1.047	0.683	0.526	0.345	0.314	0.569	1.174	1.830	3.935
	20	U	0.730	0.460	0.348	0.229	0.216	0.367	0.726	1.038	2.014
		Exp	0.581	0.345	0.257	0.195	0.260	0.516	0.975	1.384	2.468
	50	U	0.510	0.320	0.230	0.155	0.133	0.207	0.370	0.522	0.918
		Exp	0.478	0.294	0.216	0.140	0.134	0.228	0.409	0.578	0.995
	100	U	0.384	0.243	0.179	0.107	0.087	0.141	0.243	0.337	0.595
		Exp	0.324	0.187	0.135	0.088	0.105	0.191	0.346	0.472	0.788
	500	U	0.201	0.115	0.084	0.051	0.044	0.060	0.093	0.141	0.225
		Exp	0.067	0.025	0.015	0.019	0.048	0.109	0.193	0.258	0.417
	1000	U	0.121	0.075	0.051	0.036	0.027	0.039	0.076	0.109	0.161
		Exp	0.036	0.022	0.011	0.005	0.028	0.075	0.138	0.194	0.316
10^{-4}	10	U	1.098	0.716	0.554	0.387	0.415	0.843	1.816	2.870	6.197
		Exp	1.021	0.658	0.505	0.356	0.386	0.770	1.626	2.501	5.281
	20	U	0.871	0.561	0.427	0.285	0.272	0.482	0.93	1.384	2.688
		Exp	0.866	0.547	0.421	0.274	0.256	0.461	0.915	1.369	2.699
	50	U	0.624	0.385	0.295	0.180	0.151	0.249	0.457	0.658	1.207
		Exp	0.641	0.395	0.299	0.188	0.154	0.234	0.447	0.638	1.153
	100	U	0.493	0.301	0.226	0.138	0.113	0.176	0.322	0.446	0.798
		Exp	0.474	0.299	0.225	0.141	0.105	0.162	0.300	0.420	0.789
	500	U	0.240	0.144	0.109	0.059	0.041	0.066	0.118	0.159	0.303
		Exp	0.234	0.142	0.106	0.063	0.052	0.075	0.116	0.169	0.280
	1000	U	0.193	0.120	0.089	0.046	0.031	0.045	0.088	0.120	0.209
		Exp	0.174	0.105	0.073	0.038	0.029	0.048	0.086	0.121	0.218
10^{-3}	10	U	1.168	0.748	0.577	0.390	0.438	0.957	2.292	3.807	10.923
		Exp	1.182	0.762	0.584	0.403	0.453	1.002	2.350	4.009	11.562
	20	U	0.967	0.610	0.460	0.292	0.276	0.532	1.129	1.724	3.680
		Exp	0.969	0.608	0.461	0.297	0.285	0.549	1.153	1.763	3.806
	50	U	0.711	0.437	0.332	0.198	0.171	0.298	0.579	0.842	1.607
		Exp	0.721	0.447	0.331	0.198	0.166	0.283	0.553	0.808	1.545
	100	U	0.548	0.329	0.245	0.146	0.116	0.195	0.369	0.525	0.979
		Exp	0.542	0.331	0.243	0.143	0.116	0.189	0.352	0.508	0.922
	500	U	0.290	0.176	0.131	0.072	0.055	0.084	0.155	0.226	0.357
		Exp	0.273	0.167	0.120	0.073	0.056	0.076	0.134	0.199	0.375
	1000	U	0.201	0.112	0.085	0.050	0.034	0.054	0.105	0.149	0.263
		Exp	0.212	0.132	0.098	0.050	0.033	0.053	0.098	0.138	0.251
10^{-2}	10	U	1.544	0.995	0.739	0.430	0.545	4.463	∞	∞	∞
		Exp	1.538	0.992	0.737	0.426	0.547	4.332	∞	∞	∞
	20	U	1.346	0.846	0.619	0.337	0.313	1.188	6.562	136.20	∞
		Exp	1.343	0.845	0.615	0.332	0.308	1.187	6.658	115.21	∞
	50	U	1.085	0.658	0.470	0.243	0.180	0.502	1.496	2.823	11.113
		Exp	1.076	0.655	0.469	0.239	0.178	0.511	1.537	2.905	10.617
	100	U	0.875	0.524	0.371	0.186	0.125	0.303	0.805	1.360	3.297
		Exp	0.882	0.524	0.373	0.184	0.123	0.310	0.821	1.382	3.289
	500	U	0.512	0.285	0.198	0.099	0.057	0.112	0.258	0.399	0.836
		Exp	0.489	0.277	0.193	0.093	0.055	0.119	0.281	0.420	0.842
	1000	U	0.368	0.208	0.144	0.064	0.034	0.071	0.176	0.272	0.521
		Exp	0.371	0.205	0.142	0.066	0.037	0.077	0.178	0.285	0.538

Table B.9.: *Relative mean, relative standard deviation, and relative square root of inverse Fisher information of $\hat{\mu}_m$; based on 10^5 replications (cf. Section 5.4)*

			$\mu = 10^{-4}$			$\mu = 10^{-3}$			$\mu = 10^{-2}$		
ϱ	m	km	$\frac{\text{mean}(\hat{\mu}_m)}{\mu}$	$\frac{\text{std}(\hat{\mu}_m)}{\mu}$	$\frac{1}{\mu\sqrt{I_\mu}}$	$\frac{\text{mean}(\hat{\mu}_m)}{\mu}$	$\frac{\text{std}(\hat{\mu}_m)}{\mu}$	$\frac{1}{\mu\sqrt{I_\mu}}$	$\frac{\text{mean}(\hat{\mu}_m)}{\mu}$	$\frac{\text{std}(\hat{\mu}_m)}{\mu}$	$\frac{1}{\mu\sqrt{I_\mu}}$
10^{-5}	10	U	0.998	0.583	0.585	1.000	0.571	0.571	0.999	0.633	0.634
		Exp	1.001	0.827	0.826	1.000	0.712	0.711	0.998	0.420	0.421
	20	U	1.004	0.458	0.455	1.002	0.412	0.412	0.999	0.429	0.430
		Exp	1.000	0.463	0.464	1.001	0.406	0.405	1.000	0.456	0.456
	50	U	0.999	0.289	0.290	0.998	0.268	0.268	1.000	0.288	0.287
		Exp	0.999	0.305	0.307	1.000	0.262	0.263	1.001	0.267	0.267
	100	U	0.998	0.207	0.208	1.000	0.206	0.206	0.999	0.196	0.196
		Exp	0.999	0.205	0.204	1.000	0.205	0.206	1.001	0.186	0.186
	500	U	1.001	0.093	0.093	1.000	0.089	0.089	1.000	0.088	0.087
		Exp	1.000	0.095	0.095	1.000	0.091	0.091	1.000	0.088	0.088
	1000	U	1.000	0.065	0.065	1.000	0.063	0.063	1.000	0.063	0.063
		Exp	1.000	0.063	0.064	1.000	0.061	0.061	1.000	0.060	0.060
10^{-4}	10	U	1.000	0.314	0.313	0.999	0.260	0.260	1.001	0.190	0.190
		Exp	1.000	0.277	0.276	1.001	0.198	0.198	1.000	0.216	0.216
	20	U	1.000	0.178	0.178	0.999	0.141	0.141	1.000	0.136	0.136
		Exp	1.000	0.195	0.194	1.000	0.140	0.141	1.001	0.125	0.125
	50	U	1.000	0.138	0.138	1.000	0.096	0.096	1.000	0.092	0.092
		Exp	1.000	0.145	0.145	1.000	0.101	0.101	1.000	0.092	0.091
	100	U	1.000	0.085	0.086	1.000	0.065	0.065	1.000	0.064	0.064
		Exp	1.000	0.086	0.086	1.000	0.067	0.067	1.000	0.067	0.067
	500	U	1.000	0.039	0.039	1.000	0.030	0.030	1.000	0.028	0.028
		Exp	1.000	0.040	0.040	1.000	0.030	0.030	1.000	0.028	0.029
	1000	U	1.000	0.028	0.028	1.000	0.020	0.020	1.000	0.020	0.020
		Exp	1.000	0.028	0.028	1.000	0.020	0.020	1.000	0.020	0.020
10^{-3}	10	U	1.000	0.230	0.231	1.000	0.087	0.087	1.000	0.059	0.059
		Exp	0.999	0.213	0.212	1.000	0.100	0.100	1.000	0.067	0.067
	20	U	1.001	0.156	0.156	1.000	0.062	0.062	1.000	0.045	0.045
		Exp	0.999	0.171	0.171	1.000	0.062	0.062	1.000	0.048	0.048
	50	U	1.000	0.095	0.094	1.000	0.039	0.039	1.000	0.030	0.030
		Exp	1.000	0.087	0.087	1.000	0.042	0.042	1.000	0.029	0.029
	100	U	1.000	0.068	0.068	1.000	0.027	0.027	1.000	0.021	0.021
		Exp	1.000	0.064	0.064	1.000	0.025	0.025	1.000	0.021	0.021
	500	U	1.000	0.029	0.029	1.000	0.012	0.012	1.000	0.009	0.009
		Exp	1.000	0.030	0.029	1.000	0.012	0.012	1.000	0.009	0.009
	1000	U	1.000	0.020	0.020	1.000	0.009	0.009	1.000	0.007	0.007
		Exp	1.000	0.021	0.021	1.000	0.009	0.009	1.000	0.006	0.006
10^{-2}	10	U	1.000	0.184	0.185	1.000	0.066	0.066	1.000	0.026	0.026
		Exp	0.999	0.201	0.201	1.000	0.071	0.071	1.000	0.032	0.032
	20	U	0.999	0.136	0.135	1.000	0.042	0.042	1.000	0.021	0.021
		Exp	1.000	0.144	0.144	1.000	0.038	0.038	1.000	0.021	0.021
	50	U	1.000	0.081	0.081	1.000	0.030	0.030	1.000	0.013	0.013
		Exp	1.000	0.078	0.078	1.000	0.026	0.026	1.000	0.012	0.012
	100	U	1.000	0.062	0.062	1.000	0.020	0.020	1.000	0.008	0.008
		Exp	1.000	0.059	0.059	1.000	0.021	0.020	1.000	0.009	0.009
	500	U	1.000	0.028	0.028	1.000	0.009	0.009	1.000	0.004	0.004
		Exp	1.000	0.029	0.029	1.000	0.009	0.009	1.000	0.004	0.004
	1000	U	1.000	0.020	0.020	1.000	0.007	0.007	1.000	0.003	0.003
		Exp	1.000	0.020	0.020	1.000	0.007	0.007	1.000	0.003	0.003
∞	10	U	1.000	0.231	0.232	1.000	0.056	0.056	1.000	0.020	0.020
		Exp	1.001	0.223	0.224	1.000	0.086	0.086	1.000	0.018	0.018
	20	U	1.000	0.143	0.143	1.000	0.043	0.043	1.000	0.013	0.013
		Exp	1.000	0.128	0.128	1.000	0.043	0.043	1.000	0.012	0.012
	50	U	1.000	0.089	0.088	1.000	0.029	0.029	1.000	0.009	0.009
		Exp	1.000	0.096	0.096	1.000	0.028	0.027	1.000	0.009	0.009
	100	U	1.000	0.067	0.066	1.000	0.018	0.018	1.000	0.006	0.006
		Exp	1.000	0.066	0.066	1.000	0.019	0.019	1.000	0.006	0.006
	500	U	1.000	0.027	0.027	1.000	0.009	0.009	1.000	0.003	0.003
		Exp	1.000	0.027	0.027	1.000	0.008	0.008	1.000	0.003	0.003
	1000	U	1.000	0.019	0.020	1.000	0.006	0.006	1.000	0.002	0.002
		Exp	1.000	0.020	0.020	1.000	0.006	0.006	1.000	0.002	0.002

Table B.10.: *First four cumulants of* $\sqrt{I_\mu}\,(\hat{\mu}_m - \mu) = \sqrt{I_{\text{num}}(\varrho,\mu)_{22}}\,(\hat{\mu}_m - \mu)$*; based on* 10^5 *replications (cf. Section 5.4)*

			$\mu = 10^{-4}$				$\mu = 10^{-3}$				$\mu = 10^{-2}$			
ϱ	m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
10^{-5}	10	U	−0.004	0.994	1.108	1.829	−0.001	1.002	1.144	1.998	−0.001	0.998	1.272	2.504
		Exp	0.001	1.002	1.592	3.865	0.000	1.003	1.415	2.919	−0.005	0.994	0.816	0.971
	20	U	0.008	1.012	0.867	1.044	0.005	1.004	0.813	0.967	−0.002	0.994	0.857	1.156
		Exp	−0.001	0.997	0.876	1.113	0.003	1.006	0.811	0.981	0.000	1.000	0.902	1.173
	50	U	−0.005	0.997	0.557	0.467	−0.007	0.998	0.534	0.400	−0.001	1.006	0.594	0.538
		Exp	−0.003	0.992	0.581	0.530	−0.002	0.994	0.505	0.373	0.004	1.004	0.541	0.445
	100	U	−0.008	0.993	0.389	0.226	−0.001	0.992	0.411	0.286	−0.003	0.996	0.390	0.222
		Exp	−0.003	1.009	0.411	0.264	0.001	0.993	0.404	0.279	0.003	0.997	0.375	0.201
	500	U	0.013	0.999	0.181	0.077	0.003	1.001	0.176	0.046	−0.001	1.008	0.173	0.028
		Exp	−0.004	0.999	0.187	0.063	−0.001	0.998	0.191	0.045	−0.001	1.001	0.176	0.034
	1000	U	−0.002	1.004	0.126	0.048	0.001	0.999	0.129	0.040	0.001	0.998	0.120	0.019
		Exp	0.005	0.998	0.129	0.025	−0.006	0.991	0.112	0.034	0.000	1.006	0.113	0.038
10^{-4}	10	U	0.001	1.004	0.488	0.367	−0.005	1.005	0.512	0.385	0.005	1.002	0.369	0.195
		Exp	−0.001	1.002	0.413	0.263	0.004	1.003	0.369	0.174	0.002	1.004	0.443	0.292
	20	U	0.001	0.998	0.256	0.095	−0.005	0.997	0.265	0.088	0.000	1.006	0.275	0.108
		Exp	−0.001	1.006	0.304	0.143	0.000	0.995	0.258	0.075	0.005	1.005	0.247	0.118
	50	U	0.004	1.009	0.205	0.055	0.002	1.004	0.200	0.068	0.005	1.005	0.197	0.061
		Exp	−0.002	0.998	0.209	0.061	0.000	1.004	0.200	0.038	0.001	1.001	0.189	0.033
	100	U	0.000	0.997	0.125	0.012	0.004	1.003	0.118	0.020	0.000	1.003	0.142	0.027
		Exp	0.002	0.997	0.118	0.004	−0.001	1.003	0.124	0.046	−0.004	0.996	0.133	0.034
	500	U	−0.003	1.002	0.062	−0.012	0.001	1.002	0.074	−0.010	−0.004	0.999	0.065	0.021
		Exp	0.000	0.995	0.054	0.015	0.002	1.004	0.065	0.003	−0.002	0.993	0.058	0.009
	1000	U	0.003	1.001	0.035	0.011	−0.002	1.000	0.034	0.023	0.001	0.995	0.049	0.025
		Exp	0.001	0.995	0.044	0.015	0.006	1.001	0.033	−0.007	−0.003	0.998	0.038	−0.017
10^{-3}	10	U	0.000	0.993	0.251	0.063	−0.001	1.000	0.130	0.016	0.002	0.998	0.123	0.033
		Exp	−0.003	1.002	0.251	0.072	−0.001	1.000	0.142	0.004	0.004	1.001	0.134	0.015
	20	U	0.003	1.004	0.178	0.021	0.002	0.995	0.094	−0.005	0.002	1.004	0.101	0.022
		Exp	−0.007	1.000	0.188	0.051	0.001	1.000	0.081	0.017	−0.002	0.998	0.087	0.003
	50	U	0.001	1.009	0.104	0.016	0.001	0.995	0.062	0.001	−0.001	1.003	0.056	−0.001
		Exp	0.004	1.002	0.086	0.028	0.000	1.000	0.060	0.005	0.003	1.006	0.047	−0.012
	100	U	0.001	1.009	0.071	0.003	0.006	0.994	0.034	0.014	−0.004	1.003	0.027	0.013
		Exp	0.004	1.002	0.076	0.015	0.002	1.001	0.041	0.014	−0.001	1.011	0.035	−0.009
	500	U	−0.006	0.995	0.023	0.020	−0.002	0.996	0.013	−0.013	−0.003	0.993	0.011	0.013
		Exp	−0.002	1.009	0.029	−0.004	0.003	1.002	0.007	−0.004	−0.003	1.003	0.018	0.001
	1000	U	−0.002	0.999	0.013	0.002	−0.008	1.005	0.023	−0.010	0.003	0.995	0.018	0.003
		Exp	0.002	1.004	0.027	0.008	0.002	1.010	0.019	−0.006	0.000	0.999	0.013	0.001
10^{-2}	10	U	0.003	0.999	0.190	0.024	0.002	0.999	0.073	−0.001	−0.001	0.998	0.046	−0.004
		Exp	−0.004	0.992	0.202	0.064	−0.002	0.994	0.072	0.010	0.003	1.005	0.053	−0.001
	20	U	−0.007	1.003	0.145	0.020	−0.003	0.992	0.050	0.006	−0.002	1.005	0.048	−0.002
		Exp	0.001	1.002	0.150	0.048	0.002	1.011	0.025	0.015	−0.003	1.013	0.035	0.007
	50	U	0.002	1.004	0.079	0.017	−0.004	1.001	0.035	0.001	−0.003	0.999	0.011	−0.003
		Exp	0.000	0.998	0.091	0.025	−0.005	0.994	0.014	0.007	0.002	1.006	0.015	0.008
	100	U	−0.004	0.998	0.062	0.003	0.001	0.998	0.015	−0.020	0.000	0.999	0.016	0.015
		Exp	−0.002	0.998	0.058	−0.007	0.003	1.003	0.022	0.010	−0.003	1.002	0.021	0.006
	500	U	0.003	0.997	0.036	−0.015	−0.004	1.003	0.016	0.004	−0.004	1.004	0.011	0.026
		Exp	−0.003	1.000	0.025	−0.013	−0.001	0.995	0.018	−0.006	0.003	0.998	0.017	−0.019
	1000	U	0.001	0.995	0.008	0.038	−0.002	0.996	−0.001	0.006	−0.006	0.998	−0.002	−0.019
		Exp	0.004	1.003	0.010	0.004	0.000	1.000	0.002	−0.014	−0.003	0.999	0.008	−0.009
∞	10	U	0.000	0.997	0.241	0.058	−0.005	1.002	0.043	−0.003	−0.003	0.999	0.025	−0.006
		Exp	0.006	0.991	0.213	0.025	−0.002	0.998	0.076	0.007	0.001	1.005	0.022	0.009
	20	U	0.001	1.001	0.139	0.041	0.001	1.006	0.042	0.019	−0.001	1.006	0.006	−0.002
		Exp	0.001	1.009	0.128	0.029	−0.007	0.994	0.035	0.022	−0.002	0.999	0.012	0.005
	50	U	0.003	1.005	0.066	−0.011	−0.004	1.000	0.033	−0.012	−0.001	1.002	0.023	−0.006
		Exp	−0.003	0.995	0.089	−0.003	0.000	1.005	0.020	−0.021	−0.005	0.994	−0.005	0.004
	100	U	0.000	1.004	0.077	0.013	0.004	1.004	0.016	0.012	−0.001	0.993	0.011	−0.001
		Exp	−0.004	1.001	0.061	0.014	0.001	0.999	0.020	0.041	0.007	0.995	0.017	0.019
	500	U	−0.002	0.995	0.023	−0.033	−0.002	1.010	0.011	0.000	−0.005	0.999	−0.005	0.012
		Exp	−0.004	0.998	0.025	0.014	0.000	1.008	0.006	−0.013	−0.003	1.006	0.009	−0.007
	1000	U	0.003	0.988	0.010	−0.018	−0.002	1.000	0.008	−0.002	−0.005	0.997	0.006	0.023
		Exp	−0.001	0.997	0.029	−0.008	0.001	0.991	0.015	−0.010	0.004	0.997	0.011	−0.007

Table B.11.: *Mean, standard deviation, and square root of inverse Fisher information of $\hat{\xi}_m$ in the counting-model; based on 10^5 replications (cf. Section 5.5.2)*

				$(s_{j0},\ldots,s_{j4}) = (0, 4, 8, 12, \infty)$								
				$\beta = 1$			$\beta = 3$			$\beta = 5$		
ξ	μ	m	km	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	$\sqrt{J_\xi}$	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	$\sqrt{J_\xi}$	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	$\sqrt{J_\xi}$
0.5	10^{-4}	20	U	—	—	0.405	—	—	0.352	0.570	0.508	0.473
			Exp	—	—	0.488	0.546	0.430	0.379	0.587	0.566	0.507
		50	U	0.515	0.285	0.248	0.514	0.256	0.245	0.518	0.294	0.286
			Exp	—	—	0.260	0.512	0.242	0.231	0.518	0.308	0.300
		100	U	0.506	0.190	0.179	0.506	0.173	0.169	0.506	0.205	0.202
			Exp	0.508	0.208	0.194	0.506	0.167	0.163	0.507	0.207	0.203
	10^{-3}	20	U	0.503	0.125	0.122	0.503	0.123	0.122	0.503	0.156	0.154
			Exp	0.503	0.129	0.125	0.503	0.116	0.115	0.504	0.152	0.150
		50	U	0.501	0.079	0.078	0.501	0.073	0.073	0.501	0.087	0.086
			Exp	0.501	0.075	0.075	0.501	0.075	0.075	0.502	0.101	0.101
		100	U	0.501	0.056	0.056	0.500	0.052	0.052	0.501	0.063	0.063
			Exp	0.501	0.056	0.055	0.501	0.055	0.055	0.501	0.068	0.068
1	10^{-4}	20	U	—	—	0.569	—	—	0.601	—	—	0.726
			Exp	—	—	0.562	—	—	0.651	1.099	0.787	0.666
		50	U	1.050	0.408	0.356	1.030	0.366	0.345	1.033	0.461	0.434
			Exp	1.052	0.440	0.377	1.026	0.355	0.337	1.032	0.440	0.415
		100	U	1.019	0.260	0.247	1.014	0.253	0.247	1.016	0.302	0.294
			Exp	1.021	0.264	0.251	1.013	0.243	0.238	1.017	0.326	0.316
	10^{-3}	20	U	1.010	0.170	0.165	1.007	0.180	0.178	1.008	0.228	0.225
			Exp	1.008	0.163	0.159	1.006	0.169	0.167	1.010	0.241	0.236
		50	U	1.005	0.121	0.119	1.002	0.118	0.117	1.003	0.128	0.128
			Exp	1.004	0.111	0.110	1.002	0.104	0.103	1.002	0.124	0.124
		100	U	1.002	0.078	0.077	1.002	0.079	0.078	1.002	0.095	0.094
			Exp	1.002	0.075	0.075	1.001	0.080	0.080	1.001	0.090	0.090

				$(s_{j0},\ldots,s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$								
				$\beta = 1$			$\beta = 3$			$\beta = 5$		
ξ	μ	m	km	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	$\sqrt{J_\xi}$	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	$\sqrt{J_\xi}$	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	$\sqrt{J_\xi}$
0.5	10^{-4}	20	U	—	—	0.326	0.516	0.362	0.341	0.527	0.396	0.389
			Exp	—	—	0.278	0.512	0.306	0.292	0.525	0.381	0.374
		50	U	0.499	0.212	0.199	0.505	0.235	0.227	0.506	0.258	0.255
			Exp	0.497	0.210	0.198	0.503	0.222	0.215	0.505	0.232	0.229
		100	U	0.499	0.142	0.137	0.503	0.157	0.155	0.503	0.180	0.178
			Exp	0.500	0.144	0.140	0.501	0.157	0.155	0.503	0.204	0.201
	10^{-3}	20	U	0.499	0.087	0.087	0.501	0.111	0.110	0.502	0.155	0.153
			Exp	0.499	0.085	0.084	0.501	0.116	0.115	0.501	0.116	0.115
		50	U	0.500	0.063	0.063	0.500	0.064	0.064	0.500	0.083	0.083
			Exp	0.500	0.059	0.059	0.500	0.064	0.064	0.501	0.082	0.081
		100	U	0.500	0.044	0.044	0.500	0.046	0.045	0.500	0.058	0.058
			Exp	0.500	0.044	0.044	0.500	0.046	0.046	0.500	0.060	0.060
1	10^{-4}	20	U	1.047	0.524	0.434	1.032	0.476	0.440	1.042	0.619	0.570
			Exp	—	—	0.450	1.053	0.605	0.535	1.050	0.649	0.594
		50	U	1.018	0.308	0.291	1.014	0.307	0.296	1.018	0.404	0.387
			Exp	1.020	0.325	0.304	1.012	0.291	0.282	1.019	0.405	0.388
		100	U	1.008	0.209	0.203	1.007	0.224	0.220	1.006	0.261	0.257
			Exp	1.008	0.199	0.193	1.008	0.240	0.235	1.006	0.263	0.257
	10^{-3}	20	U	1.005	0.155	0.152	1.003	0.148	0.146	1.003	0.179	0.177
			Exp	1.003	0.133	0.131	1.003	0.159	0.158	1.004	0.188	0.187
		50	U	1.001	0.091	0.091	1.001	0.095	0.094	1.001	0.113	0.112
			Exp	1.002	0.088	0.088	1.001	0.099	0.099	1.002	0.124	0.124
		100	U	1.001	0.065	0.065	1.000	0.068	0.068	1.001	0.081	0.081
			Exp	1.001	0.055	0.055	1.001	0.072	0.072	1.001	0.089	0.089

Table B.12.: *Mean, standard deviation, and square root of inverse Fisher information of $\hat{\beta}_m$ in the counting-model; based on 10^5 replications (cf. Section 5.5.2)*

				$(s_{j0},\dots,s_{j4}) = (0,4,8,12,\infty)$								
				$\beta=1$			$\beta=3$			$\beta=5$		
ξ	μ	m	km	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	$\sqrt{J_\beta}$	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	$\sqrt{J_\beta}$	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	$\sqrt{J_\beta}$
0.5	10^{-4}	20	U	—	—	0.640	—	—	1.016	5.049	1.684	1.770
			Exp	—	—	0.771	3.051	1.083	1.096	5.037	1.797	1.901
		50	U	1.035	0.391	0.392	3.030	0.712	0.708	5.043	1.078	1.072
			Exp	—	—	0.412	3.025	0.676	0.668	5.053	1.119	1.125
		100	U	1.016	0.282	0.282	3.015	0.492	0.488	5.028	0.760	0.755
			Exp	1.019	0.307	0.307	3.014	0.475	0.471	5.027	0.771	0.761
	10^{-3}	20	U	1.007	0.193	0.192	3.008	0.352	0.351	5.021	0.585	0.578
			Exp	1.008	0.198	0.198	3.007	0.333	0.332	5.014	0.565	0.561
		50	U	1.003	0.124	0.123	3.002	0.211	0.211	5.005	0.323	0.323
			Exp	1.003	0.118	0.118	3.003	0.218	0.217	5.008	0.377	0.377
		100	U	1.001	0.088	0.089	3.001	0.150	0.150	5.002	0.237	0.237
			Exp	1.002	0.088	0.087	3.002	0.158	0.158	5.002	0.254	0.253
1	10^{-4}	20	U	—	—	0.711	—	—	1.418	—	—	2.257
			Exp	—	—	0.703	—	—	1.535	5.176	2.135	2.073
		50	U	1.027	0.437	0.445	3.037	0.829	0.813	5.088	1.390	1.349
			Exp	1.035	0.464	0.471	3.040	0.805	0.794	5.075	1.328	1.291
		100	U	1.014	0.307	0.309	3.021	0.585	0.582	5.037	0.926	0.913
			Exp	1.014	0.311	0.314	3.018	0.564	0.561	5.045	0.994	0.982
	10^{-3}	20	U	1.005	0.207	0.207	3.009	0.419	0.419	5.025	0.707	0.699
			Exp	1.006	0.199	0.199	3.007	0.395	0.395	5.029	0.743	0.735
		50	U	1.003	0.149	0.149	3.006	0.276	0.276	5.008	0.399	0.398
			Exp	1.002	0.137	0.137	3.002	0.244	0.244	5.008	0.386	0.386
		100	U	1.001	0.097	0.096	3.001	0.185	0.184	5.003	0.294	0.294
			Exp	1.001	0.094	0.094	3.003	0.189	0.188	5.005	0.279	0.280

				$(s_{j0},\dots,s_{j6}) = (0,2.4,4.8,7.2,9.6,12,\infty)$								
				$\beta=1$			$\beta=3$			$\beta=5$		
ξ	μ	m	km	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	$\sqrt{J_\beta}$	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	$\sqrt{J_\beta}$	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	$\sqrt{J_\beta}$
0.5	10^{-4}	20	U	—	—	0.434	3.092	0.954	0.939	5.094	1.403	1.434
			Exp	—	—	0.370	3.062	0.820	0.805	5.086	1.356	1.379
		50	U	1.022	0.269	0.265	3.047	0.641	0.627	5.065	0.947	0.938
			Exp	1.025	0.269	0.264	3.045	0.606	0.593	5.051	0.855	0.845
		100	U	1.011	0.185	0.183	3.019	0.430	0.426	5.031	0.663	0.656
			Exp	1.010	0.187	0.186	3.020	0.432	0.428	5.041	0.752	0.742
	10^{-3}	20	U	1.004	0.116	0.116	3.009	0.304	0.303	5.025	0.570	0.563
			Exp	1.004	0.112	0.112	3.012	0.319	0.319	5.015	0.427	0.424
		50	U	1.002	0.083	0.083	3.003	0.176	0.176	5.007	0.308	0.307
			Exp	1.002	0.079	0.078	3.004	0.176	0.176	5.006	0.300	0.300
		100	U	1.001	0.059	0.059	3.002	0.126	0.125	5.003	0.213	0.213
			Exp	1.001	0.058	0.058	3.002	0.126	0.126	5.003	0.219	0.220
1	10^{-4}	20	U	1.053	0.483	0.472	3.105	1.036	0.985	5.209	1.824	1.735
			Exp	—	—	0.490	3.149	1.283	1.198	5.215	1.908	1.807
		50	U	1.020	0.319	0.317	3.047	0.681	0.663	5.105	1.223	1.179
			Exp	1.023	0.334	0.331	3.041	0.646	0.631	5.093	1.218	1.180
		100	U	1.011	0.223	0.222	3.026	0.500	0.493	5.048	0.793	0.782
			Exp	1.009	0.212	0.210	3.028	0.534	0.526	5.044	0.796	0.783
	10^{-3}	20	U	1.005	0.166	0.166	3.012	0.330	0.327	5.021	0.544	0.538
			Exp	1.005	0.143	0.143	3.012	0.356	0.355	5.024	0.574	0.560
		50	U	1.003	0.099	0.099	3.004	0.212	0.212	5.008	0.342	0.342
			Exp	1.002	0.096	0.096	3.005	0.222	0.221	5.008	0.377	0.376
		100	U	1.001	0.070	0.070	3.003	0.152	0.151	5.004	0.248	0.247
			Exp	1.001	0.060	0.060	3.002	0.161	0.161	5.006	0.271	0.270

Table B.13.: *First four cumulants of* $(\hat{\xi}_m - \xi)/\sqrt{J_\xi}$ *in the counting-model; based on* 10^5 *replications (cf. Section 5.5.2)*

				$(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$							
				$\xi = 0.5$				$\xi = 1$			
β	μ	m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
1	10^{-4}	20	U	—	—	—	—	—	—	—	—
			Exp	—	—	—	—	—	—	—	—
		50	U	0.059	1.321	2.391	27.741	0.139	1.319	2.171	14.35
			Exp	—	—	—	—	0.139	1.363	2.933	39.716
		100	U	0.035	1.127	0.731	1.321	0.077	1.108	0.762	1.358
			Exp	0.039	1.150	0.819	1.741	0.083	1.112	0.777	1.244
	10^{-3}	20	U	0.026	1.054	0.384	0.361	0.060	1.052	0.445	0.449
			Exp	0.023	1.063	0.438	0.466	0.049	1.046	0.410	0.368
		50	U	0.009	1.020	0.250	0.146	0.038	1.026	0.297	0.186
			Exp	0.013	1.018	0.223	0.105	0.038	1.019	0.261	0.163
		100	U	0.012	1.005	0.166	0.076	0.025	1.015	0.189	0.085
			Exp	0.010	1.013	0.178	0.095	0.020	1.009	0.181	0.084
3	10^{-4}	20	U	—	—	—	—	—	—	—	—
			Exp	0.120	1.281	2.446	15.378	—	—	—	—
		50	U	0.057	1.093	0.765	1.113	0.087	1.128	0.836	1.420
			Exp	0.053	1.093	0.712	1.004	0.077	1.109	0.788	1.277
		100	U	0.037	1.049	0.449	0.427	0.057	1.053	0.504	0.484
			Exp	0.036	1.043	0.424	0.429	0.056	1.043	0.465	0.466
	10^{-3}	20	U	0.024	1.026	0.303	0.222	0.042	1.023	0.325	0.205
			Exp	0.026	1.022	0.280	0.134	0.039	1.024	0.317	0.180
		50	U	0.014	1.003	0.171	0.052	0.018	1.021	0.214	0.109
			Exp	0.013	1.004	0.171	0.056	0.023	1.015	0.194	0.081
		100	U	0.009	1.003	0.117	0.016	0.024	1.013	0.148	0.058
			Exp	0.010	1.003	0.122	0.028	0.012	1.009	0.151	0.032
5	10^{-4}	20	U	0.148	1.156	1.890	6.330	—	—	—	—
			Exp	0.171	1.244	2.878	21.389	0.148	1.393	3.364	42.124
		50	U	0.062	1.059	0.665	0.730	0.075	1.130	0.839	1.433
			Exp	0.059	1.051	0.732	0.860	0.076	1.121	0.792	1.338
		100	U	0.032	1.037	0.398	0.350	0.055	1.059	0.466	0.504
			Exp	0.037	1.043	0.396	0.308	0.053	1.064	0.494	0.504
	10^{-3}	20	U	0.018	1.024	0.280	0.176	0.037	1.027	0.324	0.232
			Exp	0.029	1.024	0.263	0.152	0.041	1.042	0.371	0.279
		50	U	0.014	1.005	0.144	0.040	0.021	1.005	0.180	0.062
			Exp	0.015	1.004	0.161	0.041	0.019	1.005	0.177	0.094
		100	U	0.012	1.002	0.119	0.049	0.020	1.006	0.119	0.055
			Exp	0.015	1.001	0.111	0.011	0.013	0.998	0.131	0.046

Table B.14.: *First four cumulants of $(\hat{\xi}_m - \xi)/\sqrt{J_\xi}$ in the counting-model; based on 10^5 replications (cf. Section 5.5.2)*

				$(s_{j0}, \ldots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$							
				$\xi = 0.5$				$\xi = 1$			
β	μ	m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
1	10^{-4}	20	U	—	—	—	—	0.108	1.459	3.854	51.532
			Exp	—	—	—	—	—	—	—	—
		50	U	−0.004	1.135	0.635	0.961	0.063	1.126	0.793	1.464
			Exp	−0.017	1.122	0.604	0.909	0.066	1.142	0.915	1.861
		100	U	−0.008	1.060	0.358	0.327	0.038	1.060	0.498	0.541
			Exp	−0.002	1.049	0.373	0.428	0.039	1.066	0.468	0.446
	10^{-3}	20	U	−0.008	1.012	0.202	0.143	0.031	1.034	0.350	0.281
			Exp	−0.006	1.018	0.218	0.125	0.022	1.022	0.279	0.167
		50	U	0.003	1.012	0.151	0.092	0.009	1.012	0.202	0.112
			Exp	0.002	1.008	0.137	0.072	0.018	1.007	0.182	0.063
		100	U	−0.005	0.998	0.105	0.047	0.013	1.003	0.133	0.013
			Exp	0.002	1.005	0.099	0.042	0.013	1.005	0.108	0.035
3	10^{-4}	20	U	0.046	1.127	1.253	3.349	0.072	1.171	1.088	2.486
			Exp	0.042	1.100	0.871	1.259	0.099	1.280	1.722	5.920
		50	U	0.020	1.071	0.599	0.693	0.047	1.076	0.580	0.702
			Exp	0.014	1.061	0.539	0.623	0.042	1.067	0.520	0.543
		100	U	0.016	1.028	0.349	0.284	0.030	1.037	0.395	0.310
			Exp	0.009	1.028	0.337	0.276	0.034	1.045	0.430	0.385
	10^{-3}	20	U	0.010	1.015	0.236	0.111	0.020	1.025	0.257	0.131
			Exp	0.010	1.009	0.231	0.117	0.017	1.014	0.260	0.144
		50	U	0.007	1.004	0.109	0.057	0.014	1.005	0.155	0.044
			Exp	0.003	0.998	0.131	0.044	0.014	1.004	0.168	0.066
		100	U	0.002	1.010	0.100	0.039	0.007	1.001	0.118	0.042
			Exp	0.004	1.002	0.107	0.044	0.013	1.004	0.124	0.041
5	10^{-4}	20	U	0.070	1.034	1.016	1.632	0.074	1.177	1.250	3.295
			Exp	0.068	1.039	0.951	1.413	0.085	1.195	1.331	3.418
		50	U	0.023	1.030	0.497	0.396	0.046	1.086	0.583	0.730
			Exp	0.022	1.027	0.414	0.286	0.050	1.088	0.624	0.907
		100	U	0.017	1.024	0.290	0.176	0.023	1.036	0.359	0.306
			Exp	0.017	1.024	0.342	0.218	0.024	1.042	0.355	0.301
	10^{-3}	20	U	0.014	1.023	0.242	0.143	0.017	1.022	0.223	0.091
			Exp	0.006	1.015	0.180	0.093	0.020	1.020	0.236	0.133
		50	U	0.004	1.001	0.143	0.064	0.010	1.003	0.149	0.065
			Exp	0.007	1.008	0.124	0.053	0.020	1.015	0.158	0.046
		100	U	0.005	1.002	0.078	−0.007	0.008	1.002	0.109	0.056
			Exp	0.003	1.002	0.101	0.028	0.011	1.003	0.101	0.020

Table B.15.: *First four cumulants of $(\hat{\beta}_m - \beta)/\sqrt{J_\beta}$ in the counting-model; based on 10^5 replications (cf. Section 5.5.2)*

				$(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$							
				$\xi = 0.5$				$\xi = 1$			
β	μ	m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
1	10^{-4}	20	U	—	—	—	—	—	—	—	—
			Exp	—	—	—	—	—	—	—	—
		50	U	0.088	0.993	0.100	−0.279	0.060	0.967	0.313	−0.043
			Exp	—	—	—	—	0.074	0.971	0.351	−0.047
		100	U	0.058	1.001	0.087	−0.095	0.047	0.987	0.202	−0.065
			Exp	0.063	1.002	0.102	−0.135	0.046	0.985	0.216	−0.030
	10^{-3}	20	U	0.038	1.004	0.059	−0.064	0.024	0.998	0.137	−0.011
			Exp	0.039	1.005	0.041	−0.052	0.028	0.996	0.140	−0.013
		50	U	0.028	1.005	0.024	−0.010	0.022	1.002	0.104	0.001
			Exp	0.023	0.997	0.028	−0.029	0.013	1.003	0.096	−0.006
		100	U	0.016	0.994	0.026	0.008	0.012	1.004	0.062	−0.010
			Exp	0.017	1.002	0.013	−0.015	0.016	1.001	0.057	−0.009
3	10^{-4}	20	U	—	—	—	—	—	—	—	—
			Exp	0.047	0.977	0.246	0.101	—	—	—	—
		50	U	0.042	1.012	0.211	0.090	0.045	1.038	0.343	0.252
			Exp	0.037	1.024	0.219	0.117	0.050	1.027	0.332	0.252
		100	U	0.031	1.019	0.167	0.123	0.036	1.011	0.227	0.123
			Exp	0.030	1.015	0.167	0.094	0.031	1.009	0.228	0.139
	10^{-3}	20	U	0.023	1.008	0.118	0.056	0.020	1.002	0.160	0.066
			Exp	0.022	1.005	0.108	0.025	0.018	1.000	0.131	0.036
		50	U	0.011	1.000	0.071	0.003	0.023	0.998	0.099	0.026
			Exp	0.015	1.003	0.070	0.003	0.009	0.999	0.078	0.017
		100	U	0.008	1.004	0.051	0.015	0.005	1.009	0.062	0.056
			Exp	0.011	1.008	0.059	0.005	0.017	1.006	0.063	−0.005
5	10^{-4}	20	U	0.028	0.905	0.245	0.090	—	—	—	—
			Exp	0.020	0.894	0.249	0.089	0.085	1.062	0.636	0.610
		50	U	0.040	1.012	0.279	0.055	0.065	1.062	0.489	0.485
			Exp	0.047	0.988	0.239	0.006	0.058	1.058	0.462	0.471
		100	U	0.037	1.014	0.240	0.118	0.041	1.029	0.328	0.253
			Exp	0.035	1.026	0.268	0.130	0.045	1.026	0.331	0.246
	10^{-3}	20	U	0.036	1.023	0.207	0.106	0.036	1.022	0.234	0.104
			Exp	0.025	1.011	0.178	0.081	0.039	1.022	0.241	0.126
		50	U	0.015	0.996	0.106	0.017	0.021	1.003	0.118	0.026
			Exp	0.021	0.999	0.127	0.024	0.021	1.004	0.122	0.038
		100	U	0.009	1.001	0.087	0.011	0.010	1.001	0.095	−0.002
			Exp	0.006	1.004	0.083	−0.002	0.017	0.993	0.080	0.013

Table B.16.: *First four cumulants of* $(\hat{\beta}_m - \beta)/\sqrt{J_\beta}$ *in the counting-model; based on* 10^5 *replications (cf. Section 5.5.2)*

				$(s_{j0},\ldots,s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$							
				$\xi = 0.5$				$\xi = 1$			
β	μ	m	km	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{k}_4$
1	10^{-4}	20	U	—	—	—	—	0.112	1.043	0.536	0.474
			Exp	—	—	—	—	—	—	—	—
		50	U	0.084	1.037	0.142	0.004	0.063	1.017	0.316	0.146
			Exp	0.095	1.037	0.158	0.009	0.071	1.017	0.332	0.167
		100	U	0.062	1.022	0.097	−0.002	0.051	1.012	0.209	0.089
			Exp	0.053	1.010	0.107	0.041	0.044	1.018	0.180	0.030
	10^{-3}	20	U	0.037	1.001	0.063	0.027	0.029	0.999	0.152	0.045
			Exp	0.036	0.997	0.057	0.001	0.032	1.000	0.131	0.014
		50	U	0.021	1.003	0.049	0.044	0.026	1.007	0.087	0.029
			Exp	0.022	1.004	0.056	0.017	0.021	0.999	0.080	0.030
		100	U	0.020	1.001	0.029	0.016	0.013	1.002	0.062	−0.016
			Exp	0.012	0.996	0.034	−0.003	0.011	1.002	0.061	0.012
3	10^{-4}	20	U	0.098	1.031	0.473	0.404	0.106	1.108	0.738	1.022
			Exp	0.077	1.037	0.438	0.317	0.124	1.148	0.973	1.611
		50	U	0.075	1.045	0.382	0.285	0.070	1.052	0.456	0.434
			Exp	0.076	1.043	0.368	0.282	0.065	1.047	0.412	0.353
		100	U	0.044	1.020	0.269	0.177	0.052	1.031	0.310	0.198
			Exp	0.047	1.017	0.275	0.233	0.054	0.031	0.334	0.261
	10^{-3}	20	U	0.030	1.011	0.186	0.049	0.036	1.020	0.209	0.097
			Exp	0.036	1.005	0.206	0.089	0.035	1.007	0.222	0.110
		50	U	0.020	1.009	0.099	0.002	0.018	1.001	0.120	0.023
			Exp	0.023	0.998	0.098	−0.001	0.023	1.002	0.134	0.055
		100	U	0.013	1.008	0.067	−0.018	0.018	1.002	0.087	0.012
			Exp	0.013	1.001	0.073	0.015	0.011	1.000	0.089	0.024
5	10^{-4}	20	U	0.065	0.957	0.398	0.190	0.120	1.106	0.805	0.935
			Exp	0.062	0.968	0.375	0.152	0.119	1.115	0.899	1.314
		50	U	0.069	1.018	0.364	0.170	0.089	1.076	0.618	0.725
			Exp	0.061	1.023	0.358	0.173	0.079	1.066	0.598	0.682
		100	U	0.047	1.022	0.302	0.186	0.062	1.030	0.363	0.260
			Exp	0.055	1.027	0.333	0.217	0.056	1.033	0.373	0.315
	10^{-3}	20	U	0.045	1.025	0.275	0.151	0.039	1.023	0.251	0.130
			Exp	0.035	1.016	0.202	0.086	0.042	1.022	0.267	0.144
		50	U	0.024	1.005	0.127	0.048	0.024	1.000	0.160	0.052
			Exp	0.021	1.003	0.129	0.043	0.021	1.008	0.164	0.030
		100	U	0.015	0.999	0.090	0.033	0.017	1.011	0.109	0.039
			Exp	0.015	0.997	0.087	0.018	0.022	1.004	0.123	0.053

Table B.17.: *Mean and standard deviation of $\hat{\xi}_m$ and the number of realizations with $\hat{\xi}_m = 0$ in case of $\xi = 0$ in the counting-model; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$								
			$\beta = 1$			$\beta = 3$			$\beta = 5$		
μ	m	km	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	#{$\hat{\xi}_m = 0$}	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	#{$\hat{\xi}_m = 0$}	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	#{$\hat{\xi}_m = 0$}
10^{-4}	20	U	—	—	68507	0.083	0.152	58153	0.116	0.196	53532
		Exp	—	—	65386	0.082	0.150	58060	0.115	0.195	53764
	50	U	—	—	84750	0.046	0.074	53819	0.079	0.126	52545
		Exp	—	—	87798	0.052	0.086	54118	0.067	0.107	52483
	100	U	—	—	90849	0.036	0.056	52805	0.046	0.070	51588
		Exp	—	—	90778	0.037	0.058	53240	0.053	0.082	51542
	500	U	0.025	0.043	65410	0.016	0.024	51402	0.022	0.032	50461
		Exp	0.025	0.042	64651	0.016	0.025	51443	0.022	0.033	50615
	1000	U	0.019	0.031	55471	0.011	0.017	51143	0.015	0.022	50635
		Exp	0.019	0.031	54708	0.011	0.017	50932	0.015	0.022	50422
10^{-3}	20	U	—	—	83956	0.025	0.039	52268	0.034	0.051	50809
		Exp	—	—	84074	0.022	0.034	52007	0.033	0.050	50943
	50	U	0.025	0.043	64912	0.016	0.024	51195	0.021	0.031	50581
		Exp	0.025	0.042	64474	0.017	0.027	51732	0.019	0.029	50678
	100	U	0.020	0.032	54257	0.011	0.016	51111	0.015	0.022	50272
		Exp	0.018	0.030	56212	0.011	0.017	50784	0.015	0.023	50354
	500	U	0.010	0.015	53853	0.005	0.007	50208	0.007	0.010	50315
		Exp	0.010	0.015	53872	0.005	0.008	50150	0.007	0.010	50184
	1000	U	0.007	0.011	52777	0.004	0.005	50496	0.005	0.007	50354
		Exp	0.007	0.011	52825	0.004	0.005	50754	0.005	0.007	50337
10^{-2}	20	U	0.013	0.020	55508	0.008	0.011	50442	0.010	0.015	50337
		Exp	0.016	0.025	57658	0.008	0.012	50919	0.011	0.017	50212
	50	U	0.010	0.015	53995	0.005	0.008	50524	0.007	0.010	50242
		Exp	0.009	0.015	53845	0.005	0.008	50452	0.006	0.009	50329
	100	U	0.007	0.011	52692	0.004	0.005	50205	0.005	0.007	49953
		Exp	0.007	0.011	52957	0.004	0.005	50125	0.005	0.007	50208
	500	U	0.003	0.005	51092	0.002	0.002	50079	0.002	0.003	50246
		Exp	0.003	0.005	51392	0.002	0.002	49906	0.002	0.003	50267
	1000	U	0.002	0.004	50988	0.001	0.002	49862	0.001	0.002	50055
		Exp	0.002	0.004	50960	0.001	0.002	50199	0.002	0.002	49830

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$								
			$\beta = 1$			$\beta = 3$			$\beta = 5$		
μ	m	km	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	#{$\hat{\xi}_m = 0$}	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	#{$\hat{\xi}_m = 0$}	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	#{$\hat{\xi}_m = 0$}
10^{-4}	20	U	—	—	73469	0.068	0.120	57306	0.112	0.189	54065
		Exp	—	—	72043	0.061	0.105	56482	0.098	0.162	53592
	50	U	0.041	0.075	61901	0.040	0.065	54274	0.063	0.100	52196
		Exp	0.041	0.075	61923	0.039	0.062	53825	0.065	0.103	52310
	100	U	0.030	0.052	59318	0.029	0.046	53078	0.044	0.068	51970
		Exp	0.030	0.051	58827	0.032	0.051	53073	0.044	0.068	51662
	500	U	0.015	0.024	54591	0.014	0.021	51240	0.020	0.029	50889
		Exp	0.015	0.023	54630	0.014	0.022	51308	0.020	0.031	50917
	1000	U	0.011	0.017	53320	0.010	0.015	50746	0.014	0.021	50675
		Exp	0.011	0.017	53191	0.010	0.015	51012	0.014	0.021	50251
10^{-3}	20	U	0.023	0.038	56892	0.023	0.035	52368	0.033	0.049	51261
		Exp	0.023	0.038	56677	0.022	0.034	52328	0.034	0.052	51242
	50	U	0.016	0.025	54663	0.013	0.020	51240	0.018	0.027	50528
		Exp	0.014	0.022	54483	0.013	0.019	51064	0.020	0.030	50506
	100	U	0.011	0.017	53261	0.010	0.015	51030	0.014	0.021	50287
		Exp	0.011	0.017	53273	0.011	0.016	51346	0.014	0.021	50801
	500	U	0.005	0.008	51227	0.004	0.007	50415	0.006	0.009	50373
		Exp	0.006	0.008	51210	0.004	0.007	50350	0.006	0.009	49953
	1000	U	0.004	0.006	50921	0.003	0.005	50374	0.004	0.007	50086
		Exp	0.004	0.006	50889	0.003	0.005	50443	0.004	0.007	50144
10^{-2}	20	U	0.008	0.012	52448	0.007	0.010	50584	0.010	0.014	50062
		Exp	0.008	0.011	52268	0.007	0.011	50818	0.011	0.017	50434
	50	U	0.006	0.009	51602	0.004	0.006	50522	0.006	0.009	50402
		Exp	0.005	0.008	51571	0.005	0.008	50564	0.006	0.010	50134
	100	U	0.004	0.005	50952	0.003	0.005	50356	0.005	0.007	50360
		Exp	0.004	0.005	51134	0.003	0.005	50324	0.004	0.006	49966
	500	U	0.002	0.003	50568	0.001	0.002	50077	0.002	0.003	50056
		Exp	0.002	0.002	50301	0.001	0.002	50080	0.002	0.003	50063
	1000	U	0.001	0.002	50138	0.001	0.001	50301	0.001	0.002	50219
		Exp	0.001	0.002	50482	0.001	0.001	50084	0.001	0.002	49985

Table B.18.: *Mean and standard deviation of $\hat{\beta}_m$ in case of $\xi = 0$ in the counting-model; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$					
			$\beta = 1$		$\beta = 3$		$\beta = 5$	
μ	m	km	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)
10^{-4}	20	U	—	—	2.760	0.593	4.613	0.950
		Exp	—	—	2.764	0.587	4.611	0.942
	50	U	—	—	2.861	0.335	4.717	0.677
		Exp	—	—	2.841	0.375	4.756	0.584
	100	U	—	—	2.890	0.260	4.828	0.402
		Exp	—	—	2.886	0.268	4.803	0.461
	500	U	0.947	0.098	2.949	0.118	4.915	0.196
		Exp	0.948	0.096	2.948	0.120	4.914	0.198
	1000	U	0.961	0.072	2.965	0.082	4.941	0.136
		Exp	0.961	0.072	2.963	0.084	4.941	0.135
10^{-3}	20	U	—	—	2.921	0.184	4.869	0.300
		Exp	—	—	2.931	0.162	4.872	0.298
	50	U	0.948	0.096	2.949	0.116	4.917	0.188
		Exp	0.948	0.096	2.945	0.129	4.924	0.176
	100	U	0.959	0.076	2.965	0.081	4.940	0.136
		Exp	0.963	0.069	2.965	0.082	4.939	0.139
	500	U	0.980	0.034	2.984	0.036	4.974	0.061
		Exp	0.980	0.034	2.984	0.038	4.972	0.063
	1000	U	0.985	0.025	2.989	0.026	4.981	0.044
		Exp	0.985	0.025	2.989	0.026	4.981	0.044
10^{-2}	20	U	0.973	0.046	2.975	0.056	4.960	0.091
		Exp	0.967	0.058	2.974	0.061	4.955	0.102
	50	U	0.980	0.034	2.983	0.039	4.973	0.062
		Exp	0.980	0.034	2.984	0.038	4.976	0.054
	100	U	0.985	0.024	2.989	0.026	4.981	0.043
		Exp	0.985	0.025	2.988	0.026	4.980	0.046
	500	U	0.993	0.012	2.995	0.011	4.991	0.019
		Exp	0.993	0.012	2.995	0.012	4.991	0.020
	1000	U	0.995	0.008	2.996	0.008	4.994	0.014
		Exp	0.995	0.008	2.996	0.008	4.994	0.014

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$					
			$\beta = 1$		$\beta = 3$		$\beta = 5$	
μ	m	km	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)
10^{-4}	20	U	—	—	2.821	0.515	4.649	0.960
		Exp	—	—	2.834	0.462	4.686	0.849
	50	U	0.939	0.152	2.888	0.304	4.781	0.561
		Exp	0.938	0.151	2.894	0.290	4.777	0.579
	100	U	0.954	0.107	2.917	0.221	4.842	0.402
		Exp	0.956	0.104	2.909	0.241	4.842	0.398
	500	U	0.977	0.049	2.960	0.105	4.926	0.182
		Exp	0.978	0.048	2.958	0.108	4.923	0.188
	1000	U	0.984	0.035	2.971	0.074	4.947	0.129
		Exp	0.984	0.036	2.971	0.075	4.946	0.130
10^{-3}	20	U	0.966	0.077	2.935	0.170	4.879	0.297
		Exp	0.966	0.078	2.938	0.166	4.874	0.311
	50	U	0.977	0.052	2.961	0.101	4.931	0.168
		Exp	0.979	0.046	2.963	0.096	4.924	0.187
	100	U	0.984	0.034	2.971	0.075	4.945	0.131
		Exp	0.984	0.035	2.969	0.080	4.946	0.132
	500	U	0.992	0.016	2.987	0.033	4.976	0.059
		Exp	0.992	0.016	2.987	0.033	4.976	0.058
	1000	U	0.995	0.011	2.991	0.024	4.983	0.041
		Exp	0.994	0.011	2.991	0.023	4.983	0.042
10^{-2}	20	U	0.989	0.024	2.981	0.050	4.964	0.088
		Exp	0.989	0.023	2.979	0.054	4.957	0.104
	50	U	0.992	0.017	2.987	0.032	4.976	0.058
		Exp	0.992	0.016	2.984	0.041	4.975	0.060
	100	U	0.995	0.011	2.991	0.024	4.982	0.043
		Exp	0.995	0.011	2.991	0.023	4.984	0.038
	500	U	0.997	0.005	2.996	0.011	4.992	0.018
		Exp	0.998	0.005	2.996	0.010	4.992	0.019
	1000	U	0.998	0.004	2.997	0.007	4.995	0.013
		Exp	0.998	0.004	2.997	0.007	4.995	0.013

Table B.19.: *Quantiles of $\hat{\xi}_m$ in case of $\xi = 0$ and the inverse of Φ_ξ; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$									
			$p = 0.5$		$p = 0.75$		$p = 0.9$		$p = 0.95$		$p = 0.99$	
β	μ	m	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$
1	10^{-4}	20	0	0	0	0.273	0	0.519	0	0.666	0.821	0.942
		50	0	0	0	0.204	0	0.387	0	0.497	0.503	0.703
		100	0	0	0	0.139	0	0.263	0.208	0.338	0.393	0.478
		500	0	0	0.044	0.062	0.089	0.118	0.121	0.151	0.175	0.214
	10^{-3}	20	0	0	0	0.098	0.127	0.186	0.185	0.238	0.284	0.337
		50	0	0	0.044	0.061	0.088	0.116	0.116	0.149	0.176	0.211
		100	0	0	0.030	0.046	0.070	0.087	0.092	0.112	0.134	0.159
		500	0	0	0.016	0.019	0.032	0.036	0.042	0.046	0.058	0.065
	10^{-2}	20	0	0	0.021	0.027	0.044	0.051	0.057	0.066	0.080	0.093
		50	0	0	0.016	0.019	0.032	0.036	0.041	0.046	0.058	0.065
		100	0	0	0.012	0.013	0.023	0.025	0.030	0.032	0.042	0.046
		500	0	0	0.006	0.006	0.011	0.012	0.014	0.015	0.020	0.021
3	10^{-4}	20	0	0	0.115	0.141	0.283	0.268	0.397	0.343	0.672	0.486
		50	0	0	0.072	0.078	0.152	0.149	0.204	0.191	0.315	0.271
		100	0	0	0.057	0.061	0.117	0.116	0.155	0.149	0.233	0.210
		500	0	0	0.026	0.027	0.052	0.052	0.069	0.067	0.099	0.094
	10^{-3}	20	0	0	0.041	0.043	0.083	0.082	0.109	0.105	0.159	0.148
		50	0	0	0.026	0.027	0.052	0.051	0.067	0.066	0.098	0.093
		100	0	0	0.018	0.019	0.036	0.036	0.047	0.046	0.066	0.065
		500	0	0	0.008	0.008	0.016	0.016	0.021	0.021	0.030	0.029
	10^{-2}	20	0	0	0.013	0.013	0.025	0.025	0.032	0.032	0.046	0.045
		50	0	0	0.009	0.009	0.017	0.017	0.022	0.022	0.032	0.031
		100	0	0	0.006	0.006	0.011	0.011	0.015	0.015	0.021	0.021
		500	0	0	0.003	0.003	0.005	0.005	0.006	0.006	0.009	0.009
5	10^{-4}	20	0	0	0.173	0.182	0.381	0.345	0.526	0.443	0.848	0.626
		50	0	0	0.123	0.129	0.258	0.245	0.348	0.314	0.535	0.444
		100	0	0	0.075	0.076	0.149	0.144	0.196	0.184	0.293	0.261
		500	0	0	0.036	0.036	0.070	0.069	0.091	0.089	0.131	0.125
	10^{-3}	20	0	0	0.056	0.056	0.110	0.107	0.143	0.137	0.210	0.193
		50	0	0	0.034	0.035	0.067	0.066	0.087	0.085	0.126	0.120
		100	0	0	0.025	0.025	0.048	0.048	0.062	0.062	0.089	0.087
		500	0	0	0.011	0.011	0.021	0.021	0.028	0.027	0.039	0.039
	10^{-2}	20	0	0	0.017	0.017	0.032	0.032	0.041	0.041	0.059	0.058
		50	0	0	0.011	0.011	0.022	0.022	0.028	0.028	0.040	0.040
		100	0.000	0	0.008	0.008	0.015	0.015	0.019	0.019	0.027	0.027
		500	0	0	0.004	0.004	0.007	0.007	0.009	0.009	0.012	0.012

Table B.20.: *Quantiles of $\hat{\xi}_m$ in case of $\xi = 0$ and the inverse of Φ_ξ; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$									
			$p = 0.5$		$p = 0.75$		$p = 0.9$		$p = 0.95$		$p = 0.99$	
β	μ	m	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$	p-Q$(\hat{\xi}_m)$	$\Phi_\xi^{-1}(p)$
1	10^{-4}	20	0	0	0.006	0.154	0.222	0.292	0.323	0.375	0.527	0.531
		50	0	0	0.057	0.095	0.148	0.180	0.204	0.231	0.323	0.327
		100	0	0	0.046	0.067	0.108	0.127	0.145	0.163	0.218	0.230
		500	0	0	0.025	0.030	0.052	0.057	0.068	0.073	0.098	0.104
	10^{-3}	20	0	0	0.036	0.047	0.079	0.090	0.105	0.116	0.156	0.163
		50	0	0	0.026	0.032	0.054	0.060	0.071	0.077	0.103	0.109
		100	0	0	0.018	0.020	0.036	0.039	0.047	0.050	0.067	0.070
		500	0	0	0.009	0.009	0.017	0.018	0.022	0.023	0.031	0.032
	10^{-2}	20	0	0	0.013	0.014	0.026	0.027	0.033	0.035	0.047	0.049
		50	0	0	0.010	0.010	0.019	0.020	0.024	0.025	0.034	0.035
		100	0	0	0.006	0.006	0.012	0.012	0.015	0.015	0.021	0.021
		500	0	0	0.003	0.003	0.006	0.006	0.007	0.007	0.010	0.010
3	10^{-4}	20	0	0	0.100	0.123	0.236	0.234	0.325	0.301	0.518	0.425
		50	0	0	0.063	0.071	0.134	0.136	0.180	0.174	0.272	0.246
		100	0	0	0.047	0.052	0.097	0.098	0.129	0.126	0.189	0.179
		500	0	0	0.023	0.024	0.046	0.046	0.060	0.059	0.086	0.084
	10^{-3}	20	0	0	0.037	0.040	0.075	0.076	0.098	0.097	0.144	0.137
		50	0	0	0.022	0.023	0.044	0.044	0.057	0.057	0.082	0.081
		100	0	0	0.017	0.017	0.032	0.032	0.042	0.042	0.060	0.059
		500	0	0	0.008	0.008	0.015	0.015	0.019	0.019	0.027	0.026
	10^{-2}	20	0	0	0.011	0.011	0.022	0.022	0.028	0.028	0.040	0.039
		50	0	0	0.007	0.007	0.014	0.014	0.018	0.018	0.026	0.025
		100	0	0	0.005	0.005	0.010	0.010	0.013	0.013	0.019	0.019
		500	0	0	0.002	0.002	0.005	0.005	0.006	0.006	0.008	0.008
5	10^{-4}	20	0	0	0.167	0.185	0.373	0.352	0.515	0.451	0.819	0.638
		50	0	0	0.101	0.106	0.208	0.201	0.277	0.258	0.417	0.365
		100	0	0	0.071	0.075	0.145	0.143	0.191	0.183	0.282	0.259
		500	0	0	0.033	0.033	0.064	0.063	0.083	0.081	0.119	0.115
	10^{-3}	20	0	0	0.054	0.055	0.106	0.105	0.138	0.134	0.201	0.190
		50	0	0	0.031	0.031	0.060	0.059	0.077	0.076	0.110	0.107
		100	0	0	0.024	0.024	0.046	0.046	0.059	0.059	0.086	0.083
		500	0	0	0.011	0.011	0.020	0.020	0.026	0.026	0.037	0.037
	10^{-2}	20	0	0	0.016	0.016	0.031	0.031	0.040	0.039	0.057	0.055
		50	0	0	0.010	0.011	0.020	0.020	0.026	0.026	0.036	0.036
		100	0	0	0.008	0.008	0.015	0.015	0.019	0.019	0.027	0.027
		500	0	0	0.003	0.003	0.006	0.006	0.008	0.008	0.011	0.012

Table B.21.: *Lower quantiles of $\hat{\beta}_m$ in case of $\xi = 0$ and inverse of Φ_β; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$									
			$p = 0.01$		$p = 0.05$		$p = 0.1$		$p = 0.25$		$p = 0.5$	
β	μ	m	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$
1	10^{-4}	20	0.200	−1.021	0.910	−0.429	0.924	−0.113	0.948	0.409	0.998	0.799
		50	0.294	−0.510	0.813	−0.067	0.827	0.168	0.851	0.558	0.983	0.850
		100	0.335	−0.026	0.702	0.274	0.764	0.435	0.890	0.700	0.968	0.898
		500	0.634	0.541	0.745	0.676	0.809	0.747	0.908	0.866	0.969	0.954
	10^{-3}	20	0.481	0.277	0.654	0.489	0.777	0.602	0.897	0.788	0.971	0.928
		50	0.637	0.547	0.751	0.680	0.813	0.750	0.908	0.867	0.969	0.955
		100	0.719	0.660	0.807	0.759	0.853	0.813	0.925	0.900	0.974	0.966
		500	0.874	0.860	0.911	0.901	0.931	0.923	0.965	0.959	0.988	0.986
	10^{-2}	20	0.828	0.801	0.879	0.859	0.906	0.890	0.952	0.942	0.985	0.980
		50	0.876	0.860	0.912	0.901	0.932	0.923	0.965	0.959	0.989	0.986
		100	0.910	0.901	0.936	0.930	0.950	0.946	0.974	0.971	0.992	0.990
		500	0.956	0.954	0.969	0.968	0.976	0.975	0.987	0.987	0.996	0.995
3	10^{-4}	20	1.193	1.098	1.738	1.648	2.011	1.926	2.399	2.341	2.781	2.716
		50	1.970	1.940	2.277	2.246	2.429	2.402	2.656	2.633	2.880	2.842
		100	2.203	2.177	2.435	2.415	2.551	2.536	2.731	2.715	2.904	2.877
		500	2.634	2.630	2.741	2.737	2.794	2.791	2.877	2.872	2.957	2.945
	10^{-3}	20	2.435	2.419	2.595	2.587	2.679	2.672	2.808	2.799	2.932	2.913
		50	2.641	2.635	2.746	2.741	2.797	2.794	2.878	2.874	2.957	2.946
		100	2.752	2.746	2.822	2.819	2.859	2.857	2.915	2.912	2.970	2.962
		500	2.886	2.886	2.920	2.919	2.937	2.935	2.962	2.960	2.987	2.983
	10^{-2}	20	2.825	2.823	2.875	2.874	2.901	2.900	2.941	2.939	2.979	2.974
		50	2.879	2.877	2.913	2.913	2.931	2.931	2.959	2.958	2.986	2.982
		100	2.919	2.919	2.943	2.942	2.955	2.954	2.973	2.972	2.990	2.988
		500	2.965	2.965	2.975	2.975	2.980	2.980	2.988	2.988	2.996	2.995
5	10^{-4}	20	2.255	1.902	3.010	2.798	3.407	3.253	4.015	3.929	4.625	4.539
		50	3.004	2.803	3.560	3.439	3.846	3.761	4.289	4.241	4.743	4.673
		100	3.791	3.710	4.126	4.083	4.305	4.273	4.578	4.554	4.848	4.808
		500	4.398	4.381	4.569	4.560	4.657	4.651	4.794	4.786	4.928	4.908
	10^{-3}	20	4.075	4.044	4.345	4.320	4.479	4.461	4.683	4.669	4.887	4.858
		50	4.423	4.405	4.584	4.577	4.670	4.664	4.802	4.794	4.929	4.911
		100	4.582	4.570	4.699	4.694	4.762	4.757	4.857	4.851	4.950	4.936
		500	4.811	4.808	4.865	4.864	4.894	4.892	4.937	4.934	4.978	4.971
	10^{-2}	20	4.718	4.715	4.800	4.798	4.842	4.839	4.905	4.902	4.967	4.958
		50	4.805	4.804	4.862	4.861	4.891	4.890	4.935	4.932	4.977	4.971
		100	4.867	4.866	4.905	4.905	4.925	4.924	4.955	4.954	4.984	4.980
		500	4.939	4.939	4.957	4.957	4.966	4.966	4.980	4.979	4.993	4.991

Table B.22.: *Lower quantiles of $\hat{\beta}_m$ in case of $\xi = 0$ and inverse of Φ_β; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$									
			$p = 0.01$		$p = 0.05$		$p = 0.1$		$p = 0.25$		$p = 0.5$	
β	μ	m	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$
1	10^{-4}	20	0.320	0.127	0.576	0.382	0.634	0.515	0.788	0.713	0.929	0.879
		50	0.503	0.462	0.664	0.619	0.743	0.701	0.851	0.823	0.954	0.926
		100	0.647	0.621	0.759	0.732	0.813	0.790	0.893	0.875	0.966	0.948
		500	0.837	0.830	0.887	0.879	0.912	0.905	0.949	0.944	0.982	0.976
	10^{-3}	20	0.744	0.731	0.824	0.810	0.865	0.851	0.922	0.912	0.974	0.963
		50	0.829	0.821	0.881	0.873	0.908	0.901	0.947	0.941	0.982	0.975
		100	0.889	0.885	0.922	0.918	0.938	0.936	0.964	0.962	0.988	0.984
		500	0.948	0.947	0.963	0.962	0.971	0.971	0.983	0.983	0.994	0.993
	10^{-2}	20	0.922	0.919	0.945	0.943	0.957	0.955	0.975	0.973	0.991	0.989
		50	0.943	0.942	0.960	0.959	0.969	0.968	0.982	0.981	0.994	0.992
		100	0.965	0.965	0.975	0.975	0.981	0.980	0.989	0.988	0.996	0.995
		500	0.983	0.983	0.988	0.988	0.991	0.991	0.995	0.994	0.998	0.998
3	10^{-4}	20	1.576	1.384	1.972	1.844	2.171	2.074	2.483	2.422	2.820	2.748
		50	2.134	2.063	2.375	2.330	2.498	2.464	2.692	2.665	2.896	2.854
		100	2.362	2.322	2.541	2.515	2.632	2.611	2.776	2.758	2.925	2.894
		500	2.693	2.681	2.780	2.772	2.825	2.818	2.894	2.886	2.964	2.950
	10^{-3}	20	2.507	2.478	2.644	2.627	2.715	2.701	2.827	2.813	2.941	2.919
		50	2.707	2.693	2.788	2.780	2.830	2.824	2.897	2.890	2.965	2.952
		100	2.779	2.776	2.842	2.840	2.874	2.872	2.924	2.920	2.974	2.965
		500	2.900	2.900	2.929	2.928	2.944	2.943	2.966	2.964	2.989	2.984
	10^{-2}	20	2.852	2.850	2.895	2.893	2.916	2.914	2.949	2.947	2.983	2.977
		50	2.904	2.903	2.931	2.931	2.945	2.945	2.967	2.965	2.989	2.985
		100	2.929	2.929	2.950	2.949	2.960	2.959	2.976	2.975	2.992	2.989
		500	2.969	2.968	2.977	2.977	2.982	2.982	2.989	2.989	2.996	2.995
5	10^{-4}	20	2.427	1.877	3.091	2.773	3.445	3.224	4.020	3.900	4.634	4.523
		50	3.405	3.213	3.837	3.726	4.063	3.984	4.416	4.371	4.791	4.727
		100	3.835	3.733	4.154	4.096	4.320	4.279	4.586	4.554	4.855	4.806
		500	4.458	4.439	4.612	4.600	4.690	4.681	4.812	4.803	4.935	4.914
	10^{-3}	20	4.129	4.070	4.367	4.337	4.494	4.471	4.691	4.673	4.892	4.858
		50	4.498	4.477	4.638	4.627	4.711	4.702	4.825	4.816	4.939	4.920
		100	4.603	4.595	4.716	4.711	4.774	4.770	4.864	4.857	4.953	4.938
		500	4.821	4.819	4.873	4.871	4.899	4.897	4.939	4.936	4.979	4.972
	10^{-2}	20	4.736	4.729	4.810	4.807	4.849	4.846	4.908	4.904	4.968	4.959
		50	4.825	4.822	4.875	4.873	4.901	4.899	4.940	4.937	4.979	4.973
		100	4.870	4.869	4.908	4.906	4.926	4.925	4.956	4.954	4.985	4.980
		500	4.944	4.944	4.960	4.960	4.968	4.968	4.981	4.980	4.993	4.991

Table B.23.: *Upper quantiles of $\hat{\beta}_m$ in case of $\xi = 0$ and inverse of Φ_β; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$									
			$p = 0.5$		$p = 0.75$		$p = 0.9$		$p = 0.95$		$p = 0.99$	
β	μ	m	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$	p-Q($\hat{\beta}_m$)	$\Phi_\beta^{-1}(p)$
1	10^{-4}	20	0.998	0.799	1.154	1	1.270	1.188	1.335	1.286	1.468	1.458
		50	0.983	0.850	1.092	1	1.185	1.140	1.239	1.213	1.335	1.342
		100	0.968	0.898	1.052	1	1.115	1.095	1.154	1.145	1.222	1.232
		500	0.969	0.954	1.012	1	1.046	1.043	1.066	1.065	1.105	1.104
	10^{-3}	20	0.971	0.928	1.030	1	1.079	1.067	1.106	1.102	1.158	1.164
		50	0.969	0.955	1.011	1	1.046	1.042	1.065	1.064	1.104	1.103
		100	0.974	0.966	1.011	1	1.038	1.032	1.052	1.048	1.077	1.077
		500	0.988	0.986	1.004	1	1.015	1.013	1.021	1.020	1.032	1.032
	10^{-2}	20	0.985	0.980	1.005	1	1.021	1.018	1.030	1.028	1.045	1.045
		50	0.989	0.986	1.004	1	1.015	1.013	1.021	1.020	1.032	1.032
		100	0.992	0.990	1.002	1	1.010	1.009	1.015	1.014	1.022	1.022
		500	0.996	0.995	1.001	1	1.005	1.004	1.007	1.006	1.010	1.010
3	10^{-4}	20	2.781	2.716	3.153	3	3.482	3.394	3.691	3.601	4.096	3.963
		50	2.880	2.842	3.086	3	3.269	3.220	3.377	3.335	3.590	3.536
		100	2.904	2.877	3.067	3	3.207	3.171	3.290	3.260	3.447	3.416
		500	2.957	2.945	3.030	3	3.092	3.077	3.128	3.117	3.196	3.187
	10^{-3}	20	2.932	2.913	3.046	3	3.145	3.121	3.203	3.184	3.311	3.294
		50	2.957	2.946	3.028	3	3.090	3.076	3.126	3.115	3.192	3.184
		100	2.970	2.962	3.020	3	3.063	3.053	3.088	3.080	3.134	3.129
		500	2.987	2.983	3.009	3	3.028	3.024	3.039	3.036	3.059	3.058
	10^{-2}	20	2.979	2.974	3.014	3	3.044	3.037	3.061	3.056	3.093	3.089
		50	2.986	2.982	3.010	3	3.031	3.025	3.043	3.039	3.065	3.062
		100	2.990	2.988	3.006	3	3.020	3.017	3.028	3.026	3.043	3.041
		500	2.996	2.995	3.003	3	3.009	3.007	3.012	3.011	3.019	3.018
5	10^{-4}	20	4.625	4.539	5.234	5	5.771	5.637	6.141	5.970	6.859	6.555
		50	4.743	4.673	5.167	5	5.554	5.452	5.779	5.688	6.250	6.103
		100	4.848	4.808	5.101	5	5.321	5.265	5.452	5.404	5.705	5.648
		500	4.928	4.908	5.049	5	5.153	5.127	5.215	5.194	5.331	5.311
	10^{-3}	20	4.887	4.858	5.073	5	5.235	5.197	5.331	5.300	5.510	5.480
		50	4.929	4.911	5.046	5	5.147	5.123	5.206	5.187	5.317	5.299
		100	4.950	4.936	5.033	5	5.105	5.089	5.147	5.135	5.224	5.216
		500	4.978	4.971	5.015	5	5.047	5.039	5.066	5.060	5.100	5.096
	10^{-2}	20	4.967	4.958	5.023	5	5.071	5.059	5.098	5.089	5.150	5.143
		50	4.977	4.971	5.016	5	5.048	5.040	5.067	5.061	5.102	5.098
		100	4.984	4.980	5.010	5	5.033	5.028	5.046	5.042	5.070	5.067
		500	4.993	4.991	5.005	5	5.015	5.013	5.021	5.019	5.032	5.031

Table B.24.: *Upper quantiles of $\hat{\beta}_m$ in case of $\xi = 0$ and inverse of Φ_β; based on 10^5 replications (cf. Section 5.5.3)*

			$\xi = 0$ and $(s_{j0}, \ldots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$									
			$p = 0.5$		$p = 0.75$		$p = 0.9$		$p = 0.95$		$p = 0.99$	
β	μ	m	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$	p-Q$(\hat{\beta}_m)$	$\Phi_\beta^{-1}(p)$
1	10^{-4}	20	0.929	0.879	1.066	1	1.196	1.153	1.265	1.232	1.401	1.372
		50	0.954	0.926	1.039	1	1.118	1.094	1.162	1.143	1.245	1.229
		100	0.966	0.948	1.028	1	1.081	1.066	1.112	1.101	1.169	1.162
		500	0.982	0.976	1.011	1	1.036	1.030	1.050	1.045	1.075	1.073
	10^{-3}	20	0.974	0.963	1.019	1	1.057	1.047	1.079	1.072	1.119	1.115
		50	0.982	0.975	1.012	1	1.038	1.031	1.052	1.048	1.080	1.076
		100	0.988	0.984	1.008	1	1.024	1.020	1.034	1.031	1.051	1.049
		500	0.994	0.993	1.003	1	1.011	1.009	1.015	1.014	1.023	1.023
	10^{-2}	20	0.991	0.989	1.005	1	1.017	1.014	1.023	1.021	1.036	1.034
		50	0.994	0.992	1.004	1	1.012	1.010	1.017	1.015	1.026	1.025
		100	0.996	0.995	1.002	1	1.007	1.006	1.010	1.009	1.015	1.015
		500	0.998	0.998	1.001	1	1.003	1.003	1.005	1.005	1.007	1.007
3	10^{-4}	20	2.820	2.748	3.160	3	3.462	3.371	3.658	3.564	4.047	3.904
		50	2.896	2.854	3.093	3	3.269	3.215	3.374	3.327	3.579	3.524
		100	2.925	2.894	3.066	3	3.192	3.156	3.267	3.237	3.410	3.380
		500	2.964	2.950	3.031	3	3.089	3.073	3.124	3.111	3.189	3.178
	10^{-3}	20	2.941	2.919	3.051	3	3.147	3.120	3.204	3.182	3.314	3.292
		50	2.965	2.952	3.030	3	3.086	3.070	3.120	3.107	3.184	3.172
		100	2.974	2.965	3.022	3	3.063	3.051	3.087	3.078	3.132	3.125
		500	2.989	2.984	3.010	3	3.028	3.023	3.039	3.035	3.059	3.056
	10^{-2}	20	2.983	2.977	3.015	3	3.042	3.034	3.058	3.052	3.087	3.084
		50	2.989	2.985	3.009	3	3.027	3.022	3.037	3.034	3.057	3.054
		100	2.992	2.989	3.007	3	3.020	3.016	3.028	3.025	3.041	3.040
		500	2.996	2.995	3.003	3	3.009	3.007	3.012	3.011	3.019	3.018
5	10^{-4}	20	4.634	4.523	5.260	5	5.860	5.682	6.251	6.039	7.016	6.664
		50	4.791	4.727	5.154	5	5.485	5.390	5.683	5.594	6.073	5.953
		100	4.855	4.806	5.113	5	5.341	5.277	5.480	5.421	5.746	5.675
		500	4.935	4.914	5.050	5	5.149	5.122	5.209	5.186	5.322	5.299
	10^{-3}	20	4.892	4.858	5.081	5	5.245	5.203	5.344	5.309	5.536	5.496
		50	4.939	4.920	5.046	5	5.139	5.114	5.194	5.174	5.297	5.279
		100	4.953	4.938	5.035	5	5.106	5.088	5.149	5.135	5.224	5.216
		500	4.979	4.972	5.016	5	5.048	5.040	5.066	5.060	5.101	5.097
	10^{-2}	20	4.968	4.959	5.024	5	5.072	5.059	5.100	5.090	5.153	5.145
		50	4.979	4.973	5.016	5	5.047	5.039	5.065	5.059	5.099	5.095
		100	4.985	4.980	5.012	5	5.035	5.029	5.048	5.044	5.073	5.070
		500	4.993	4.991	5.005	5	5.015	5.012	5.021	5.019	5.031	5.030

Table B.25.: *Mean and standard deviation of $\hat{\xi}_m$ in the counting-maximum model; based on 10^4 replications (cf. Section 5.6.2)*

				$(s_{j0}, \ldots, s_{j4}) = (0, 4, 8, 12, \infty)$					
				$\beta = 1$		$\beta = 3$		$\beta = 5$	
ξ	μ	m	km	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)
0	10^{-4}	20	U	0.039	0.075	0.043	0.083	0.042	0.078
			Exp	0.039	0.079	0.045	0.083	0.041	0.079
		50	U	0.031	0.055	0.027	0.047	0.032	0.054
			Exp	0.030	0.053	0.032	0.054	0.029	0.048
		100	U	0.023	0.038	0.022	0.036	0.020	0.033
			Exp	0.023	0.039	0.023	0.038	0.023	0.038
	10^{-3}	20	U	0.023	0.040	0.017	0.028	0.017	0.026
			Exp	0.024	0.042	0.016	0.024	0.017	0.027
		50	U	0.016	0.026	0.012	0.018	0.011	0.017
			Exp	0.016	0.026	0.013	0.020	0.010	0.015
		100	U	0.012	0.020	0.008	0.012	0.008	0.012
			Exp	0.012	0.019	0.008	0.013	0.008	0.012
0.5	10^{-4}	20	U	0.436	0.257	0.456	0.226	0.458	0.237
			Exp	0.410	0.296	0.452	0.245	0.450	0.258
		50	U	0.472	0.158	0.479	0.156	0.483	0.143
			Exp	0.473	0.169	0.481	0.145	0.482	0.150
		100	U	0.486	0.113	0.490	0.105	0.493	0.099
			Exp	0.484	0.122	0.491	0.102	0.492	0.101
	10^{-3}	20	U	0.494	0.098	0.498	0.082	0.497	0.082
			Exp	0.494	0.104	0.498	0.077	0.497	0.080
		50	U	0.499	0.064	0.498	0.048	0.499	0.046
			Exp	0.497	0.061	0.500	0.050	0.497	0.053
		100	U	0.499	0.046	0.499	0.034	0.500	0.034
			Exp	0.499	0.045	0.500	0.036	0.501	0.036
1	10^{-4}	20	U	0.933	0.349	0.951	0.343	0.951	0.338
			Exp	0.936	0.347	0.944	0.372	0.963	0.316
		50	U	0.977	0.215	0.982	0.193	0.984	0.198
			Exp	0.973	0.228	0.990	0.190	0.986	0.192
		100	U	0.988	0.148	0.992	0.137	0.992	0.133
			Exp	0.988	0.153	0.994	0.132	0.988	0.144
	10^{-3}	20	U	0.998	0.121	0.999	0.109	0.998	0.116
			Exp	0.999	0.118	1.000	0.107	0.999	0.124
		50	U	0.999	0.086	1.000	0.072	0.999	0.066
			Exp	0.999	0.081	1.000	0.064	1.000	0.065
		100	U	1.000	0.056	1.000	0.048	1.000	0.049
			Exp	1.000	0.055	1.000	0.050	0.999	0.048

Table B.26.: *Mean and standard deviation of $\hat{\xi}_m$ in the counting-maximum model; based on 10^4 replications (cf. Section 5.6.2)*

				$(s_{j0},\ldots,s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$					
				$\beta = 1$		$\beta = 3$		$\beta = 5$	
ξ	μ	m	km	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)	mean($\hat{\xi}_m$)	std($\hat{\xi}_m$)
0	10^{-4}	20	U	0.043	0.086	0.042	0.079	0.044	0.084
			Exp	0.043	0.086	0.040	0.072	0.041	0.075
		50	U	0.030	0.053	0.027	0.045	0.028	0.048
			Exp	0.030	0.053	0.026	0.044	0.030	0.051
		100	U	0.023	0.039	0.021	0.034	0.021	0.034
			Exp	0.023	0.038	0.022	0.036	0.021	0.033
	10^{-3}	20	U	0.022	0.035	0.017	0.027	0.017	0.026
			Exp	0.021	0.035	0.016	0.026	0.017	0.028
		50	U	0.015	0.024	0.010	0.016	0.010	0.015
			Exp	0.014	0.021	0.010	0.015	0.011	0.017
		100	U	0.010	0.016	0.008	0.012	0.008	0.012
			Exp	0.010	0.016	0.008	0.013	0.008	0.012
0.5	10^{-4}	20	U	0.441	0.258	0.459	0.241	0.465	0.210
			Exp	0.452	0.234	0.468	0.207	0.471	0.204
		50	U	0.478	0.160	0.480	0.156	0.485	0.136
			Exp	0.477	0.162	0.480	0.147	0.489	0.122
		100	U	0.489	0.109	0.491	0.106	0.493	0.093
			Exp	0.488	0.113	0.491	0.104	0.489	0.105
	10^{-3}	20	U	0.496	0.079	0.497	0.078	0.497	0.083
			Exp	0.496	0.076	0.496	0.082	0.499	0.068
		50	U	0.498	0.055	0.499	0.045	0.499	0.047
			Exp	0.498	0.053	0.499	0.045	0.500	0.047
		100	U	0.499	0.039	0.500	0.032	0.499	0.033
			Exp	0.499	0.039	0.499	0.032	0.500	0.034
1	10^{-4}	20	U	0.947	0.322	0.970	0.278	0.968	0.286
			Exp	0.949	0.341	0.953	0.340	0.963	0.305
		50	U	0.976	0.213	0.988	0.182	0.985	0.192
			Exp	0.978	0.221	0.991	0.175	0.985	0.193
		100	U	0.988	0.147	0.993	0.134	0.993	0.127
			Exp	0.990	0.142	0.991	0.144	0.992	0.127
	10^{-3}	20	U	0.995	0.121	1.001	0.098	1.000	0.099
			Exp	0.999	0.107	0.997	0.107	0.999	0.107
		50	U	0.998	0.073	0.999	0.063	1.000	0.063
			Exp	1.000	0.071	0.999	0.066	0.999	0.069
		100	U	0.999	0.052	1.001	0.045	1.000	0.045
			Exp	1.000	0.044	1.001	0.049	1.000	0.049

Table B.27.: *Mean and standard deviation of $\hat{\beta}_m$ in the counting-maximum model; based on 10^4 replications (cf. Section 5.6.2)*

				$(s_{j0},\ldots,s_{j4}) = (0,4,8,12,\infty)$					
				$\beta = 1$		$\beta = 3$		$\beta = 5$	
ξ	μ	m	km	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)
0	10^{-4}	20	U	0.957	0.161	2.857	0.502	4.789	0.805
			Exp	0.953	0.173	2.845	0.502	4.785	0.790
		50	U	0.965	0.119	2.907	0.290	4.839	0.572
			Exp	0.962	0.115	2.893	0.331	4.850	0.495
		100	U	0.972	0.082	2.926	0.225	4.891	0.339
			Exp	0.972	0.085	2.927	0.237	4.880	0.391
	10^{-3}	20	U	0.961	0.093	2.940	0.170	4.911	0.259
			Exp	0.957	0.098	2.946	0.148	4.915	0.260
		50	U	0.972	0.061	2.959	0.107	4.946	0.164
			Exp	0.970	0.062	2.956	0.119	4.947	0.153
		100	U	0.978	0.048	2.971	0.075	4.961	0.117
			Exp	0.978	0.045	2.970	0.076	4.957	0.120
0.5	10^{-4}	20	U	1.113	0.381	3.178	0.843	5.294	1.414
			Exp	1.153	0.432	3.198	0.910	5.326	1.535
		50	U	1.050	0.226	3.085	0.573	5.117	0.842
			Exp	1.050	0.241	3.078	0.533	5.117	0.884
		100	U	1.023	0.156	3.045	0.385	5.046	0.587
			Exp	1.028	0.172	3.040	0.376	5.053	0.591
	10^{-3}	20	U	1.017	0.164	3.015	0.294	5.032	0.453
			Exp	1.017	0.168	3.011	0.278	5.021	0.442
		50	U	1.003	0.105	3.008	0.177	5.008	0.254
			Exp	1.007	0.100	3.005	0.180	5.014	0.292
		100	U	1.004	0.076	3.004	0.125	5.004	0.185
			Exp	1.002	0.074	3.003	0.130	5.003	0.197
1	10^{-4}	20	U	1.139	0.464	3.271	1.166	5.436	1.889
			Exp	1.140	0.463	3.319	1.277	5.336	1.707
		50	U	1.049	0.274	3.087	0.623	5.131	1.017
			Exp	1.061	0.294	3.064	0.618	5.123	1.003
		100	U	1.025	0.185	3.039	0.439	5.061	0.678
			Exp	1.025	0.190	3.040	0.426	5.097	0.734
	10^{-3}	20	U	1.010	0.167	3.017	0.336	5.037	0.537
			Exp	1.009	0.162	3.013	0.320	5.034	0.574
		50	U	1.006	0.119	3.006	0.223	5.012	0.308
			Exp	1.004	0.112	3.006	0.194	5.008	0.299
		100	U	1.002	0.077	3.005	0.149	5.005	0.225
			Exp	1.003	0.075	3.002	0.151	5.007	0.218

Table B.28.: *Mean and standard deviation of $\hat{\beta}_m$ in the counting-maximum model; based on 10^4 replications (cf. Section 5.6.2)*

				$(s_{j0}, \dots, s_{j6}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$					
				$\beta = 1$		$\beta = 3$		$\beta = 5$	
ξ	μ	m	km	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)	mean($\hat{\beta}_m$)	std($\hat{\beta}_m$)
0	10^{-4}	20	U	0.951	0.177	2.869	0.483	4.786	0.847
			Exp	0.949	0.177	2.870	0.435	4.792	0.752
		50	U	0.967	0.112	2.921	0.284	4.858	0.492
			Exp	0.963	0.113	2.921	0.270	4.849	0.512
		100	U	0.972	0.081	2.934	0.209	4.891	0.354
			Exp	0.971	0.081	2.930	0.227	4.898	0.350
	10^{-3}	20	U	0.968	0.074	2.949	0.163	4.922	0.264
			Exp	0.970	0.073	2.947	0.159	4.916	0.279
		50	U	0.978	0.050	2.969	0.096	4.951	0.149
			Exp	0.980	0.044	2.969	0.092	4.944	0.165
		100	U	0.985	0.033	2.975	0.070	4.962	0.116
			Exp	0.985	0.034	2.974	0.076	4.962	0.118
0.5	10^{-4}	20	U	1.096	0.354	3.173	0.866	5.235	1.254
			Exp	1.077	0.311	3.137	0.743	5.207	1.203
		50	U	1.035	0.206	3.083	0.559	5.104	0.791
			Exp	1.036	0.214	3.072	0.524	5.074	0.700
		100	U	1.020	0.143	3.031	0.372	5.047	0.540
			Exp	1.019	0.148	3.038	0.374	5.069	0.607
	10^{-3}	20	U	1.007	0.109	3.013	0.265	5.037	0.467
			Exp	1.008	0.105	3.020	0.282	5.014	0.355
		50	U	1.004	0.077	3.005	0.153	5.009	0.249
			Exp	1.003	0.073	3.004	0.155	5.007	0.247
		100	U	1.002	0.056	3.001	0.111	5.004	0.173
			Exp	1.002	0.054	3.002	0.112	5.003	0.182
1	10^{-4}	20	U	1.099	0.388	3.157	0.888	5.275	1.520
			Exp	1.108	0.416	3.216	1.077	5.316	1.614
		50	U	1.042	0.246	3.072	0.573	5.142	0.986
			Exp	1.043	0.260	3.066	0.544	5.130	0.993
		100	U	1.022	0.170	3.048	0.419	5.058	0.640
			Exp	1.018	0.164	3.047	0.446	5.056	0.647
	10^{-3}	20	U	1.013	0.148	3.010	0.283	5.023	0.446
			Exp	1.008	0.129	3.018	0.311	5.024	0.474
		50	U	1.003	0.087	3.005	0.182	5.009	0.283
			Exp	1.002	0.085	3.003	0.192	5.014	0.308
		100	U	1.002	0.063	3.000	0.131	5.005	0.201
			Exp	1.000	0.054	3.001	0.140	5.006	0.223

Table B.29.: *Optimal class limit configuration and corresponding quantiles with regard to the generalized Pareto distribution (cf. Section 5.5.4)*

		$\beta = 1$	
d	ξ	optimal class limits $(s_{11}, \ldots, s_{1,d-1})$	quantiles of class limits in %
3	0.0	(0.90, 4.38)	(59.16, 98.74)
	0.1	(0.86, 5.05)	(56.26, 98.32)
	0.2	(0.83, 5.81)	(53.63, 97.88)
	0.3	(0.80, 6.67)	(51.24, 97.44)
	0.4	(0.77, 7.65)	(49.07, 96.99)
	0.5	(0.75, 8.75)	(47.08, 96.54)
4	0.0	(0.83, 3.17, 5.61)	(56.35, 95.81, 99.64)
	0.1	(0.79, 3.43, 6.89)	(53.46, 94.75, 99.47)
	0.2	(0.76, 3.70, 8.45)	(50.91, 93.73, 99.29)
	0.3	(0.74, 3.99, 10.37)	(48.64, 92.76, 99.10)
	0.4	(0.71, 4.31, 12.74)	(46.60, 91.84, 98.91)
	0.5	(0.69, 4.66, 15.65)	(44.75, 90.97, 98.72)
5	0.0	(0.58, 1.54, 3.71, 6.01)	(43.84, 78.51, 97.54, 99.76)
	0.1	(0.54, 1.50, 4.09, 7.51)	(40.90, 75.26, 96.76, 99.63)
	0.2	(0.51, 1.46, 4.50, 9.38)	(38.38, 72.28, 95.97, 99.49)
	0.3	(0.48, 1.43, 4.95, 11.71)	(36.18, 69.53, 95.18, 99.34)
	0.4	(0.46, 1.40, 5.42, 14.61)	(34.23, 67.00, 94.40, 99.18)
	0.5	(0.43, 1.36, 5.93, 18.23)	(32.51, 64.66, 93.64, 99.02)
6	0.0	(0.51, 1.29, 2.96, 4.67, 6.82)	(40.08, 72.53, 94.84, 99.06, 99.89)
	0.1	(0.48, 1.26, 3.15, 5.41, 8.88)	(37.43, 69.34, 93.51, 98.68, 99.83)
	0.2	(0.45, 1.22, 3.33, 6.26, 11.60)	(35.17, 66.48, 92.23, 98.28, 99.75)
	0.3	(0.43, 1.19, 3.53, 7.24, 15.18)	(33.21, 63.91, 91.00, 97.86, 99.67)
	0.4	(0.41, 1.17, 3.74, 8.35, 19.91)	(31.49, 61.57, 89.84, 97.45, 99.58)
	0.5	(0.39, 1.14, 3.96, 9.63, 26.18)	(29.97, 59.44, 88.75, 97.04, 99.50)
7	0.0	(0.44, 1.06, 2.15, 3.65, 5.22, 7.31)	(35.55, 65.24, 88.36, 97.40, 99.46, 99.93)
	0.1	(0.41, 1.00, 2.13, 3.97, 6.18, 9.71)	(32.81, 61.46, 85.47, 96.46, 99.19, 99.89)
	0.2	(0.38, 0.95, 2.10, 4.29, 7.29, 12.92)	(30.49, 58.12, 82.68, 95.48, 98.88, 99.83)
	0.3	(0.35, 0.91, 2.07, 4.62, 8.56, 17.24)	(28.51, 55.15, 80.02, 94.50, 98.56, 99.77)
	0.4	(0.33, 0.87, 2.04, 4.97, 10.04, 23.02)	(26.79, 52.50, 77.50, 93.53, 98.23, 99.70)
	0.5	(0.31, 0.83, 2.01, 5.34, 11.76, 30.81)	(25.28, 50.11, 75.13, 92.57, 97.89, 99.63)
8	0.0	(0.38, 0.88, 1.65, 2.95, 4.26, 5.74, 7.78)	(31.57, 58.59, 80.82, 94.79, 98.59, 99.68, 99.96)
	0.1	(0.35, 0.83, 1.61, 3.11, 4.82, 7.02, 10.63)	(29.19, 55.12, 77.61, 93.33, 98.04, 99.51, 99.93)
	0.2	(0.33, 0.79, 1.58, 3.27, 5.42, 8.56, 14.61)	(27.20, 52.13, 74.72, 91.92, 97.46, 99.32, 99.89)
	0.3	(0.31, 0.76, 1.55, 3.43, 6.10, 10.43, 20.17)	(25.51, 49.51, 72.09, 90.56, 96.88, 99.12, 99.85)
	0.4	(0.29, 0.73, 1.53, 3.61, 6.84, 12.71, 27.98)	(24.05, 47.20, 69.68, 89.27, 96.29, 98.90, 99.81)
	0.5	(0.28, 0.70, 1.51, 3.79, 7.67, 15.49, 38.99)	(22.78, 45.14, 67.48, 88.05, 95.72, 98.69, 99.76)

Table B.30.: *Optimal class length Λ_{opt} and quantiles of the corresponding class limits with regard to the generalized Pareto distribution (cf. Section 5.5.4)*

		$\beta = 1$	
d	ξ	Λ_{opt}	quantiles of class limits in %
3	0.0	1.76	(82.72, 97.01)
	0.1	1.79	(80.75, 95.32)
	0.2	1.81	(78.68, 93.44)
	0.3	1.82	(76.61, 91.47)
	0.4	1.83	(74.61, 89.47)
	0.5	1.83	(72.68, 87.48)
4	0.0	1.45	(76.53, 94.49, 98.71)
	0.1	1.49	(75.12, 92.66, 97.53)
	0.2	1.51	(73.36, 90.63, 96.05)
	0.3	1.52	(71.51, 88.53, 94.39)
	0.4	1.53	(69.67, 86.43, 92.61)
	0.5	1.53	(67.88, 84.37, 90.78)
5	0.0	1.24	(71.17, 91.69, 97.60, 99.31)
	0.1	1.30	(70.42, 90.01, 96.25, 98.46)
	0.2	1.32	(69.07, 88.04, 94.61, 97.29)
	0.3	1.34	(67.51, 85.96, 92.82, 95.89)
	0.4	1.34	(65.89, 83.88, 90.94, 94.32)
	0.5	1.35	(64.27, 81.83, 89.03, 92.66)
6	0.0	1.09	(66.53, 88.80, 96.25, 98.74, 99.58)
	0.1	1.15	(66.42, 87.45, 94.87, 97.75, 98.95)
	0.2	1.19	(65.50, 85.64, 93.19, 96.44, 98.00)
	0.3	1.20	(64.23, 83.68, 91.35, 94.92, 96.79)
	0.4	1.21	(62.83, 81.67, 89.43, 93.27, 95.40)
	0.5	1.22	(61.39, 79.68, 87.49, 91.54, 93.89)
7	0.0	0.98	(62.50, 85.94, 94.73, 98.02, 99.26, 99.72)
	0.1	1.04	(62.98, 85.00, 93.45, 96.95, 98.50, 99.23)
	0.2	1.08	(62.45, 83.44, 91.80, 95.57, 97.44, 98.44)
	0.3	1.10	(61.46, 81.62, 89.97, 93.99, 96.14, 97.39)
	0.4	1.12	(60.27, 79.72, 88.06, 92.28, 94.68, 96.15)
	0.5	1.12	(58.99, 77.81, 86.12, 90.51, 93.10, 94.76)
8	0.0	0.89	(58.97, 83.16, 93.09, 97.17, 98.84, 99.52, 99.80)
	0.1	0.96	(59.96, 82.69, 92.01, 96.11, 98.01, 98.94, 99.41)
	0.2	1.00	(59.79, 81.39, 90.45, 94.70, 96.87, 98.06, 98.74)
	0.3	1.02	(59.07, 79.75, 88.67, 93.09, 95.51, 96.93, 97.82)
	0.4	1.04	(58.07, 77.97, 86.79, 91.36, 93.99, 95.62, 96.69)
	0.5	1.05	(56.94, 76.15, 84.88, 89.57, 92.37, 94.18, 95.41)

C. Figures

Figure C.1.: ***Frequency distribution and normal probability plot of*** $\sqrt{\frac{m}{2}}(\hat{D}_2 - 1)$ ***under Poisson hypothesis with sample sizes*** $m = 10$, 50, 1000; ***based on*** 10^6 ***replications (cf. Section 5.2.1).***

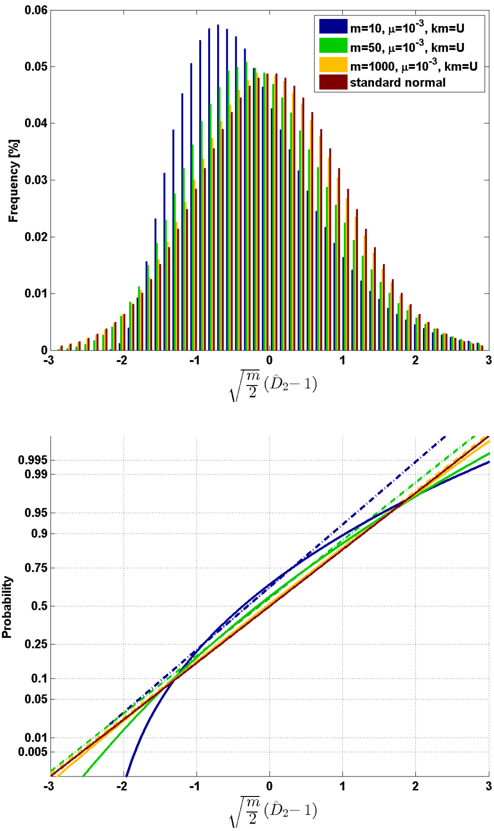
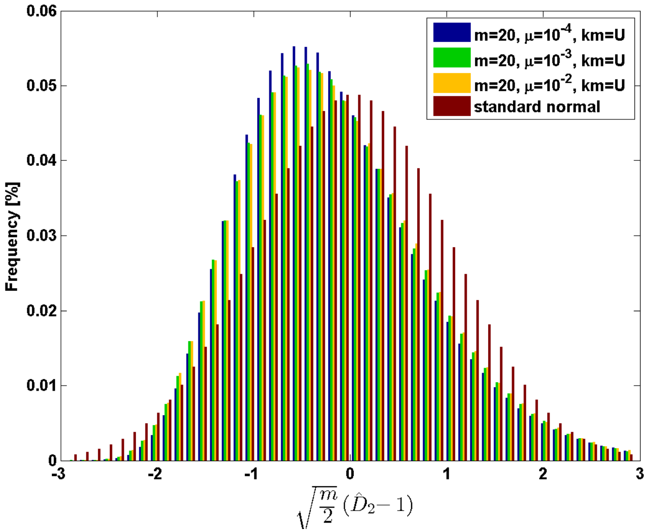


Figure C.2.: *Frequency distribution and normal probability plot of* $\sqrt{\frac{m}{2}}(\hat{D}_2 - 1)$ *under Poisson hypothesis with mean numbers of SOLEs per kilometer* $\mu = 10^{-4}$, 10^{-3}, 10^{-2}; *based on* 10^6 *replications (cf. Section 5.2.1).*



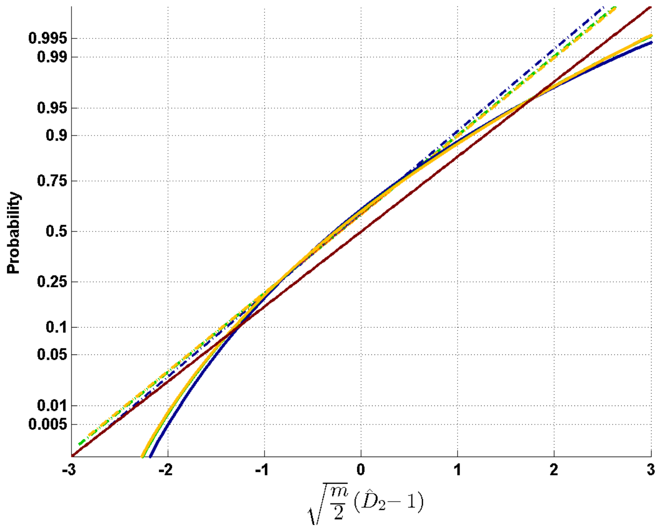
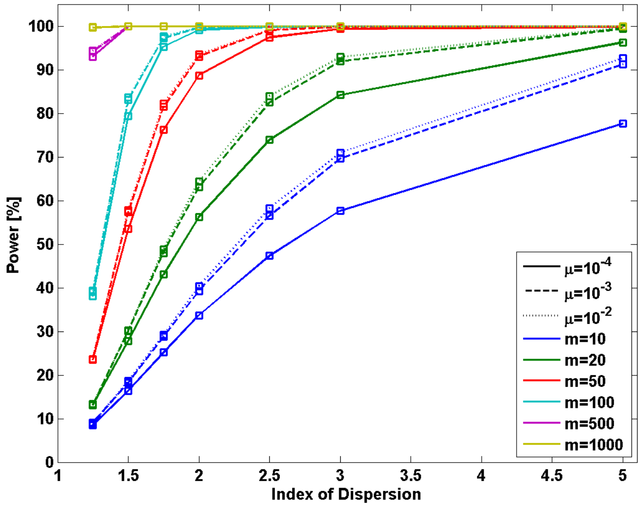


Figure C.3.: *Power of the test statistic* $\sqrt{\frac{m}{2}}(\hat{D}_2 - 1)$ *in % under several alternative hypotheses (*$\mathrm{IOD} > 1$ *and* $\mathrm{IOD} < 1$*) for significance level* $1 - \alpha = 0.95$*; based on* 10^5 *replications (cf. Section 5.2.1, Table B.3, Table B.4, Table B.5, "km=U").*



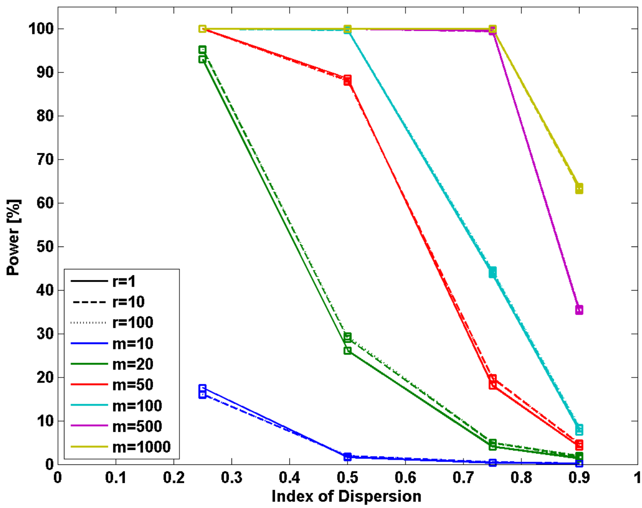
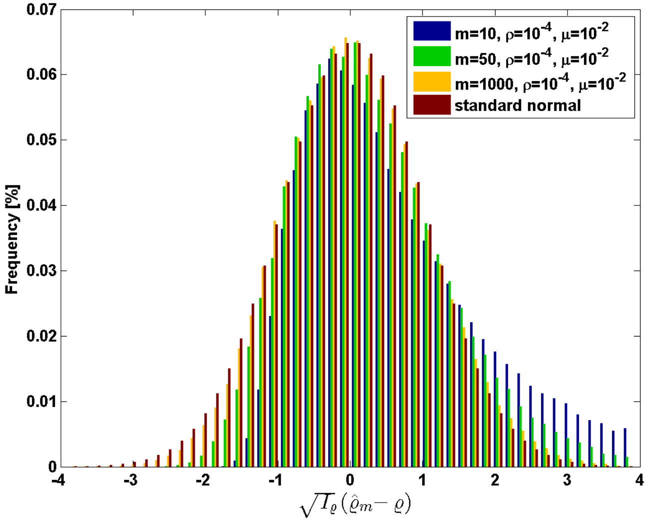


Figure C.4.: *Frequency distribution and normal probability plot of* $\sqrt{I_\varrho}\,(\hat{\varrho}_m - \varrho)$ *with sample sizes* $m = 10,\ 50,\ 1000$*; based on* 10^5 *replications (cf. Section 5.3.2, uniformly distributed mileages).*



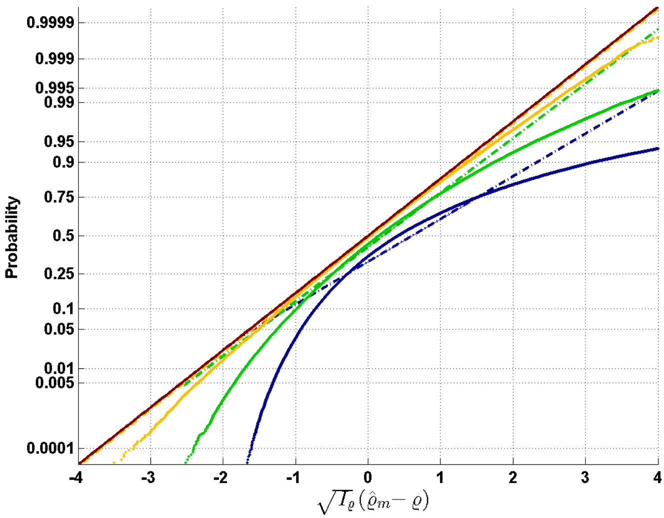
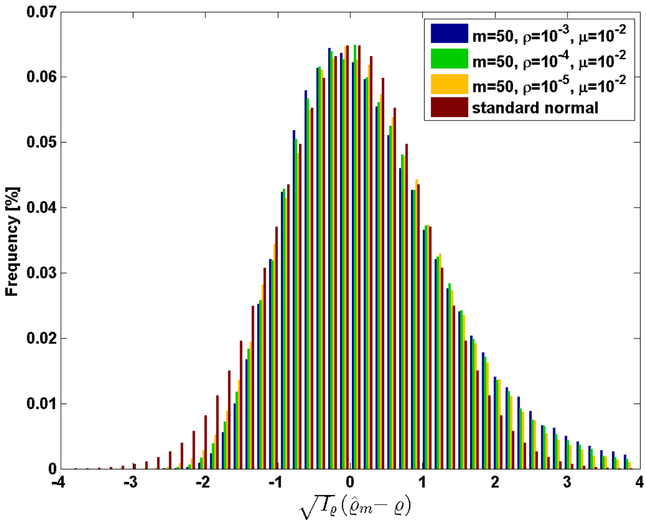


Figure C.5.: *Frequency distribution and normal probability plot of* $\sqrt{I_\varrho}\,(\hat{\varrho}_m - \varrho)$ *with exponents* $\varrho = 10^{-3}$, 10^{-4}, 10^{-5}*; based on* 10^5 *replications (cf. Section 5.3.2, uniformly distributed mileages).*



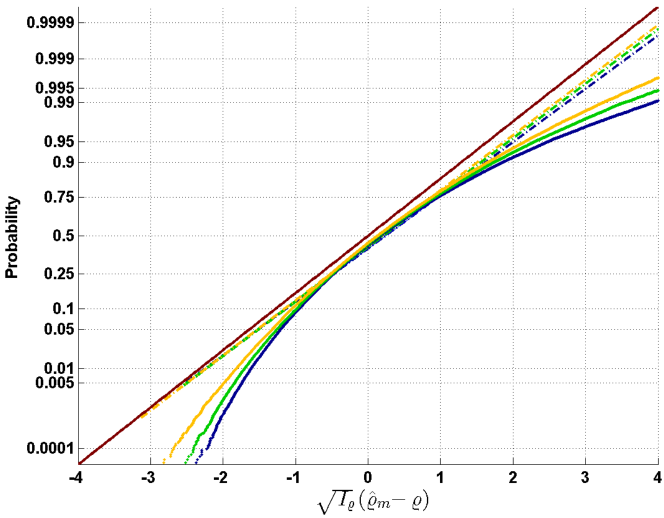


Figure C.6.: *Frequency distribution and normal probability plot of $\sqrt{I_\varrho}\,(\hat{\varrho}_m - \varrho)$ with means $\mu = 10^{-4},\ 10^{-3},\ 10^{-2}$; based on 10^5 replications (cf. Section 5.3.2, uniformly distributed mileages).*

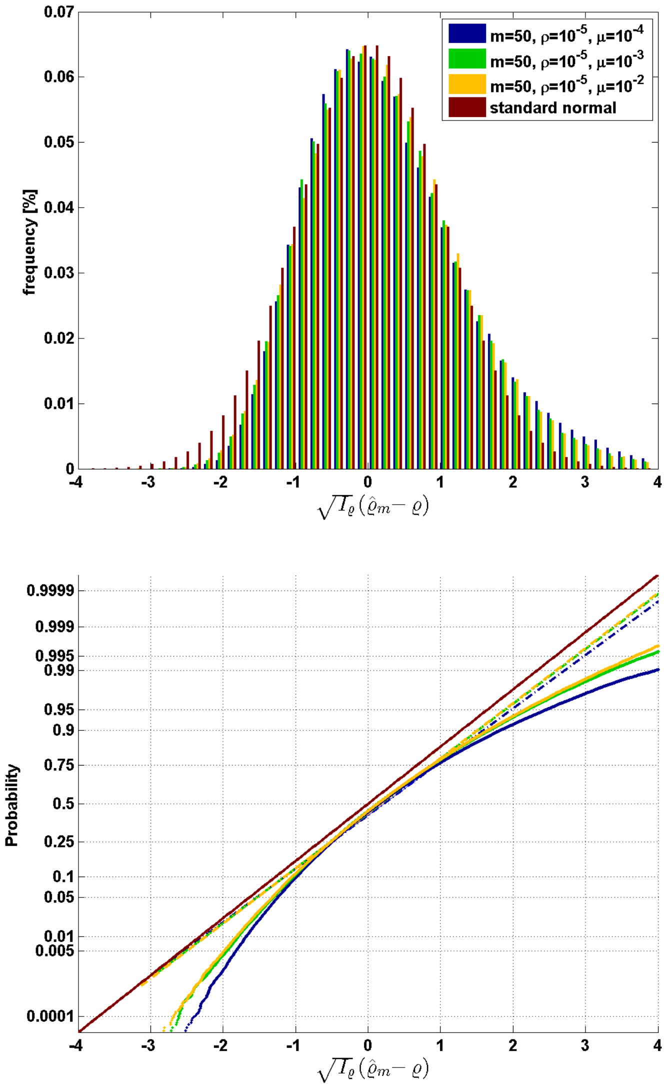
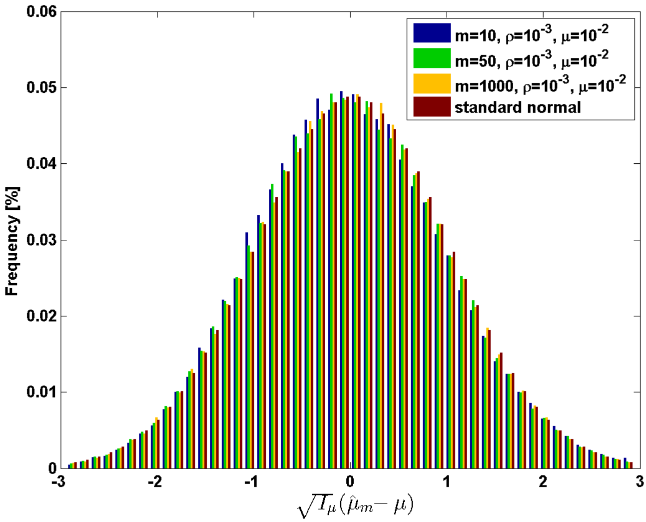


Figure C.7.: *Frequency distribution and normal probability plot of* $\sqrt{I_\mu}\,(\hat{\mu}_m - \mu)$ *with sample sizes* $m = 10,\ 50,\ 1000$*; based on* 10^5 *replications (cf. Section 5.4, uniformly distributed mileages).*



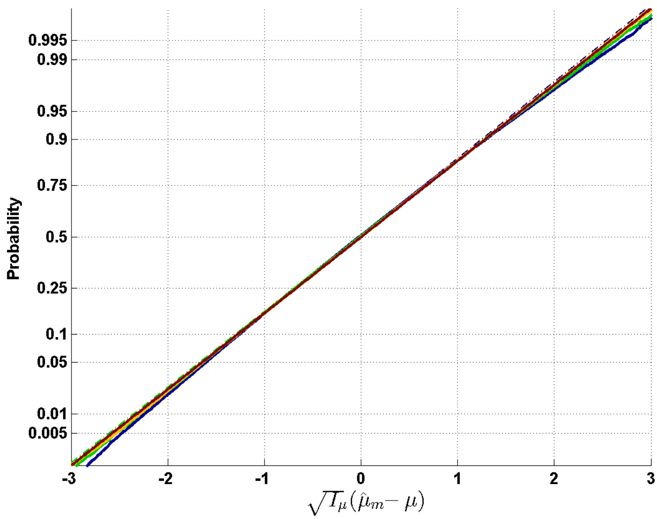
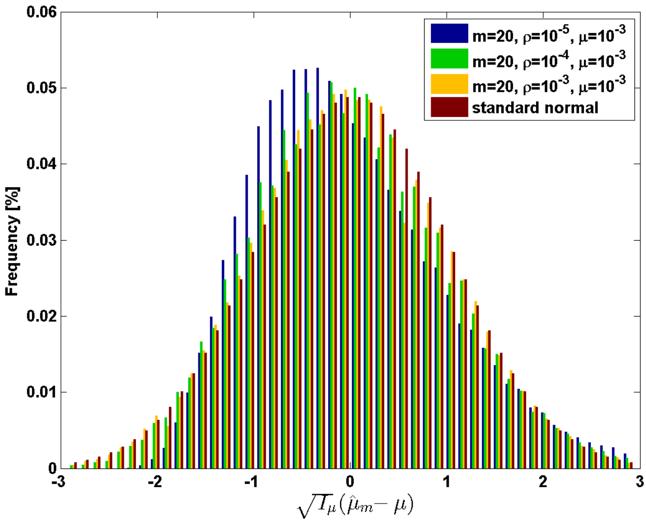


Figure C.8.: *Frequency distribution and normal probability plot of* $\sqrt{I_\mu}\,(\hat{\mu}_m - \mu)$ *with exponents* $\varrho = 10^{-5}$, 10^{-4}, 10^{-3}; *based on* 10^5 *replications (cf. Section 5.4, uniformly distributed mileages).*



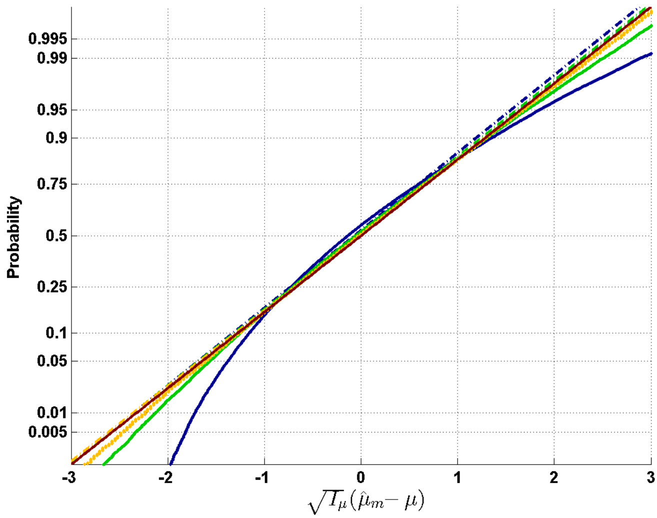
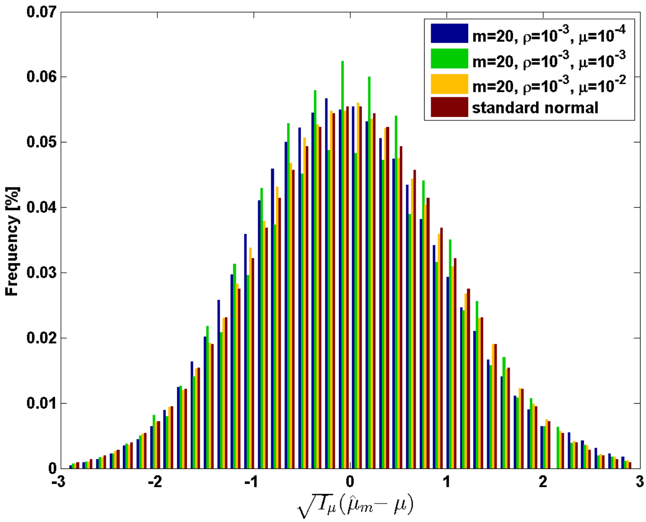


Figure C.9.: *Frequency distribution and normal probability plot of* $\sqrt{I_\mu}\,(\hat{\mu}_m - \mu)$ *with means* $\mu = 10^{-4}$, 10^{-3}, 10^{-2}; *based on* 10^5 *replications (cf. Section 5.4, uniformly distributed mileages).*



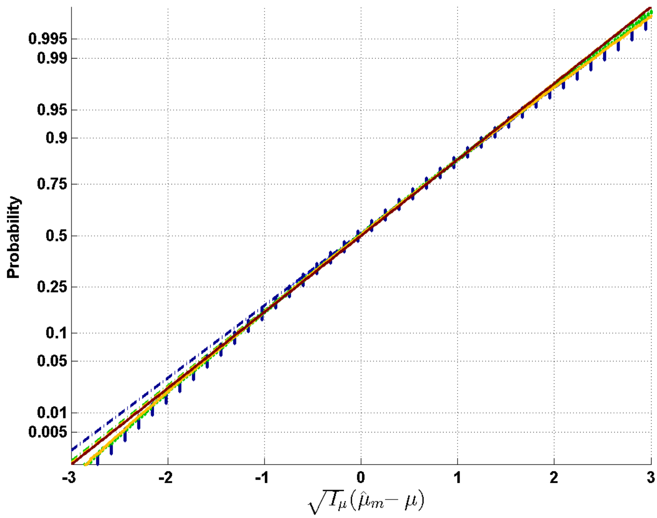


Figure C.10.: *Relative square root of inverse Fisher information concerning ϱ, $\frac{1}{\varrho\sqrt{I_\varrho}} = \frac{1}{\varrho\sqrt{I_{\mathrm{num}}(\varrho,\mu)_{11}}}$; based on 10^5 replications (cf. Section 5.3.2, Table B.6, "km=U").*

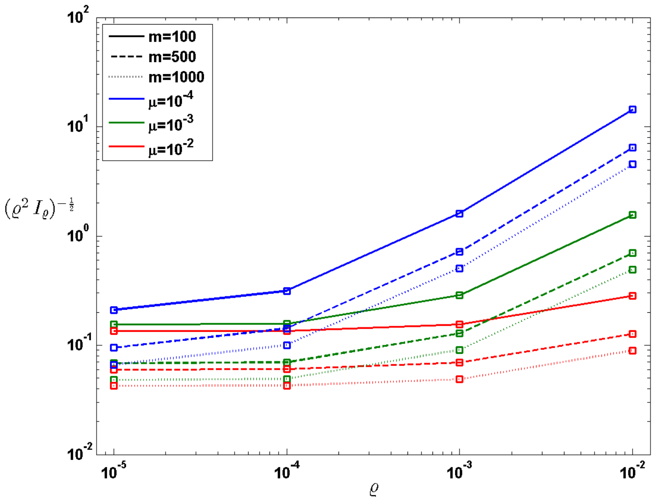


Figure C.11.: *Relative square root of inverse Fisher information concerning* μ, $\frac{1}{\mu\sqrt{I_\mu}} = \frac{1}{\mu\sqrt{I_{\text{num}(\varrho,\mu)22}}}$; *based on* 10^5 *replications (cf. Section 5.4, Table B.9, "km=U").*

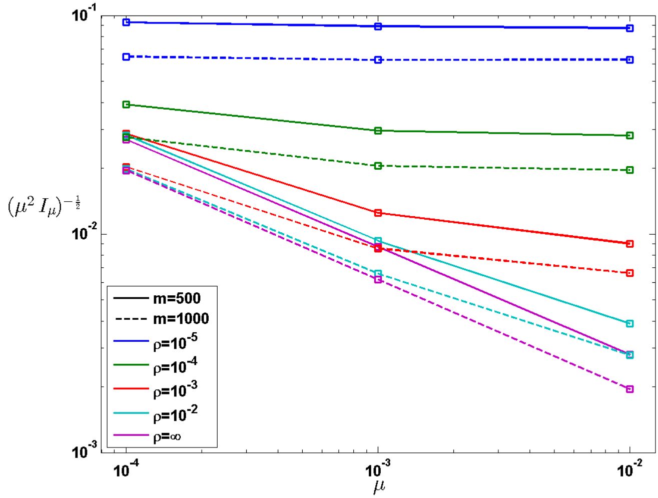


Figure C.12.: *Frequency distribution and normal probability plot of* $(\hat{\xi}_m - \xi)/\sqrt{J_\xi}$ *in the counting-model with sample sizes* $m = 20, 50, 100$*; based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages,* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$*).*

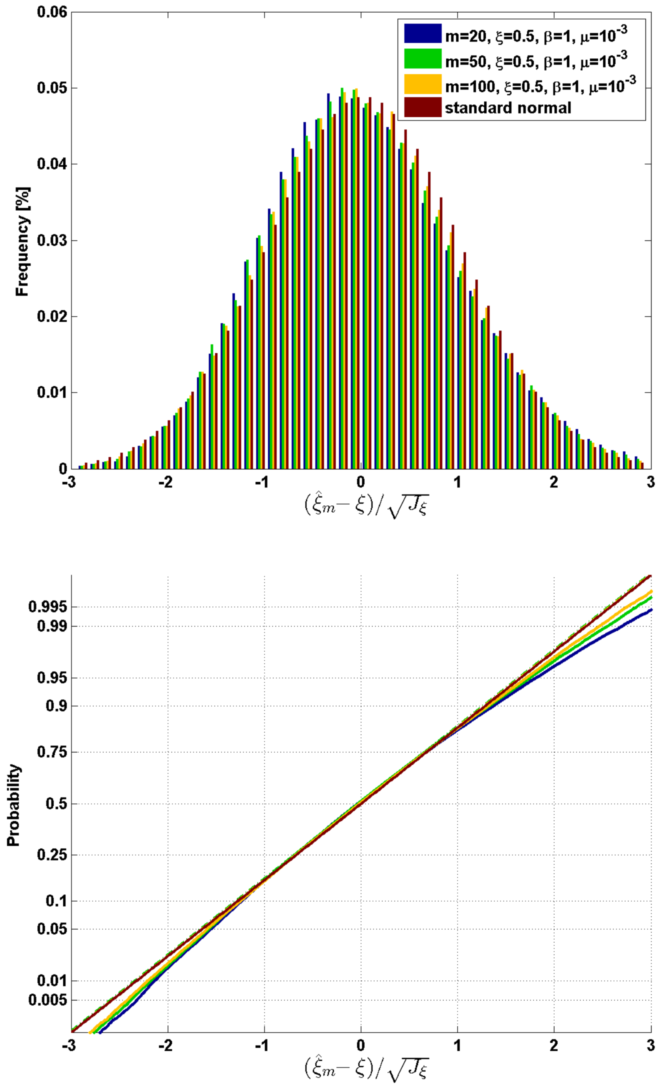
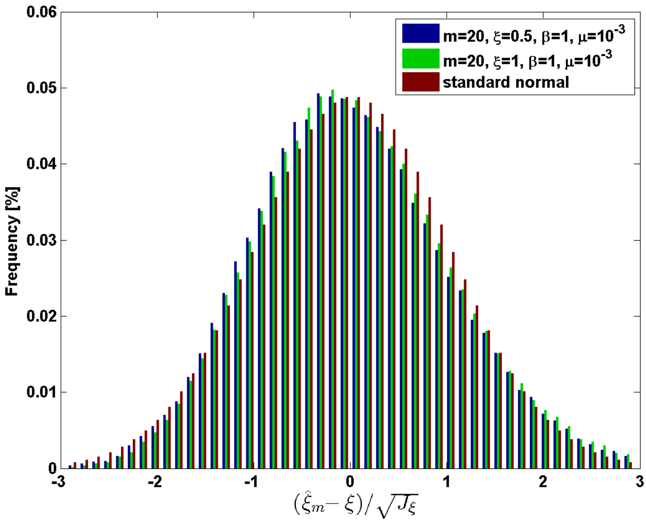


Figure C.13.: *Frequency distribution and normal probability plot of* $(\hat{\xi}_m - \xi)/\sqrt{J_\xi}$ *in the counting-model with shapes* $\xi = 0.5,\ 1$; *based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages,* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$*).*



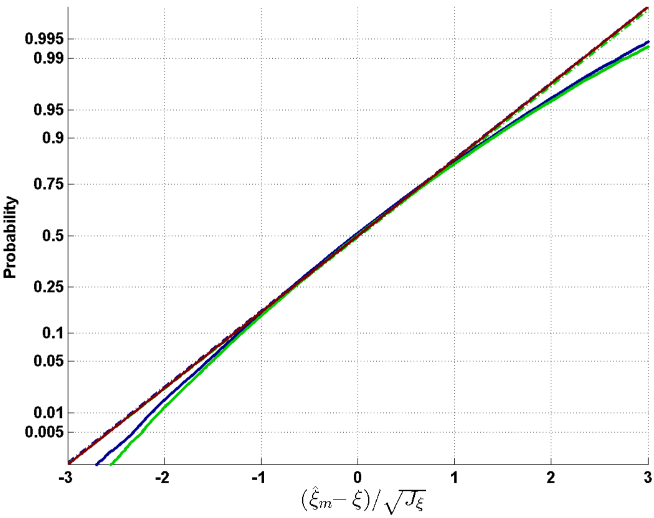


Figure C.14.: *Frequency distribution and normal probability plot of* $(\hat{\xi}_m - \xi)/\sqrt{J_\xi}$ *in the counting-model with scales* $\beta = 1, 3, 5$; *based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages,* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$*).*

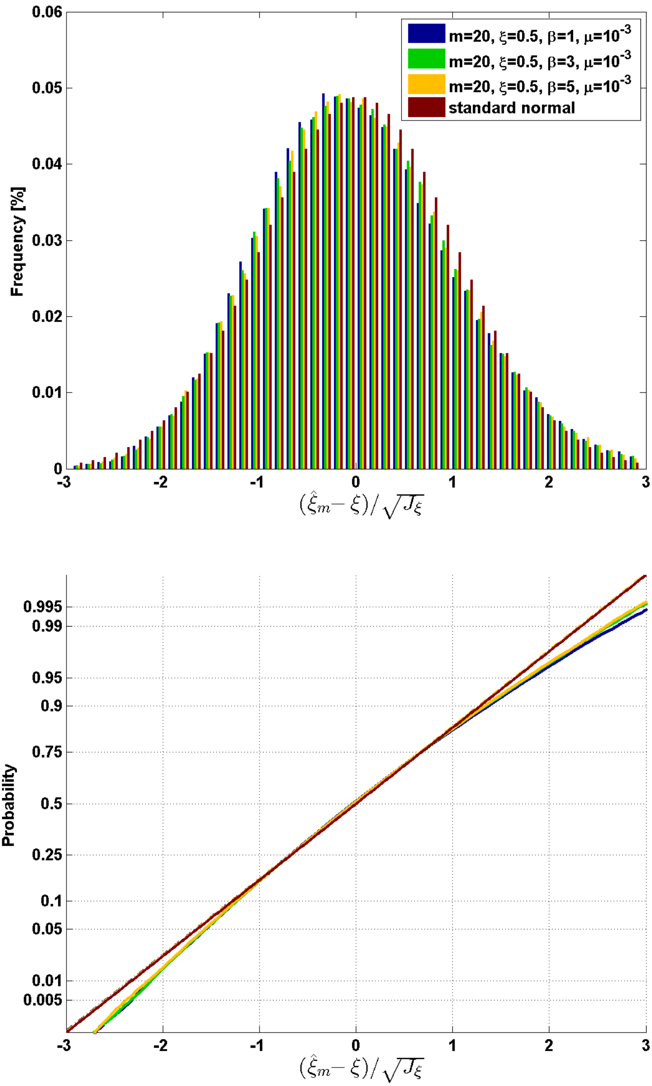


Figure C.15.: *Frequency distribution and normal probability plot of* $(\hat{\xi}_m - \xi)/\sqrt{J_\xi}$ *in the counting-model with class limits* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$, $(0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$; *based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages).*

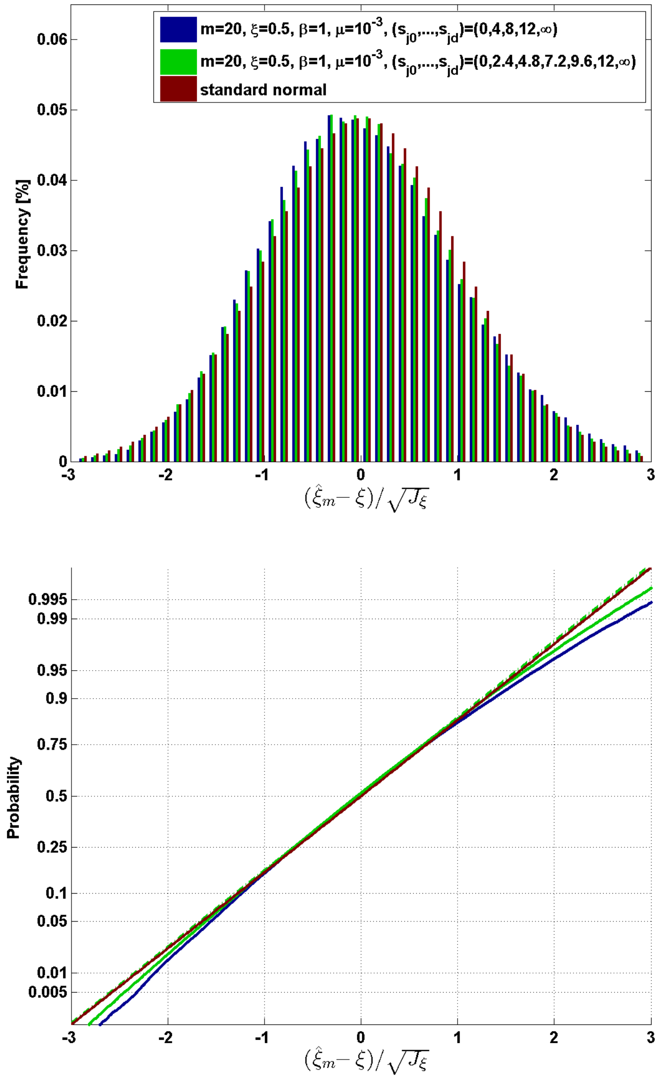
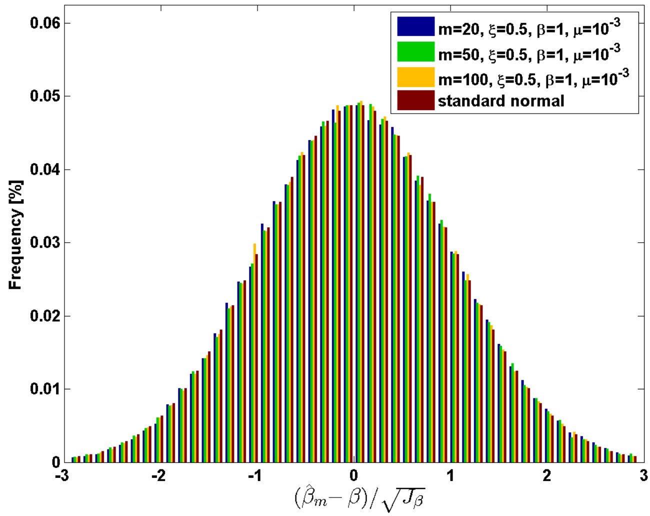


Figure C.16.: *Frequency distribution and normal probability plot of* $(\hat{\beta}_m - \beta)/\sqrt{J_\beta}$ *in the counting-model with sample sizes* $m = 20, 50, 100$*; based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages,* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$*).*



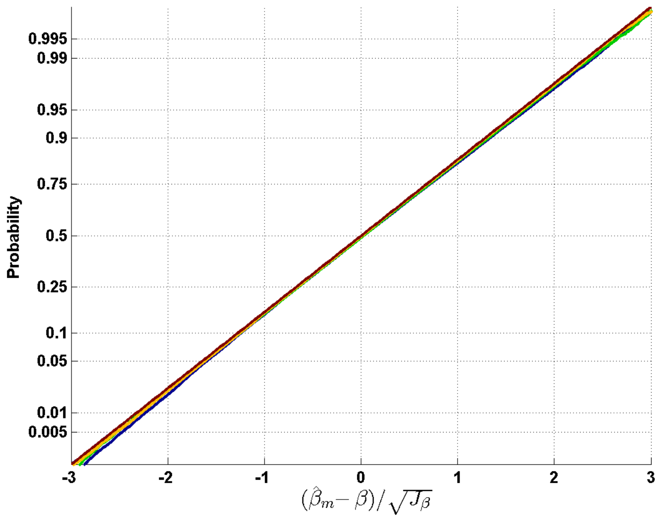
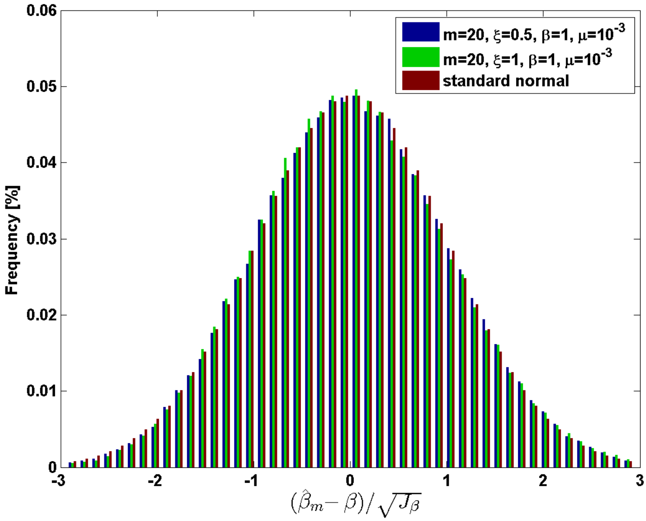


Figure C.17.: *Frequency distribution and normal probability plot of* $(\hat{\beta}_m - \beta)/\sqrt{J_\beta}$ *in the counting-model with shapes* $\xi = 0.5,\ 1$; *based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages,* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$*).*



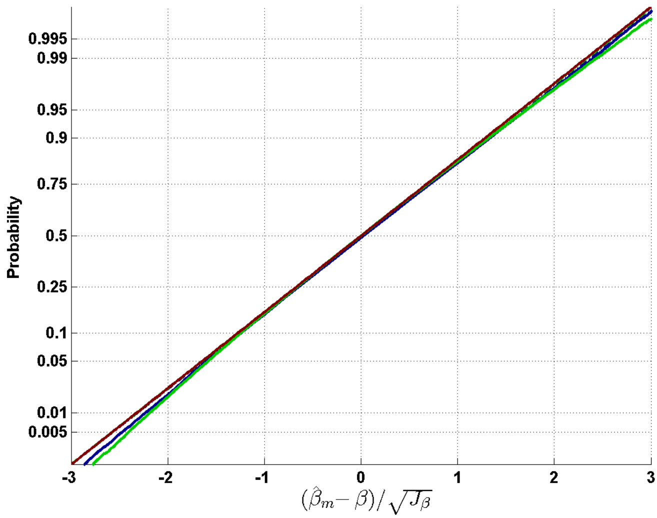
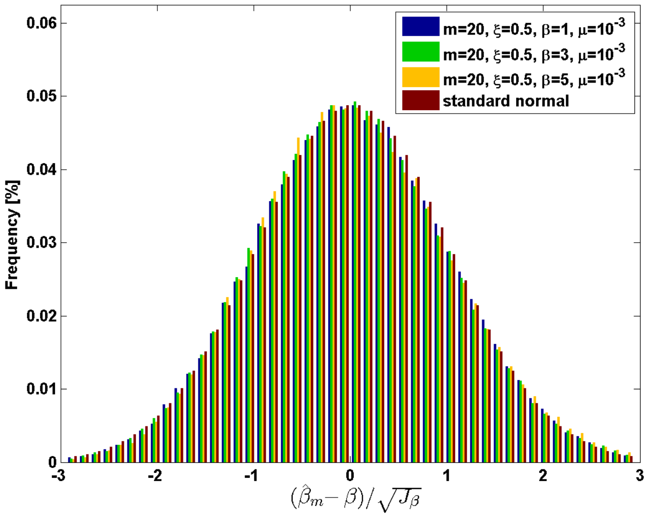


Figure C.18.: *Frequency distribution and normal probability plot of* $(\hat{\beta}_m - \beta)/\sqrt{J_\beta}$ *in the counting-model with scales* $\beta = 1,\ 3,\ 5$; *based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages,* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$*).*



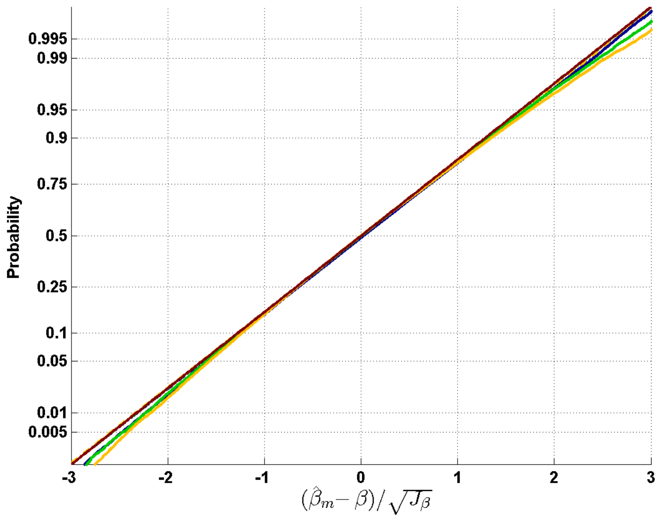


Figure C.19.: *Frequency distribution and normal probability plot of* $(\hat{\beta}_m - \beta)/\sqrt{J_\beta}$ *in the counting-model with class limits* $(s_{j0}, \ldots, s_{jd}) = (0, 4, 8, 12, \infty)$, $(0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$; *based on* 10^5 *replications (cf. Section 5.5.2, uniformly distributed mileages).*

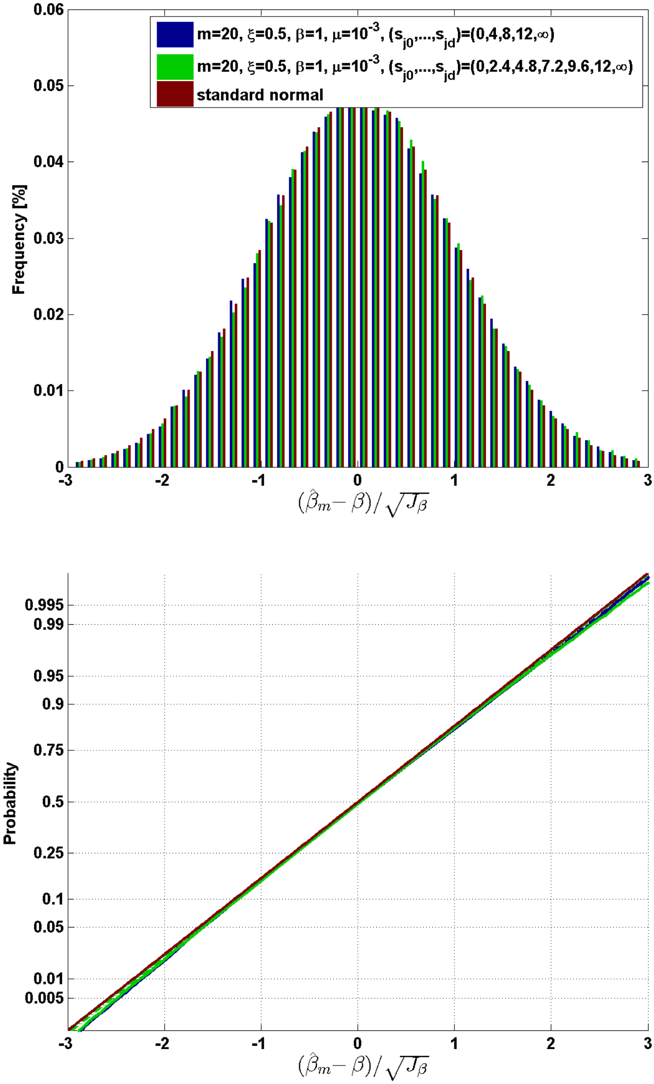


Figure C.20.: *Frequency distribution and cumulative distribution function of* $\hat{\xi}_m$ *in the counting-model: empirical (blue, Monte-Carlo simulation) and theoretical (green,* Φ_ξ*); based on* 10^5 *replications with* $\xi = 0$, $\beta = 1$, $\mu = 10^{-2}$, $m = 100$, $(s_{j0}, \ldots, s_{jd}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$, *uniformly distributed mileages (cf. Section 5.5.3).*

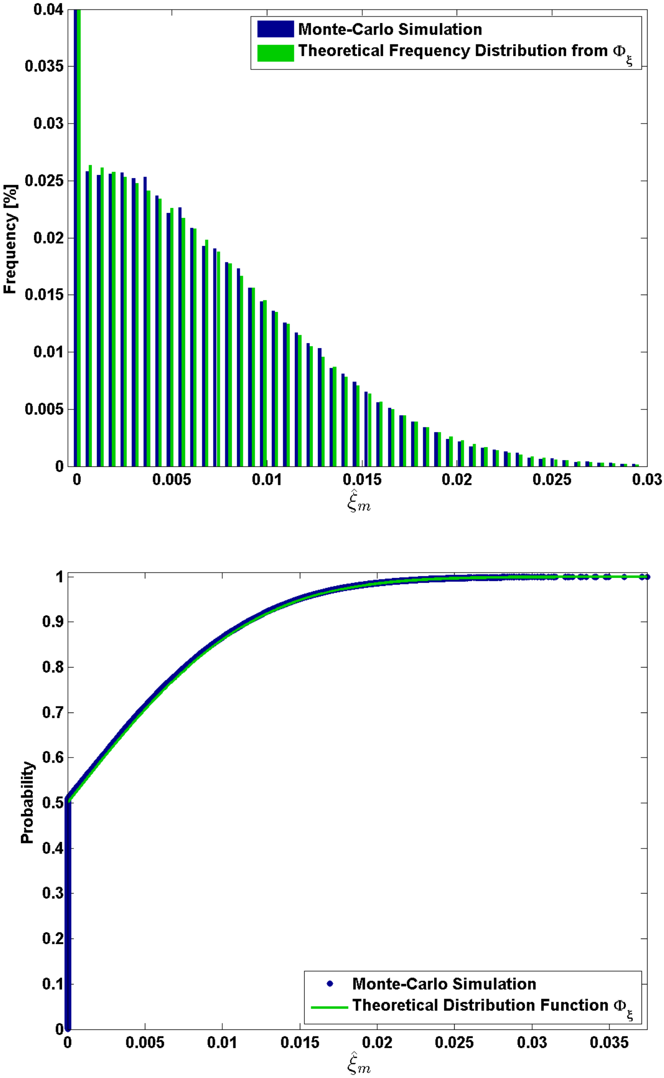
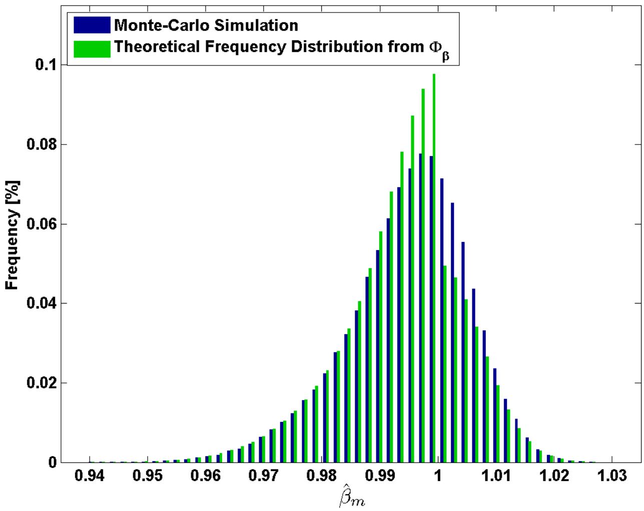


Figure C.21.: *Frequency distribution and cumulative distribution function of* $\hat{\beta}_m$ *in the counting-model: empirical (blue, Monte-Carlo simulation) and theoretical (green,* Φ_β*); based on* 10^5 *replications with* $\xi = 0$, $\beta = 1$, $\mu = 10^{-2}$, $m = 100$, $(s_{j0}, \ldots, s_{jd}) = (0, 2.4, 4.8, 7.2, 9.6, 12, \infty)$, *uniformly distributed mileages (cf. Section 5.5.3).*



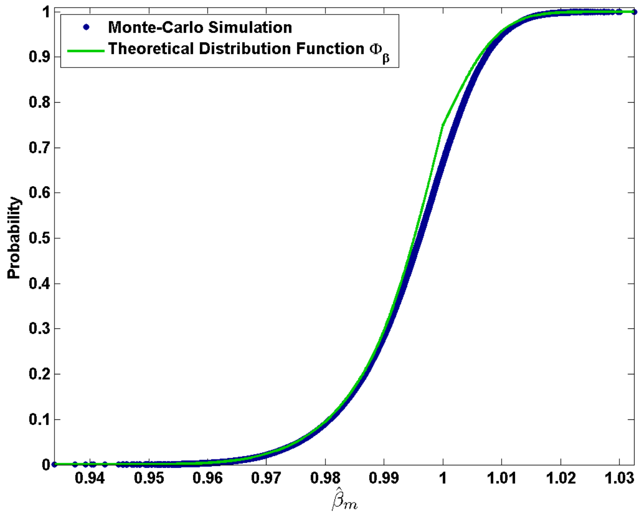
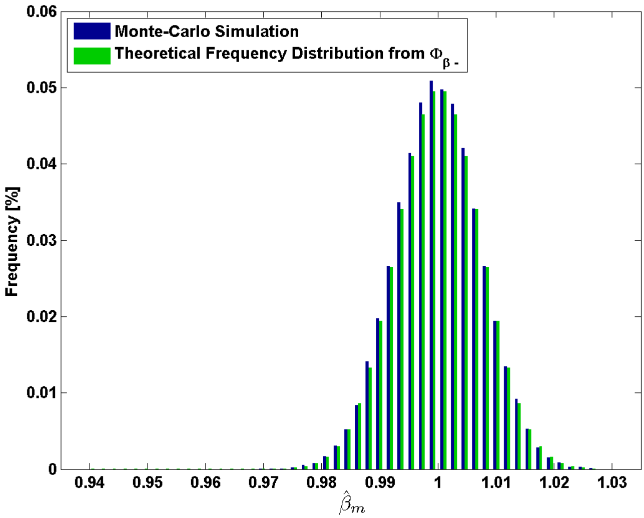


Figure C.22.: *Situation as in Figure C.21. Pictured are those realizations of $\hat{\beta}_m$ where it is $\hat{\xi}_m = 0$ simultaneously (blue), and the approximation of these realizations through $\Phi_{\beta-}$ (green, see Section 4.4.3) (cf. Section 5.5.3).*



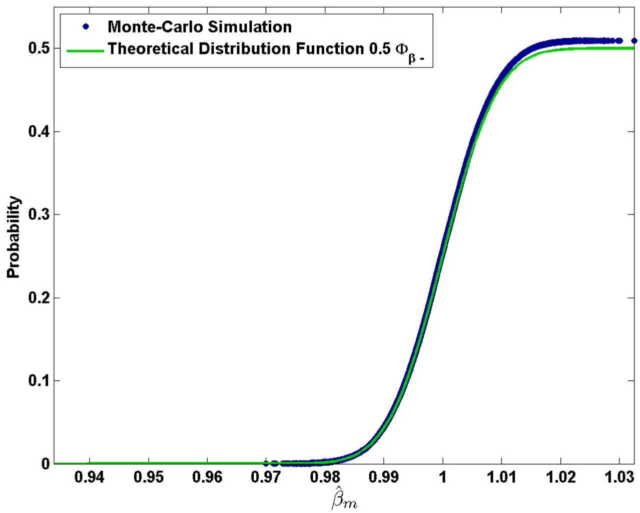
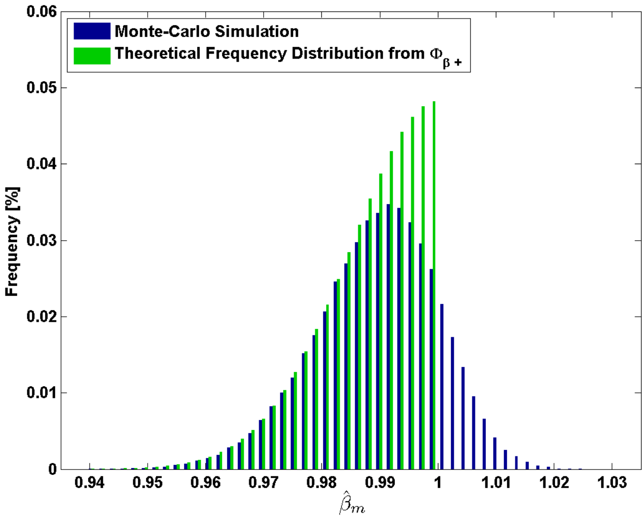
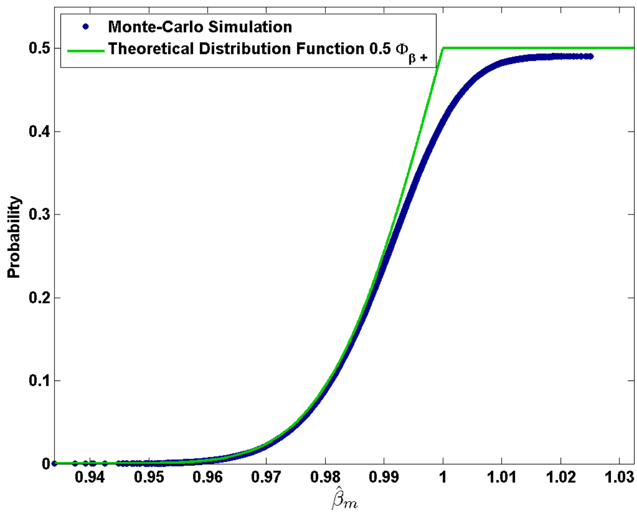


Figure C.23.: *Situation as in Figure C.21. Pictured are those realizations of $\hat{\beta}_m$ where it is $\hat{\xi}_m > 0$ simultaneously (blue), and the approximation of these realizations through $\Phi_{\beta+}$ (green, see Section 4.4.3) (cf. Section 5.5.3).*





Bibliography

[AB94] Ambagaspitiya, Rohana S.; Balakrishnan, Narayanaswamy: On the Compund Generalized Poisson Distribution. In: *ASTIN Bulletin* Vol. 24 (1994), pp. 255–263

[AD82a] Ahrens, Joachim H.; Dieter, Ulrich: Computer Generation of Poisson Deviates from Modified Normal Distributions. In: *ACM Transactions on Mathematical Software* Vol. 8 (1982, June), No. 2, pp. 163–179

[AD82b] Ahrens, Joachim H.; Dieter, Ulrich: Generating gamma variates by a modified rejection technique. In: *Communications of the ACM* Vol. 25 (1982, Jan.), No. 1, pp. 47–54

[AD92] Atkinson, A. Anthony C.; Donev, A. Alexander N.: *Optimum Experimental Designs* (*Oxford Statistical Science Series* Vol. 8). Oxford University Press, 1992

[AEE92] Aragón, Jorge; Eberly, David; Eberly, Shelly: Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution. In: *Statistics & Probability Letters* Vol. 15 (1992), pp. 375–379

[Ahm94] Ahmed, S. E.: Improved estimation of the coefficient of variation. In: *Journal of Applied Statistics* Vol. 21 (1994), No. 6, pp. 565–573

[Ald97] Aldrich, John: R. A. Fisher and the Making of Maximum Likelihood 1912-1922. In: *Statistical Science* Vol. 12 (1997, Aug.), No. 3, pp. 162–176

[Als05] Alsmeyer, Gerold: *Wahrscheinlichkeitstheorie.* Westfälische Wilhelms-Universität, Münster : Skripten zur Mathematischen Statistik Nr. 30, 2005. – 4th edn.

[Als06] Alsmeyer, Gerold: *Mathematische Statistik.* Westfälische Wilhelms-Universität, Münster : Skripten zur Mathematischen Statistik Nr. 36, 2006. – 3rd edn.

[Ans50] Anscombe, Francis J.: Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. In: *Biometrika* Vol. 37 (1950, Dec.), No. 3/4, pp. 358–382

[AS65] Abramowitz, Milton (ed.); Stegun, Irene A. (ed.): *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (*Applied Mathematics Series* Vol. 55). Dover Publications, 1965

[Bau02] Bauer, Heinz: *Wahrscheinlichkeitstheorie.* 5th edn. de Gruyter Lehrbuch, 2002

[Ben59] Bennett, B. M.: A Sampling Study on the Power Function of the χ^2 'Index of Dispersion' Test. In: *The Journal of Hygiene* Vol. 57 (1959, Sept.), No. 3, pp. 360–365

[BF93] Burden, Richard L.; Faires, J. D.: *Numerical Analysis.* 5th edn. PWS Publishing Company, 1993

[BG62] Bradley, Ralph A.; Gart, John J.: The asymptotic properties of ML estimators when sampling from associated populations. In: *Biometrika* Vol. 49 (1962, June), No. 1/2, pp. 205–214

[Bil69] Billinger, David R.: The calculation of cumulants via conditioning. In: *Annals of the Institute of Statistical Mathematics* Vol. 21 (1969), pp. 215–218

[Bor99] Borovkov, Alexandr A.: *Mathematical Statistics.* Taylor & Francis, 1999

[CGH+96] Corless, Robert M.; Gonnet, Gaston H.; Hare, D. E. G.; Jeffrey, David J.; Knuth, Donald E.: On the Lambert W function. In: *Advances in Computational Mathematics* Vol. 5 (1996), No. 1, pp. 329–359

[CJ73] Consul, Prem C.; Jain, Gopi C.: A Generalization of the Poisson Distribution. In: *Technometrics* Vol. 15 (1973, Nov.), No. 4, pp. 791–799

[Col07] Coles, Stuart: *An Introduction to Statistical Modeling of Extreme Values.* 4th edn. Springer-Verlag, 2007 (Springer Series in Statistics)

[Con89] Consul, Prem C.: *Generalized Poisson Distributions - Properties and Applications* (*Statistics: Textbooks and Monographs* Vol. 99). Marcel Dekker, Inc., 1989

[CR87] Cox, David R.; Reid, Nancy: Parameter Orthogonality and Approximate Conditional Inference. In: *Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 49 (1987), No. 1, pp. 1–39

[Cra62] CRAMÉR, Harald: *Mathematical Methods of Statistics.* Asia Publishing House, 1962

[Cro86] CROWDER, Martin: On Consistency and Inconsistency of Estimating Equations. In: *Econometric Theory* Vol. 2 (1986, Dec.), No. 3, pp. 305–330

[CS05] CERVANTES SAAVEDRA, Miguel de: *El ingenioso hidalgo don Quijote de la Mancha.* Vol. 1. Madrid, 1605

[Dev86] DEVROYE, Luc: *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986

[Dur10] DURRETT, Rick: *Probability: Theory and Examples.* 4th edn. Cambridge University Press, 2010 (Cambridge Series in Statistical and Probabilistic Mathematics)

[EH78] EFRON, Bradley; HINKLEY, David V.: Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information. In: *Biometrika* Vol. 65 (1978, Dec.), No. 3, pp. 457–482

[Eve02] EVERITT, Brian S.: *The Cambridge Dictionary of Statistics.* 2nd edn. Cambridge University Press, 2002

[FCW43] FISHER, Ronald A.; CORBET, A. S.; WILLIAMS, Carrington B.: The Relation between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. In: *Journal of Animal Ecology* Vol. 12 (1943), pp. 42–58

[Fis29] FISHER, Ronald A.: Moments and Product Moments of Sampling Distributions. In: *Proceedings of the London Mathematical Society* Vol. 30 (1929), No. 2, pp. 199–238

[Fis96] FISHMAN, George: *Monte Carlo: Concepts, Algorithms, and Applications.* Springer-Verlag, 1996 (Springer Series in Operations Research)

[For99] FORSTER, Otto: *Analysis 2: Differentialrechnung im* $\mathbb{R}^n$*, gewöhnliche Differentialgleichungen.* 5th edn. Vieweg, 1999

[For04] FORSTER, Otto: *Analysis 1: Differential- und Integralrechnung einer Veränderlichen.* 7th edn. Vieweg, 2004

[Geo07] GEORGII, Hans-Otto: *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik.* 3rd edn. Walter de Gruyter, 2007 (Gruyter Lehrbuch)

[GY20] Greenwood, Major; Yule, G. U.: An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Diseas or of Repeated Accidents. In: *Journal of the Royal Statistical Society* Vol. 83 (1920, Mar.), No. 2, pp. 255–279

[Hal41] Haldane, John B. S.: The Fitting Of Binomial Distributions. In: *Annals of Eugenics* Vol. 11 (1941, Jan.), pp. 179–181

[Hal99] Hald, Anders: On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares. In: *Statistical Science* Vol. 14 (1999, May), No. 2, pp. 214–222

[Hal00] Hald, Anders: The Early History of the Cumulants and the Gram-Charlier Series. In: *International Statistical Review / Revue Internationale de Statistique* Vol. 68 (2000, Aug.), No. 2, pp. 137–153

[HCBNM02] Huber-Carol, C. (ed.); Balakrishnan, N. (ed.); Nikulin, M. S. (ed.); Mesbah, M. (ed.): *Goodness-of-Fit Tests and Model Validity.* Birkhäuser, 2002 (Statistics for Industry and Technology)

[HF96] Hyndman, Rob J.; Fan, Yanan: Sample Quantiles in Statistical Packages. In: *The American Statistician* Vol. 50 (1996, Nov.), No. 4, pp. 361–365

[HF06] Haan, Laurens de; Ferreira, Ana: *Extreme Value Theory – An Introduction.* Springer-Verlag, 2006 (Springer Series in Operations Research and Financial Engineering)

[Hoa71] Hoadley, Bruce: Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case. In: *The Annals of Mathematical Statistics* Vol. 42 (1971), No. 6, pp. 1977–1991

[Hoe43] Hoel, Paul G.: On Indices of Dispersion. In: *Annals of Mathematical Statistics* Vol. 14 (1943, June), No. 2, pp. 155–162

[HW87] Hosking, Jonathan R. M.; Wallis, James R.: Parameter and Quantile Estimation for the Generalized Pareto Distribution. In: *Technometrics* Vol. 29 (1987, Aug.), No. 3, pp. 339–349

[IJ07] Ismail, Noriszura; Jemain, Abdul A.: Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. In: *Casualty Actuarial Society Forum* (2007, Winter), pp. 103–157

[Ina73] INAGAKI, Nobuo: Asymptotic Relations between the Likelihood Estimating Function and the Maximum Likelihood Estimator. In: *Annals of the Institute of Statistical Mathematics* Vol. 25 (1973), No. 1, pp. 1–26

[JKB94] JOHNSON, Norman L.; KOTZ, Samuel; BALAKRISHNAN, Narayanaswamy: *Continuous Univariate Distributions: Volume 1* (*Wiley Series in Probability and Mathematical Statistics* Vol. 288). 2nd edn. Wiley, 1994

[JKK05] JOHNSON, Norman L.; KOTZ, Samuel; KEMP, Adrienne W.: *Univariate Discrete Distributions* (*Wiley Series in Probability and Statistics* Vol. 444). 3rd edn. Wiley, 2005

[JZ05] JOE, Harry; ZHU, Rong: Generalized Poisson Distribution: the Property of Mixture of Poisson and Comparison with negative Binomial Distribution. In: *Biometrical Journal* Vol. 47 (2005), No. 2, pp. 219–229

[KK65] KENNEY, John F.; KEEPING, E. S.: *Mathematics of Statistics, Part II.* 2nd edn. Van Nostrand, 1965

[KN00] KOTZ, Samuel; NADARAJAH, Saralees: *Extreme Value Distributions: Theory and Applications.* Imperial College Press, 2000

[KS88] KACHITVICHYANUKUL, Voratas; SCHMEISER, Bruce W.: Binomial random variate generation. In: *Communications of the ACM* Vol. 31 (1988, Feb.), No. 2, pp. 216–222

[Kul61] KULLDORFF, Gunnar: *Estimation from Grouped and Partially Grouped Samples.* John Wiley & Sons, Almquist & Wiksell, 1961

[Law87] LAWLESS, Jerald F.: Negative Binomial and Mixed Poisson Regression. In: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* Vol. 15 (1987, Sep.), No. 3, pp. 209–225

[LC98] LEHMANN, Erich L.; CASELLA, George: *Theroy of Point Estimation.* 2nd edn. Springer-Verlag, 1998 (Springer Texts in Statistics)

[Lin05] LIN'KOV, Yu. N.: *Lectures in Mathematical Statistics: Parts 1 and 2* (*Translations of Mathematical Monographs* Vol. 229). American Mathematical Society, 2005

[LR77] LEVIN, Bruce; REED, James: Compound Multinomial Likelihood Functions are Unimodal: Proof of a Conjecture of I. J. Good. In: *The Annals of Statistics* Vol. 5 (1977, Jan.), No. 1, pp. 79–87

[LSWY02] Lee, Andy H.; Stevenson, Mark R.; Wang, Kui; Yau, Kelvin K. W.: Modeling young driver motor vehicle crashes: data with extra zeros. In: *Accident Analysis and Prevention* Vol. 34 (2002), No. 4, pp. 515–521

[Luk70] Lukacs, Eugene: *Characteristic Functions.* 2nd edn. Griffin, 1970

[LWI05] Lord, Dominique; Washington, Simon P.; Ivan, John N.: Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. In: *Accident Analysis and Prevention* Vol. 37 (2005), No. 1, pp. 53–46

[MAT12] MATLAB: *Version 7.14.0.739 (R2012a), Statistical Toolbox Version 8.0.* Natick, Massachusetts : The Mathworks, Inc., 2012

[Nel75] Nelson, David L.: Some Remarks on Generalizations of the Negative Binomial and Poisson Distributions. In: *Technometrics* Vol. 17 (1975, Feb.), No. 1, pp. 135–136

[NW06] Nocedal, Jorge; Wright, Stephen J.: *Numerical Optimization.* 2nd edn. Springer Verlag, 2006 (Springer Series in Operations Research and Financial Engineering)

[OPZ81] Olkin, Ingram; Petkau, A. J.; Zidek, James V.: A Comparison of n Estimators for the Binomial Distribution. In: *Journal of the American Statistical Association* Vol. 76 (1981, Sept.), No. 375, pp. 637–642

[OR02] Orme, Chris D.; Ruud, Paul A.: On the uniqueness of the maximum likelihood estimator. In: *Economics Letters* Vol. 75 (2002), pp. 209–217

[Pic75] Pickands, James I.: Statistical Inference Using Extreme Order Statistics. In: *The Annals of Statistics* Vol. 3 (1975, Jan.), No. 1, pp. 119–131

[Pie90] Piegorsch, Walter W.: Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. In: *Biometrics* Vol. 46 (1990, Sep.), No. 3, pp. 863–867

[Poi37] Poisson, Siméon D.: *Recherches sur la probabilité des jugements en matière criminelle et en matière civile.* Bachelier, 1837

[Puk06] Pukelsheim, Friedrich: *Optimal Design of Experiments* (*Classics in Applied Mathematics* Vol. 50). Society for Industrial and Applied Mathematics, 2006

[RB04] Ridout, Martin S.; Besbeas, Panagiotis: An empirical model for underdispersed count data. In: *Statistical Modelling* Vol. 4 (2004, Apr.), No. 1, pp. 77–89

[Sel65] Selby, George B.: The Index of Dispersion as a Test Statistic. In: *Biometrika* Vol. 52 (1965, Dec.), No. 3/4, pp. 627–629

[Slu25] Slutsky, Eugenius E.: Über stochastische Asymptoten und Grenzwerte. In: *Metron* Vol. 5 (1925), No. 3, pp. 3–89

[Smi84] Smith, Richard L.: Threshold Methods for Sample Extremes. In: Oliveira, J. T. (ed.): *Statistical Extremes and Applications*, D. Reidel Publishing Company, 1984, pp. 621–638

[Sor04] Sornette, Didier: *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. 2nd edn. Springer-Verlag, 2004 (Springer Series in Synergetics)

[SZ06] Savani, Vippal; Zhigljavsky, Anatoly A.: Efficient Estimation of Parameters of the Negative Binomial Distribution. In: *Communications in Statistics: Theory and Methods* Vol. 35 (2006), No. 5, pp. 767–783

[UC11] Upton, Graham; Cook, Ian: *Oxford Dictionary of Statistics*. 2nd edn. Oxford University Press, 2011

[Wan96] Wang, Yining: Estimation problems for the two-parameter negative binomial distribution. In: *Statistics & Probability Letters* Vol. 26 (1996), pp. 113–114

[Wei05] Weiss, Neil A.: *A Course in Probability*. Addison Wesley, 2005

[YHAV07] Yang, Zhao; Hardin, James W.; Addy, Cheryl L.; Vuong, Quang H.: Testing Approches for Overdispersion in Poisson Regression versus the Generalized Poisson Model. In: *Biometrical Journal* Vol. 47 (2007), No. 4, pp. 565–584

[ZGKW08] Zeichfüssl, Bernd; Grabenstein, Martin; Kirschner, Stefan; Witt, Robert: Ableitung von Sonderereignis-Lastannahmen für die Fahrzeugauslegung / Derivation of special event load assumptions for vehicle design. In: *MP Materialprüfung - MP material testing* Vol. 59 (2008), pp. 665–670

Index

Wissenschaftlicher Werdegang

Persönliche Daten

Name	Seifen
Vorname	Sebastian
Geburtsdatum	05. Januar 1985
Geburtsort	Gronau (Westf.)

Schulische Ausbildung

08/1991–07/1995	Buterlandschule (Grundschule) in Gronau (Westf.)
08/1995–06/2004	Alexander-Hegius-Gymnasium in Ahaus Abschluss: Abitur

Studium

10/2004–09/2009	Mathematikstudium an der Westfälischen Wilhelms-Universität WWU Münster Nebenfach: Physik Schwerpunktfächer: Stochastik, Mathematische Statistische Mechanik Abschluss: Diplom (Diplom-Mathematiker, Dipl.-Math.) Abschlussarbeit: „Verdünnte Curie-Weiss Modelle"
01/2010–09/2013	Promotionsstudium an der Technischen Universität Kaiserslautern am Fachbereich Mathematik, Arbeitsgruppe Statistik
01/2010–06/2013	Promotionsstipendium der Fraunhofer-Gesellschaft Stipendiat am Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM in Kaiserslautern, Abteilung Mathematische Methoden in Dynamik und Festigkeit MDF

Scientific Career

Personal Data

Surname	Seifen
First Name	Sebastian
Date of Birth	January 5, 1985
Birthplace	Gronau (Westf.)

School Education

08/1991–07/1995	Buterlandschule (elementary school) in Gronau (Westf.)
08/1995–06/2004	Alexander-Hegius-Gymnasium (grammer school) in Ahaus Graduation: Abitur

Scientific Education

10/2004–09/2009	Studies in mathematics at Westphalian Wilhelm University WWU of Münster Minor field of study: physics Special subject: stochastics, mathematical statistical mechanics Degree: Diplom (Diplom-Mathematiker, Dipl.-Math.) Diplom thesis: “Verdünnte Curie-Weiss Modelle” (Diluted Curie–Weiss Models)
01/2010–09/2013	Doctoral studies at Technical University Kaiserslautern, Department of Mathematics, Statistics Group
01/2010–06/2013	Doctoral scholarship from Fraunhofer Society Scholar at Fraunhofer Institute for Industrial Mathematics ITWM in Kaiserslautern, department Mathematical Methods in Dynamics and Durability MDF